Ex 1

6. JULI 1982

ISSN 0332-5903

Norwegian Geophysical Society

Geophysica Norvegica

VOL. 32 NO. 3 . 1982
UNIVERSITETSFORLAGET

Geophysica Norvegica

is a journal of geophysics, issued under the auspices of the Norwegian Geophysical Society. *Geophysica Norvegica* is mainly intended as a journal for Norwegian authors, but papers from other authors may be accepted provided that the work has been carried out at a Norwegian institution or its content has a special relevance to Norway.

PUBLISHER

Universitetsforlaget: P. O. Box 2959, Tøyen, Oslo 6, Norway
P. O. Box 258, Irvington-on-Hudson, New York 10533

SUBSCRIPTION

Geophysica Norvegica is published at irregular intervals. Order from the Publisher, Universitetsforlaget.

CONTRIBUTIONS

Manuscripts conforming with the rules on page 3 of the cover should be addressed to the editor. The editorial board will appoint referees, who will ensure that the paper meets a sufficiently high scientific standard.

A study of the wave climate in the Norwegian Sea. Algorithms in Markov models for deriving probabilities of certain events

R. FJØRTOFT
The Norwegian Meteorological Institute

Fjørtoft, R. A study of the wave climate in the Norwegian Sea. Algorithms in Markov models for deriving probabilities of certain events. *Geophysica Norvegica*, Vol. 32, No. 3, pp. 45–76, 1982.

A 'gamma'-probability distribution is used to fit the cumulative frequencies of *observed* significant wave-heights H. At different times of the year it is shown that these distributions differ only by scale factors on the H-axis, the annual variation of which is satisfactorily represented by an annual average plus an annual and semiannual periodic component. The frequency distributions for significant wave-heights *computed* from the model presently in use at the Norwegian Meteorological Institute, differ significantly from those based on instrument measured wave heights. A method is given by means of which these distributions may be corrected. – In Part II algorithms are developed, chiefly of iterative nature by which probabilities of certain events may be obtained in MARKOV models. In Part III this is applied to the study of the wave climate.

R. Fjørtoft, The Norwegian Meteorological Institute, P.O. Box 320 Blindern, Oslo 3, Norway.

INTRODUCTION

Statistical models are important tools to meet the needs of users wishing to estimate probabilities of important climatological events. On the continental shelf of the Norwegian Sea, wave height is *one* of the important geophysical variables. This study is limited to this variable.

Numerous events of different types connected with wave-height and its variation in time over shorter or longer intervals are of interest to users. Since statistical models, apart from a usually small number of parameters, have well-defined mathematical forms or implications, they provide means of great flexibility for arriving at estimates of probabilities. Furthermore, they are indispensable when dealing with events which occur so rarely that they are hardly observed in the existing series of observations. Finally, probabilities of all events covered by a model are completely defined after numerical values have been assigned to the parameters in the model, which will usually require fewer observations than if all these probabilities were to be estimated from a direct counting of the corresponding frequencies. However, obviously the choice of models must utilize all existing knowledge based upon observations as well as on theoretical evidence.

This paper falls into three parts:

In the first part, a *model* probability distribution of significant wave-height is developed, based upon the study of *observed* waves. It concludes with a method for correcting the frequency distributions of *computed* wave-height in the wave model at the Meteorological Institute.

In the second part, *Markov* models, chiefly of first order, are introduced for successive events. Certain algorithms are developed by means of which probabilities of a number of compound events can be obtained:

a) The duration of 'runs' of certain events, a run being defined as a successive sequence of occurrences of an event flanked by the complementary event.
b) Extremes in periods of arbitrary lengths.
c) The frequency of occurrence of certain events within periods of arbitrary lengths.
d) The means of variates taken over periods of arbitrary lengths.

In the third part the results of part one and two are applied to the waveclimate.

PART I

1. *The probability distribution function of significant wave height*

1.1 *Definitions*

Let H denote significant wave-height considered as a random variable, and h particular values of it. The model distribution function, F(h), of H, at a certain day τ of the year and at a certain location l, is supposed to satisfy the following general conditions:

$$f = F(h);\ 0 \leq h \tag{1)a}$$

F is continuous and strictly increasing in

$$[0, \infty > \tag{1)b}$$

$$F(0) = 0;\ F(\infty) = 1 \tag{1)c}$$

It follows that the inverse of F(h) exists:

$$h = H(f);\ 0 \leq f \leq 1 \tag{2}$$

The connection between F(h) and probability is

$$\Pr[H \leq h] = F(h) \tag{3}$$

If $h'' > h'$ are two fixed, but otherwise arbitrary values of H, we may write the event $[H \leq h'']$ as the union of the exclusive events $[H \leq h']$ and $[h' < H \leq h'']$. From the definition of probability we then obtain:

$$\Pr[H \leq h''] = \Pr[H \leq h'] + \Pr[h' < H \leq h'']$$

Using (3), it follows

$$\Pr[h' < H \leq h''] = F(h'') - F(h') \tag{4}$$

If we wish to express the dependency on the day of the year and on the location, we shall write

$$f = F(h;\ \tau, l) \tag{5}$$

The probabilities also depend on the hour of the day. This dependency is so weak, however, that it may be ignored except perhaps in summer close to the coast, where random variations in the sea-breeze wind system may influence the probability distribution of H significantly.

1.2 *The sampling problem*

Consider a sample of N observations from which we wish to estimate F for a given day of the year and hour of the day. Since the wave climate is known to vary over the year, only *one* observation, to be strict, should be taken from each year, and of course refer to the considered time of the year. In this time series over N years, suppose a number $n(H \leq h)$ of the N observations are observed to have values less or equal to a fixed value h, and a number $n(h' < H \leq h'')$ observed to have values between two arbitrary values h′ and h″. The corresponding observed relative frequencies are then

$$\frac{n(H \leq h)}{N} \equiv F_{obs}(h) \qquad (6)$$

$$\frac{n(h'' < H \leq h')}{N} = F_{obs}(h'') - F_{obs}(h') \qquad (7)$$

The notation on the right side of (7) is consistent with the notation defined in (6) because we have

$$n(H \leq h') + n(h' < H \leq h'') = n(H \leq h'')$$

As supported to some extent by experience, we adopt as a working hypothesis that these observed relative frequencies, by successive series of N observations for increasing N, gain sufficient stability to justify the relations

$$F_{obs}(h) \approx \Pr[H \leq h) = F(h) \qquad (8)$$

$$F_{obs}(h'') - F_{obs}(h') \approx \Pr[h' < H \leq h''] = F(h'') - F(h') \qquad (9)$$

when N is becoming large.

Besides serving to give the probability concept introduced more substance, these relations as written are meant to state that the *probability* of F_{obs} differing from the true value F of the underlaying probability by more than a given magnitude, however small, goes to zero when N goes to infinity. Relations (8), (9) represent an abstraction from the real world. In fact the increased stability of F_{obs} by increased N will certainly again get lost when N becomes so large that the influence of systematic long-periodic trends on F_{obs} can no longer be ignored.

1.3 *The estimation problem*

Relations (8), (9) tell us that the observed relative frequencies are estimates of the corresponding probabilities, and that the risk of making errors beyond a chosen limit decreases to zero with increasing N. How fast this decrease is, can be evaluated when we have a probability model for the *compound* event of N h-values. However, even in the favourable case of a *random* sample (statistical independence between the N results), the number N of years for which we have observations (3–25) are far from sufficient for estimates obtained in this way to have acceptable risks of errors.

In many estimation problems it is known that the distribution function belongs to a family $G(c_1, c_2, \ldots, c_2)$ of distributions. If the number of parameters is small, say one or two, even a small number of random observations may be sufficient to estimate the F with acceptable risks. In the case of waveheight, however, there is so far no or little such a priori knowledge. We must therefore, at least initially, be content with a display of the estimated cumulative frequencies, for instance by the F_{obs} defined in (8), for a limited number of h-values.

1.4 *The relation between distributions at different times of the year*

Consider the ratio between the inverses of the distribution functions at time τ and τ_1 of the year: $H(f;\tau)/H(f;\tau_1)$. This ratio must be a function of f, τ, τ_1, which we shall denote by $Q(f, \tau, \tau_1)$. Then

$$H(f;\tau) = Q(f, \tau, \tau_1)\, H(f; \tau_1) \qquad (10)$$

In the simplest case Q is independent of f:

$$H(f;\tau) = c(\tau, \tau_1)\, H(f;\tau_1) \qquad (11)$$

It will be shown in the next section that observations confirm these linear relations to a satisfactory degree. In doing so, it is convenient first to eliminate $c(\tau,\tau_1)$. This can be accomplished by applying an operator L on the H(f) values for $0 \leq f \leq 1$,

$$h_o(\tau) = L:[H(f;\tau)] \qquad (12)a$$

with the linear property

$$L:[cH] = cL:[H] \qquad (12)b$$

Using this operator on both sides of (11), we get

$$h_o(\tau) = c(\tau,\tau_1)h_o(\tau_1) \qquad (13)$$

Then, by dividing on both sides of (11) with the corresponding terms in (13), we obtain

$$\frac{H(f;\tau)}{h_o(\tau)} = \frac{H(f;\tau_1)}{h_o(\tau_1)} \qquad (14)$$

Hence $H(f;\tau)/h_o(\tau)$ is a function of f alone. Denoting this by $H^*(f)$ and its values by h^*, we obtain

$$h^* = H^*(f);\ 0 \leq f \leq 1 \qquad (15)$$

with the inverse

$$f^* = F(h^*),\ 0 \leq h^* \qquad (16)$$

Between the distributions $F(h;\tau)$ and $F^*(h^*)$ and between their inverses we have the relations

$$F(h;\tau) = F^*(h/h_o(\tau)) \qquad (17)$$

$$H(f;\tau) = h_o(\tau)H^*(f) \qquad (18)$$

In order to estimate $F^*(h^*)$ we need observations of h^*. Since $h^* = h(\tau)/h_o(\tau)$, and only h is observed, $h_o(\tau)$ needs to be estimated.

The linear operators which have been used are

$$h_o(\tau) = \mu(\tau) \equiv \int_0^1 H(f;\tau)df \qquad (19)a$$

$$h_o(\tau) = \mu'^{\frac{1}{2}}(\tau) \equiv \left(\int_0^1 H^2(f;\tau)df\right)^{\frac{1}{2}} \qquad (19)b$$

$$h_o(\tau) = \nu(\tau) \equiv \text{solution of}$$
$$0.5 = F(h;\tau) \qquad (19)c$$

which, when estimated by the corresponding sample means in (19)a, (19)b, and by the sample median in (19)c, have resulted in no significant differences in the final results.

1.5 *The observations*

The observations of wave-height in the open sea fall roughly into three categories:

a) Observations of measured wave-heights over a few years in a few geographical positions. The corresponding significant wave-heights are recorded for every 3 hours and in the unit 0.01 m.

b) A relatively long series of *visually* observed heights of wind-sea and swell at the weathership in position 66° N, 2° E, hereafter named MIKE. The series started in January 1949. The heights are recorded for every 3 hours, and given in the unit 0.5 m.

c) *Computed* significant wave-heights every 6 hours from the wave model of The Norwegian Meteorological Institute, using analyzed weather maps to get the input windfields in the model. (Haug 1968.) The unit is 0.1 m.

From these sources we have selected data as follows: Measured significant wave-heights every 24 hours as shown in Table 1 together with all-year frequencies of measured wave-heights at BRENT at 60°10′N, 1°40′ E.

Visually estimated significant wave-heights every 24 hours at MIKE at 60° N 2° E for the period Oct. 1953–Sept. 1977.

Computed significant wave-heights every 24 hours for the period Oct. 3, 1972–Sept. 30, 1977 in grid points named and located as follows: Ekofisk 56°37′ N, 2°43′ E. Statfjord 60°37′ N, 1°40′ E. Haltenbanken 64°59′ N, 1°44′ E. Mike 66°14′ N, 3°05′ E. Tromsøflaket 71°47′ N, 19°05′ E.

Table 1
Number of instrument-measured significant waveheights every 24 hours in different months and years at FAMITA, UTSIRA and AMI.

FAMITA 57°30′ N, 3° E	J	F	M	A	M	J	J	A	S	O	N	D
1969										31	30	31
70	31	28	31									
71											25	28
72	30		29	28					30	31	20	31
73	31	28	31	29						28	23	27
74	22	24	26						16	29	25	28
75	19	25	28							31	30	31
76	31	29	31									

UTSIRA 59°18′ N, 4°53′ E	J	F	M	A	M	J	J	A	S	O	N	D
1970							8	30	30	31	27	31
71	9	7	30			3	17					
72					5	30	31	31	30	31	30	31
73	31	28	31	30	31	8			30	31	30	31
74	31	28	30	30	30	24	5			20	30	31
75	26	27	16	22	21	3	15		7	31	2	2
76	5	28	28	17	26	30	19	9	7			4
77	25	24	15	30	23				7			

AMI 71°30′ N, 19° E	J	F	M	A	M	J	J	A	S	O	N	D
1976									20	28	24	26
77	25	25	27	29	29	26	25	21	22	25	26	25
78	28	25	24	26	26	26	28	20	25	27	23	27
79	24	25	23	27	21	25	28	26	19	22	27	26

1.6 *Observed cumulative frequencies*

In this section we shall test the 0-hypothesis that samples taken at different days of the year of the random variable h/h_o are from identical distributions $F^*(h/h_o)$. For h_o we shall take the first moment $\mu(\tau)$ estimated by the sample mean $\bar{h}(\tau)$. Because of the small number of years of observations we shall, at least initially, consider monthly groups of observations. Because of the slow variation of the distribution over the year, we may consider this test as applying approximately to the distributions at mid-month times τ_m, $m = 1,2,\ldots,12$.

These monthly grouped cumulative relative frequencies, denoted by $F_{obs}\left(\frac{h}{\bar{h}};m\right)$, are tabulated below for $\frac{h}{\bar{h}} = 0.3, 0.6,\ldots$ for FAMITA, UTSIRA, and for $\frac{h}{\bar{h}} = 0.2, 0.4,\ldots$, for AMI.

These tables show no obvious sign of any significant annual trend in the distribution of $h/\bar{h}$. The same is demonstrated by the curves in Fig. 1 a–d which connect the monthly $h/\bar{h}$-values corresponding to a number of selected values of $F^*_{obs}(h/\bar{h};m)$. The variations from month to month give the impression of rather erratic variations around the annual means. This is contrary to the observed distribution of h, where the erratic varia-

Table 2 (a)–(d)

Cumulative observed frequencies of $h/\bar{h}$ in percent based on observed waves for FAMITA, UTSIRA, AMI, and MIKE for calendar months. The next to bottom row gives the monthly sample means, $\bar{h}$, of sign. wave height, h, in metres, and the bottom row the total number of observations used for each month.

(a) FAMITA

$h/\bar{h}$	J	F	M	A	M	J	J	A	S	O	N	D
0.3	3	1	3						0	5	3	2
0.6	20	22	25	19					22	28	20	18
0.9	52	51	53	54					50	56	52	47
1.2	66	69	71	70					72	71	72	70
1.5	84	84	81	86					85	80	86	90
1.8	94	92	91	95					93	88	94	95
2.1	96	98	95	95					100	95	98	99
2.4	99	100	97	97						97	98	99
2.7	99		98	98						97	99	100
3.0	100		99	100						99	99	
3.3			99							99	100	
3.6			99							100		
3.9			99									
4.2			100									
$\bar{h}$	3.29	2.31	2.42	2.60					1.91	2.16	3.28	3.39 m
N	163	134	174	57					46	152	162	177

(b) UTSIRA

$h/\bar{h}$	J	F	M	A	M	J	J	A	S	O	N	D
0.3	2	3	3	3	8	1	0	9	4	2	1	4
0.6	32	29	28	25	32	19	24	24	29	29	24	24
0.9	49	48	56	50	56	49	55	47	51	57	50	44
1.2	69	69	72	71	71	70	77	66	67	74	72	69
1.5	83	85	84	83	81	89	87	77	85	85	85	85
1.8	90	90	91	92	91	96	92	87	90	90	91	92
2.1	97	95	92	98	95	96	93	96	94	93	95	98
2.4	99	99	97	99	97	98	96	98	96	96	98	100
2.7	100	99	98	99	98	100	98	98	99	96	99	
3.0		99	99	99	98		98	100	100	97	99	
3.3		100	99	99	98		99			99	100	
3.6			99	100	98		100			99		
3.9			99		98					100		
4.2			99		98							
4.5			100		99							
4.8					99							
5.1					100							
$\bar{h}$	2.30	1.87	1.77	1.91	1.22	1.17	1.03	0.98	1.55	1.91	2.51	2.82 m
N	127	143	149	129	136	98	95	70	113	144	119	130

(c) AMI

$h/\bar{h}$	J	F	M	A	M	J	J	A	S	O	N	D
0.2	0	0	0	0	0	0	0	0	0	0	0	0
0.4	4	1	0	7	11	4	2	3	5	6	3	4
0.6	17	19	11	23	28	22	5	13	20	21	24	16
0.8	47	41	31	45	45	44	30	42	42	40	47	42

$h/\bar{h}$	J	F	M	A	M	J	J	A	S	O	N	D
1.0	60	63	54	55	63	56	51	64	60	57	63	62
1.2	75	71	74	73	71	71	77	76	73	74	73	74
1.4	82	80	88	78	80	81	86	82	80	82	81	86
1.6	88	85	95	88	83	88	96	88	84	89	87	90
1.8	94	95	96	93	88	92	98.8	94	90	93	90	92
2.0	97.4	98,7	97.3	94	92	96	100	95.5	92	94	92	94
2.2	97.4	98.7	98.7	95	96	96		95.5	93	95	95	94
2.4	97.4	98.7	100	97.6	98.7	98.7		97	96.5	96	96	95
2.6	98.7	98.7		98.8	98.7	100		97	98.8	98	97	97
2.8	98.7	100		100	98.7			98.5	100	100	98	98
3.0	98.7				100			98.5			99	100
3.2	98.							98.5			100	
3.4	98.7							100				
3.6	100											
$\bar{h}$	3.08	2.75	2.56	2.05	1.75	1.66	1.19	1.44	1.95	2.38	2.93	2.90 m
N	77	75	74	82	76	77	81	67	86	102	100	104

(d) MIKE

$h/\bar{h}$	J	F	M	A	M	J	J	A	S	O	N	D
0.3	1	0	3	4	12	3	1	7	6	4	1	0
0.6	21	19	25	11	24	25	14	22	22	25	15	32
0.9	56	51	44	50	53	53	44	44	64	44	50	46
1.2	69	72	81	79	71	76	69	74	79	74	71	75
1.5	85	86	92	88	83	92	86	92	92	88	88	89
1.8	90	94	93	92	96	98.6	92	98.6	96	96	94	92
2.1	94	97.2	94	97	98.6	100	97	98.6	96	97	98.6	94
2.4	97	97.2	97.2	98.6	98.6		100	100	97	100	98.6	98.6
2.7	100	98.6	98.6	98.6	98.6				98.6		100	100
3.0		100	100	98.6	98.6				100			
3.3				100	100							
$\bar{h}$	2.91	2.92	2.73	2.21	1.85	1.60	1.67	1.83	2.32	2.58	2.76	2.73 m
N	72	72	72	72	72	72	72	72	72	72	72	72

tions are superimposed upon a distinct annual trend, as shown for instance for MIKE in Fig. 2.

For an objective test of the 0-hypothesis that the distribution of $h/\bar{h}$ is independent of the time of the year, we have studied samples from MIKE of size 24 which consist of one observation a day over 24 years for the 11th of each month. For h_0 we have used the first moment estimated by the means $\bar{h}$ tabulated in Table 2. This gives the ordered sequences tabulated in Table 3. (Since the unit in the MIKE-observations is 0.5 m, many observations will have equal values. These have been made different by a randomization process, using a uniform distribution in each 1/2-m interval.)

In Table 4, 16 pairs of these samples have been combined into ordered samples of size 48, the ordering having been indicated by letters 0 and X, 0 representing a summer-month observation and X a winter-month observation.

From this table, it is seen that the observations seem to be well mixed. Using the number D of 'runs', defined in the introduction, as the test statistics of the 0-hypothesis, we find that for pairs of random and mutually independent sam-

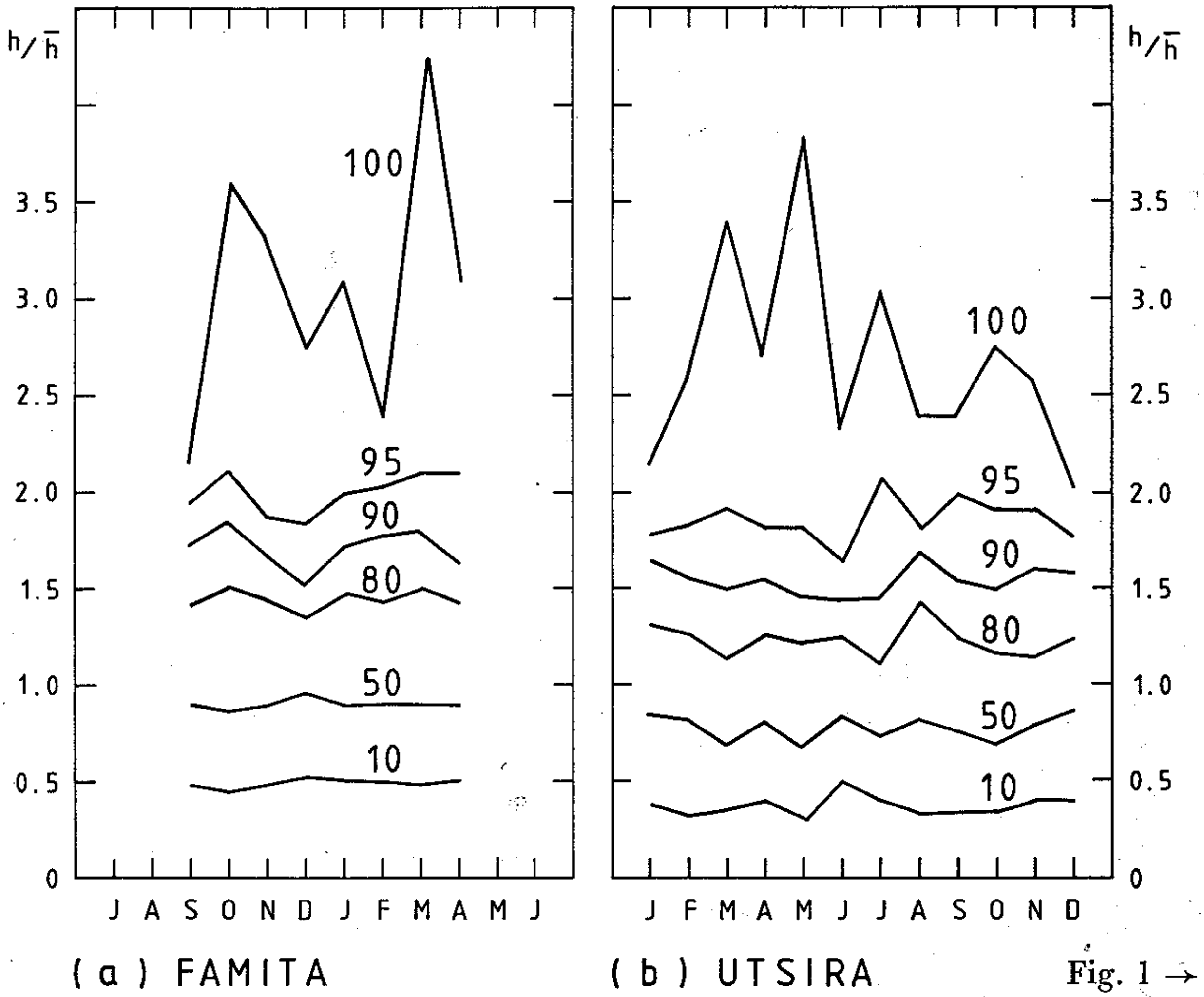


Fig. 1 →

ples, when the sample size is 24, D has expectation and standard deviation

$$E[D] = 25;\ \sigma_D \approx 3.5$$

Furthermore, this test statistics is approximately normally distributed when the sample size is greater than ten. (See, for example, Mood 1963.) The observed values of D are shown at the bottom row of the table. It is seen that 14 out of the 16 values deviate from the expected value with fractions of σ_D.

On the basis of this result we feel motivated not to reject the 0-hypothesis i.e. to consider $h/\bar{h}$, irrespective of the time of the year as samples from an identical distribution $F^*(h^*)$. This gives values of $F^*_{obs}(h^*_i) - F^*_{obs}(h^*_{i-1})$ and $F^*_{obs}(h^*_i)$ as given by the numbers in the first and second columns of Table 5.

Instead of a small number 3–5 of quasi-independent observations for estimating $F(h;\tau)$ at any particular day of the year for the stations FAMITA, UTSIRA, and AMI, we have now got from 1001–1453 values of $h/\bar{h}$ to estimate $F^*(h^*)$. These, however, do not constitute a random sample because of the statistical dependence between observations 24 hours apart. Assuming that the influence of this dependence is effectively lost after periods of approximately the length of a month, we may roughly assume an efficient increase in the size of the sample by a factor of at least twelve.

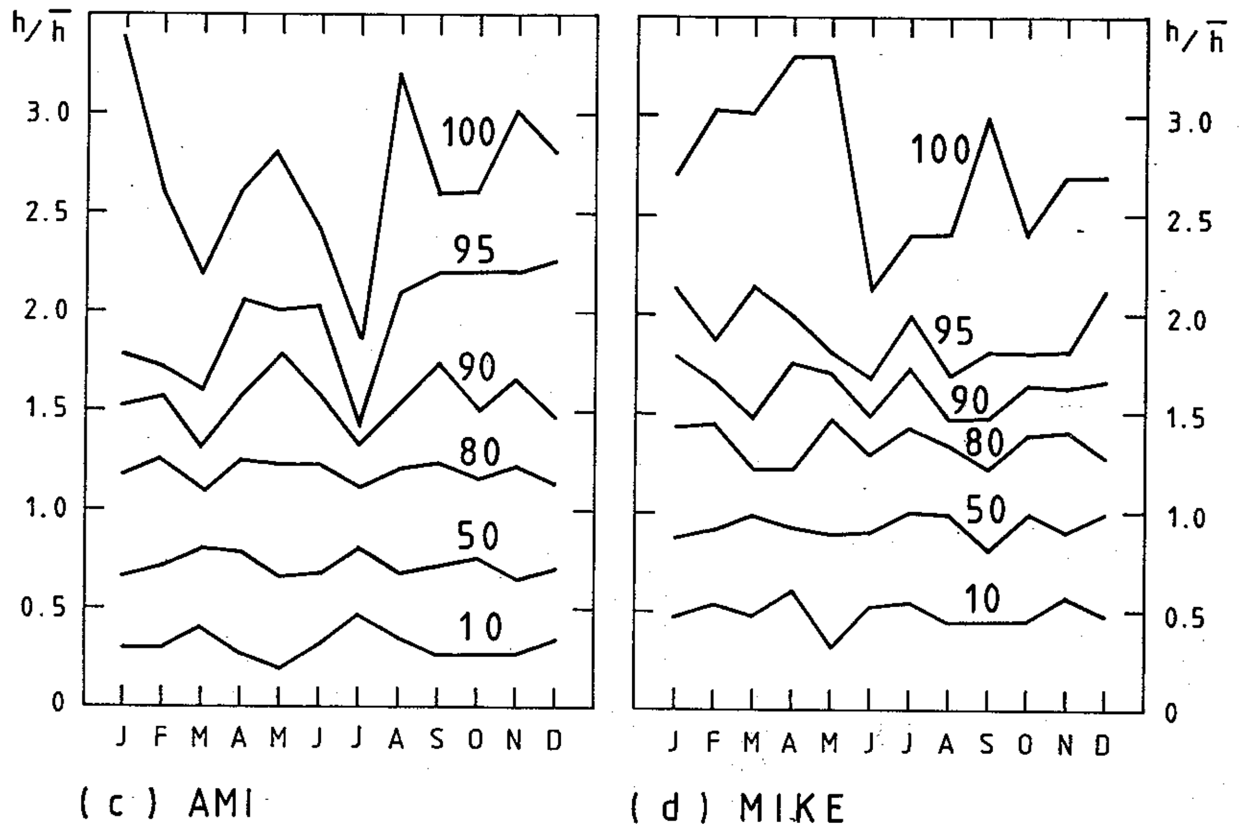


Fig. 1. Selected percentiles of monthly cumulative frequencies of instrument measured significant waveheights h divided by their monthly sample means, $\bar{h}$, at FAMITA, UTSIRA, AMI, and MIKE.

To get estimates of $F(h;\tau)$ at some day τ we use relation (17)

$$F(h;\tau) = F^*(h/\mu(\tau))$$

where $\mu(\tau)$ must be estimated from $\bar{h}(\tau)$. These estimates for mid-month days are given in Table 2 as monthly averages for FAMITA, UTSIRA and AMI, and as the average of three days each month, ten days apart, for MIKE.

1.7 *Fitting a gamma distribution to the observed frequencies*

A variate X has a gamma distribution if

$$\Pr[X \leq x] = \int_0^{x/\beta} \frac{e^{-U}\,U^{\alpha}}{\alpha!}\,dU \equiv G(x;\alpha,\beta) \qquad (20)$$

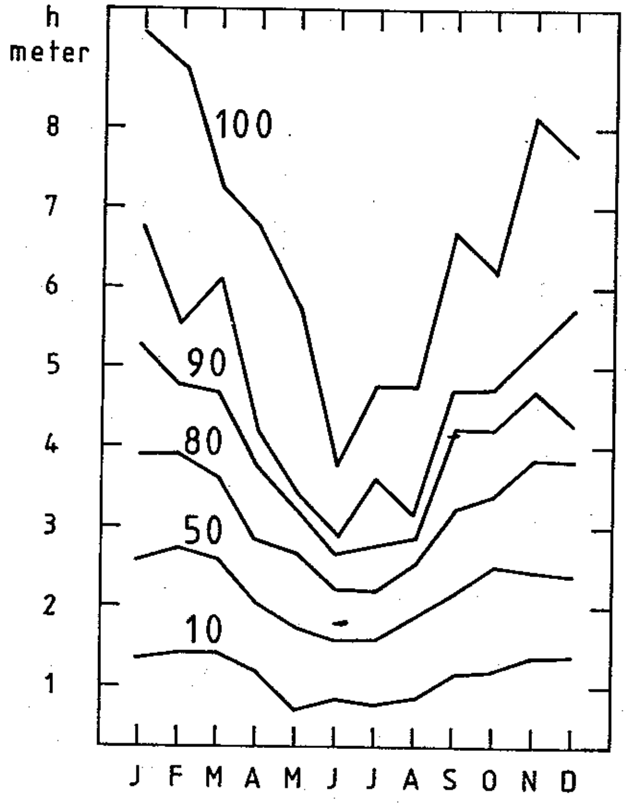


Fig. 2. Selected percentiles of monthly cumulative frequencies of visually estimated significant waveheights at MIKE.

Table 3

MIKE: Ordered samples of 10^2 h/$\bar{h}$ the 11th of each month.

J	F	M	A	M	J	J	A	S	O	N	D
34	26	20	14	20	50	30	30	25	12	28	29
56	34	46	28	33	52	33	44	30	67	33	30
59	41	55	49	34	54	50	72	49	69	51	48
73	64	65	50	48	57	51	72	57	71	55	49
74	65	67	60	50	65	51	73	65	76	59	52
74	66	70	62	52	68	52	74	66	77	63	56
76	70	71	67	54	69	52	78	66	80	68	56
77	73	74	74	75	78	54	78	70	89	73	56
81	78	74	76	83	79	67	83	74	91	73	57
82	87	83	78	90	84	67	93	75	92	75	58
91	88	86	79	101	84	72	94	86	97	77	59
95	89	87	80	102	86	74	98	86	106	78	89
98	93	88	96	104	96	75	100	95	108	84	92
98	95	91	103	107	100	82	104	100	109	84	111
99	98	92	105	109	113	87	106	108	109	102	117
103	108	93	105	114	122	94	107	118	111	107	123
105	117	95	106	122	124	106	111	125	123	139	128
131	129	96	109	135	126	115	117	136	128	144	130
136	141	101	112	137	126	126	117	142	138	153	132
149	142	124	117	140	142	126	122	154	144	175	137
151	149	132	118	144	163	127	124	154	149	176	144
180	155	140	179	163	163	128	134	178	180	189	197
291	169	223	182	164	194	141	158	231	183	206	220
310	238	270	301	202	229	146	160	286	204	214	258

$$0 \leq x, \quad 0 < \beta, \quad -1 < \alpha$$

The first moment μ is given by

$$\mu = \beta(\alpha + 1) \tag{21}$$

First moment of H^* is by definition equal to unity, since $H^* = H/\mu$. If H^* shall satisfy a gamma distribution, we must therefore have

$$\beta = \frac{1}{\alpha + 1} \tag{22}$$

and accordingly

$$G(h^*;\alpha,\beta) = G\left(h^*;\alpha, \frac{1}{\alpha + 1}\right) \equiv F^*(h^*;\alpha) \tag{23}$$

which is a one-parametric family of distributions. We have calculated $F^*(h^*;\alpha)$ with various multiples of 0.5 substituted for α, for the same intervals used in Table 5. The interval frequencies $F^*(h^*_{i+1};\alpha) - F^*(h^*_i;\alpha)$ and the observed ones are plotted in Fig. 3 a–d, together with the values of α which seem to have given the best fit. Similarly Fig. 4 a–d represent the cumulative frequencies.

As it appears, the model looks as a whole quite good. Some of the deviations are certainly erratic. However, for the stations with instrument measured wave-heights there is a discrepancy for small values of h^*, which follows the same picture at all three stations, and therefore most likely is not erratic.

It is possible that this discrepancy is a result of the particular way the instrument wave data are processed for the lowest wave-heights (Personal communication with Lars Håland).

Table 4

In this table the 24 values of $10^2\, h/\overline{h}$ for the 11th of each month for summer and winter months have been grouped together and each of the combined samples of 48 values arranged in increasing order, 0 indicating a summer value and X a winter value. The 48 values of each combined sample are obtained by first running through the first column and then continuing from the top of the second one. The bottom row gives the number of 'runs', as defined in the introduction, of the combined samples.

May	Jun	Jul	Aug	May	Jun	Jul	Aug	May	Jun	Jul	Aug	May	Jun	Jul	Aug
Nov	Nov	Nov	Nov	Des	Des	Des	Des	Jan	Jan	Jan	Jan	Feb	Feb	Feb	Feb
00	X0	XX	X0	00	XX	X0	X0	0X	X0	00	0X	0X	XX	X0	XX
XX	X0	00	00	X0	X0	00	00	00	0X	0X	XX	X0	XX	0X	00
X0	00	XX	X0	X0	X0	XX	X0	X0	0X	X0	0X	00	X0	0X	XX
00	X0	0X	0X	00	XX	0X	00	0X	0X	0X	X0	X0	0X	XX	X0
0X	0X	00	X0	0X	00	X0	X0	00	X0	0X	XX	00	00	X0	00
00	0X	X0	X0	X0	XX	X0	XX	0X	0X	0X	00	XX	0X	0X	X0
00	X0	0X	XX	0X	00	0X	X0	00	XX	0X	0X	00	00	0X	X0
X0	00	00	X0	X0	0X	00	XX	00	00	0X	X0	00	XX	00	XX
00	X0	0X	X0	0X	X0	0X	X0	X0	00	0X	00	0X	00	0X	X0
00	X0	00	00	XX	X0	XX	X0	X0	00	X0	X0	00	X0	00	0X
X0	00	00	00	0X	X0	00	X0	XX	X0	X0	00	XX	X0	0X	00
XX	XX	X0	X0	0X	0X	00	XX	X0	X0	00	X0	X0	00	X0	X0
X0	00	X0	00	X0	XX	00	X0	XX	XX	00	X0	X0	0X	X0	00
XX	0X	X0	X0	XX	XX	XX	0X	00	XX	00	X0	X0	XX	X0	00
X0	XX	0X	0X	X0	XX	X0	0X	X0	X0	X0	0X	XX	X0	00	XX
XX	X0	00	XX	X0	00	XX	0X	X0	0X	0X	00	0X	0X	0X	00
00	X0	XX	XX	XX	0X	XX	00	XX	0X	XX	XX	X0	XX	X0	0X
X0	XX	00	00	X0	00	XX	0X	XX	X0	X0	XX	0X	0X	0X	0X
XX	0X	XX	X0	00	00	X0	0X	00	X0	00	0X	XX	00	XX	XX
XX	XX	XX	0X	00	00	0X	00	00	0X	XX	X0	X0	00	00	XX
0X	00	0X	0X	XX	0X	00	X0	XX	00	XX	00	X0	0X	0X	X0
X0	XX	XX	XX	00	0X	0X	XX	X0	00	XX	0X	0X	X0	XX	00
XX	0X	0X	XX	XX	00	0X	0X	XX	XX	0X	XX	X0	X0	0X	XX
0X	X0	XX	0X	0X	XX	0X	0X	XX	XX	XX	0X	XX	XX	XX	0X
24	28	27	23	28	23	23	19	22	23	20	28	28	25	27	27

1.8 *Smoothing of* $\overline{h}(\tau)$

Let us consider the daily values of expectation $\mu(\tau)$ putting $\tau = 1$ at January 1. $\mu(\tau)$ has an annual periodicity and most probably undergoes at no time of the year great changes within periods of up to a couple of weeks. This may be expressed by assuming a *model*

$$\mu(\tau) = M(\tau) \equiv a_o + \sum_{K=1}^{KK} \vec{a_K} \cdot \vec{F_K}$$

where (24)

$$\vec{F_K} \equiv \left(\cos \frac{2\pi K\tau}{365},\ \sin \frac{2\pi K\tau}{365}\right)$$

in which relatively few terms are kept in this discrete Fourier expansion of μ. – The expansion of the estimates $\overline{h}(\tau)$ may be written

$$\overline{h}(\tau) = M'(\tau;KK) + R(\tau;KK) \tag{25}$$

where

$$M'(\tau;KK) = a'_o + \sum_{K=1}^{KK} \vec{a'_K} \cdot \vec{F_K} \tag{26}$$

and where R contains the components with $K > KK$. The coefficients $\vec{a'_K}$ are estimates of the $\vec{a_K}$ in the model, and are given by

Table 5

Interval frequencies $F^*_{obs}(h^*_i) - F^*_{obs}(h^*_{i-1})$ and cumulative frequencies $F^*_{obs}(h^*_i)$ for interval limits as shown for the variable $h^* = h/\bar{h}$. Observations every 24 hours, whenever they exist, are used for FAMITA, UTSIRA, and AMI, and one observation every 10th day for MIKE. Last row shows the corresponding sample sizes. Unit 1/1000.

i	h^*_i	FAMITA		UTSIRA		MIKE		h^*_i	AMI	
1	0.3	26	26	33	33	36	36	0.2	0	0
2	0.6	193	219	238	271	177	213	0.4	42	42
3	0.9	299	519	241	513	287	500	0.6	142	184
4	1.2	179	698	195	708	242	742	0.8	231	415
5	1.5	146	844	134	841	140	882	1.0	175	589
6	1.8	83	927	69	910	61	943	1.2	146	735
7	2.1	43	970	41	950	27	970	1.4	87	822
8	2.4	13	983	28	978	16	986	1.6	63	885
9	2.7	7	990	10	988	8.1	994	1.8	43	928
10	3.0	5.6	995	3	991	3.5	998	2.0	23	951
11	3.3	1.9	997	3	994	2.3	1000	2.2	10	961
12	3.6	0.9	998	1.4	996			2.4	14	975
13	3.9	0.9	999	2.1	998			2.6	10	985
14	4.2	0.9	1000					2.8	8	993
15	4.5			1.4	999			3.0	4	997
16								3.2	1	998
17	5.1			0.7	1000			3.4	1	999
18	5.4							3.6	1	1000
Number of obs.		1062		1453		864			1001	

$$a'_o = \frac{1}{365}\sum_{\tau=1}^{365}\bar{h}(\tau);$$

$$\vec{a'_K} = \frac{2}{365}\sum_{\tau=1}^{365}\bar{h}\,\vec{F_K} \qquad (27)$$

If, as is the case for MIKE, $\bar{h}(\tau)$ is computed only for 36 days of the year, starting January 1, we get

$$a'_o = \frac{1}{36}\sum_{i=1}^{36}\bar{h}(i) \quad \vec{a'_K} = \frac{2}{36}\sum_{i=1}^{36}\bar{h}\,\vec{F_K} \qquad (28)$$

Averages $\bar{h}$ based upon monthly grouped observations, $\bar{h}(m)$, $m = 1,2,\ldots,12$, when approximated by

$$\bar{h}(m) \approx M'(m;KK) =$$

$$a'_o + \sum_{K=1}^{KK}\vec{A'_K}\cdot\vec{F_K} \qquad (29)$$

$$\vec{F_K} = \left(\cos\frac{K2\pi m}{12},\ \sin\frac{K2\pi m}{12}\right) \qquad (30)$$

gives

$$\vec{A'_K} = \frac{12\sin\dfrac{2\pi K}{12}}{2\pi K}\vec{a'_K} \qquad (31)$$

Putting $m = 1$ at mid-January the estimates $\vec{A'_K}$ are obtained from (32)

$$a'_o = \frac{1}{12}\sum_{m=1}^{12}\bar{h}(m);$$

$$\vec{A'_K} = \frac{2}{12}\sum_{m=1}^{12}\bar{h}(m)\vec{F_K} \qquad (33)$$

Using (31), this gives the estimates of $\vec{a'_K}$:

$$\vec{a'_K} = \frac{2\pi K}{12\sin\dfrac{2\pi K}{12}}\vec{A'_K} \qquad (34)$$

The results obtained for the estimates of the annual constant and the coefficients of the two first components are given in Table 6 in the unit of metres.

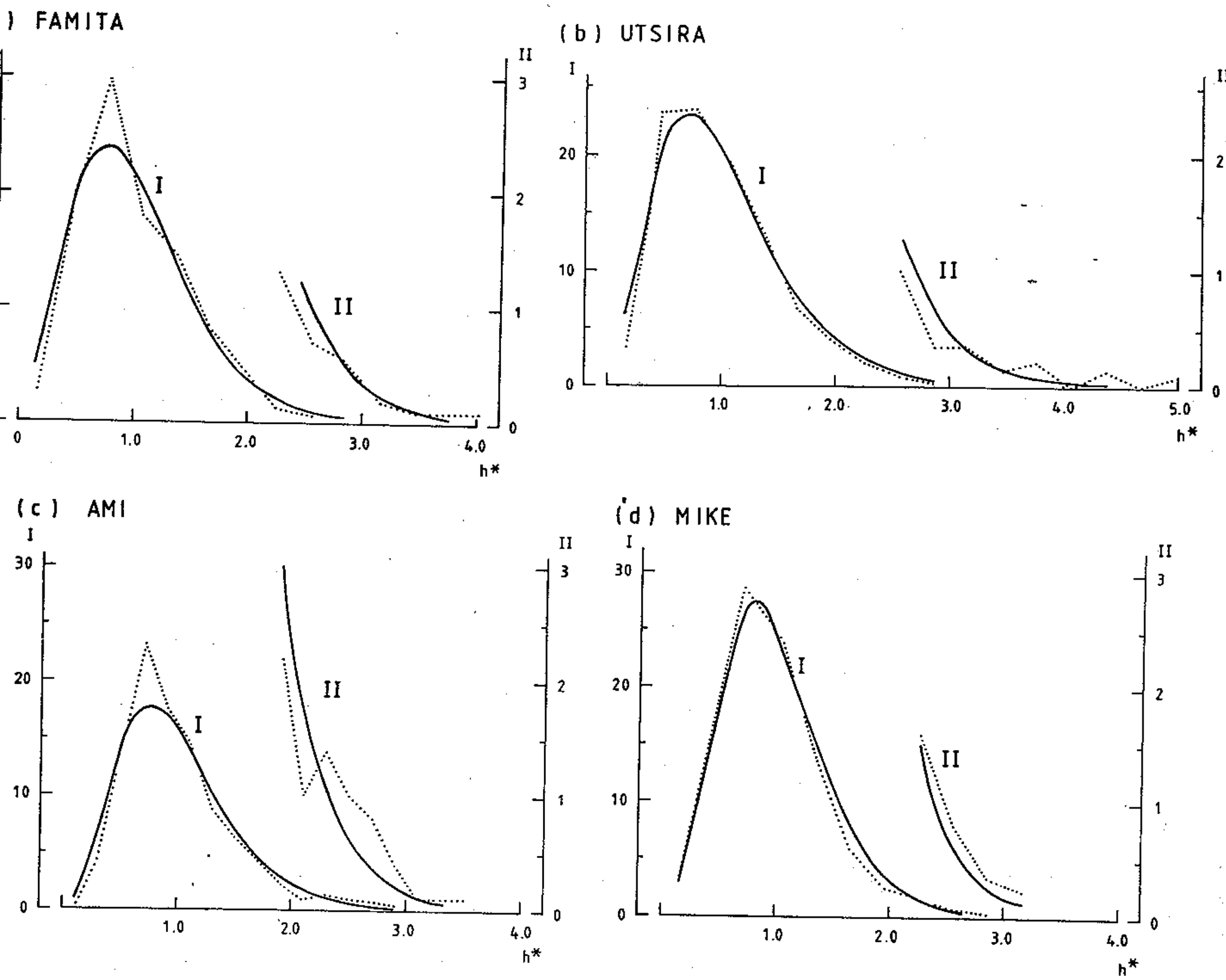


Fig. 3. Points on dotted curves represent the interval frequencies of Table 5. The unit on the vertical axis is 1/100. The curves labelled II represent the upper tails of the distributions. The corresponding points on the full lines are the same interval frequencies as obtained from the Gamma distribution with α taking values 2.5, 2.0, 3.0, and 3.5, respectively.

In order to obtain estimates $\bar{h}(m)$ for FAMITA in the months May–August, we have used the regression

$$\underset{FAM}{\bar{h}(m)} = 1.3\,\underset{UTS}{\bar{h}(m)} \tag{35}$$

obtained from the data of the months September–October. (See Figure 5.)

Substituting from the obtained estimates into (28) and (29) we have got data for the curves drawn in Figure 6, a–d which illustrate the smoothing efficiency of the model representation with the annual and semiannual components. It looks as if this model is quite satisfactory, and it will therefore be adopted in the following. The discussion of the error distribution of the coefficient estimates is difficult, and will be skipped over for the present.

1.9 *The model representation of $F(h;\tau,l)$*

We are now in the position to formulate the model probability distribution

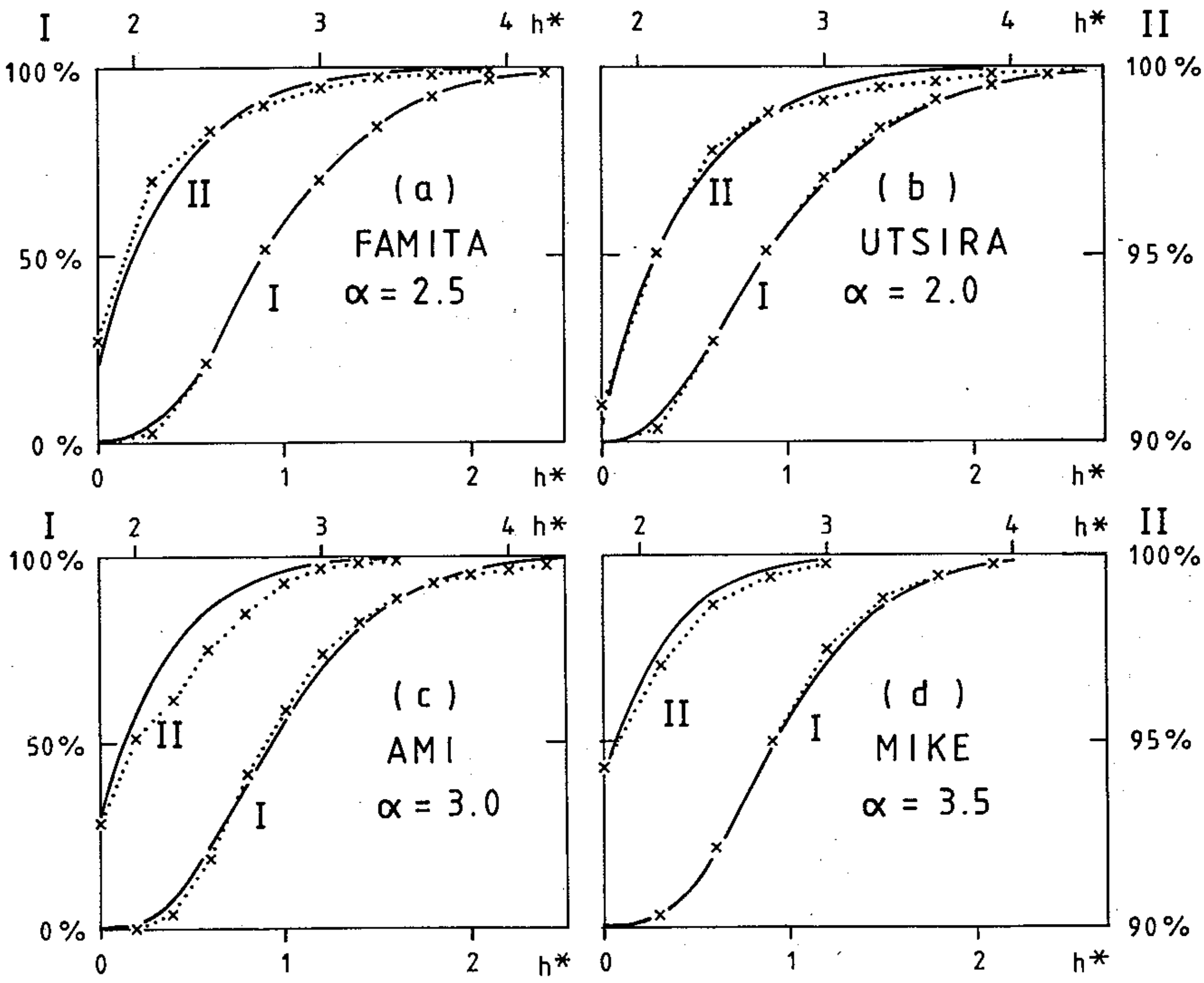


Fig. 4. These curves are the cumulative frequencies of Table 4, corresponding to the curves in Fig. 3.

F(h;τ,l). By account of the results in the two preceding sections, we get

$$\Pr[H \leq h] \equiv F(h;\tau,l) = G\left(\frac{h}{\mu};\alpha,\frac{1}{\alpha+1}\right)$$

$$\mu = a_o + a_1 \cos\frac{2\pi\tau}{365} + b_1 \sin\frac{2\pi\tau}{365} + a_2 \cos\frac{4\pi\tau}{365} + b_2 \sin\frac{4\pi\tau}{365} \quad (36)$$

In (36) α, a_o, a_1, b_1, b_2 are functions of locality, only. For integer values of α, the gamma distribution is

$$G(x;\alpha,\beta) = 1 - e^{-\frac{x}{\beta}}\left(1 + \frac{x}{\beta} + \frac{1}{2}\left(\frac{x}{\beta}\right)^2 + \ldots + \frac{1}{(\alpha-1)!}\left(\frac{x}{\beta}\right)^{\alpha-1}\right) \quad (37)$$

As we have found α to have values near the integers 2, 3, 4, we define functions G_2, G_3, G_4 from

$$G_2(h) = 1 - e^{-(\alpha+1)h/\mu}\left(1 + (\alpha+1)h/\mu + \frac{1}{2}((\alpha+1)h/\mu)^2\right)$$

$$G_3(h) = 1 - e^{-(\alpha+1)h/\mu}\left(1 + (\alpha+1)h/\mu + \frac{1}{2}((\alpha+1)h/\mu)^2 + \frac{1}{6}((\alpha+1)h/\mu)^3\right) \quad (38)$$

$$G_4(h) = 1 - e^{-(\alpha+1)h/\mu}\left(1 + (\alpha+1)h/\mu + \frac{1}{2}((\alpha+1)h/\mu)^2 + \frac{1}{6}((\alpha+1)h/\mu)^3 + \frac{1}{24}((\alpha+1)h/\mu)^4\right)$$

from which we find by quadratic interpolation

$$G\left(h/\mu; \alpha, \frac{1}{\alpha+1}\right) = G_3 + \frac{1}{2}(\alpha-3)(G_4 - G_2) + \frac{1}{2}(\alpha-3)^2(G_2 - 2G_3 + G_4) \quad (39)$$

The algorithms (36), (39) will be used extensively in Part III. They may, for instance, be used for computation on a small desk computer.

1.10 *The frequency distributions of* computed *wave-heights*

Relative interval frequencies of observed significant wave-heights have been compared with those of computed waves in gridpoints nearby, considering all observations at a station as a single group. At MIKE the visually estimated heights of wind sea and swell have been combined in accordance with the formula used to get the significant wave-height from computed wind-sea and swell. The results are shown in Figure 7, a–d.

Table 6

The first moment μ of sign. wave height as a function of the day τ of the year is estimated by

$$\mu(\tau) = a'_0 + a'_1 \sin\frac{2\pi\tau}{365} + b'_1 \cos\frac{2\pi\tau}{365} + a'_2 \sin\frac{4\pi\tau}{365} + b'_2 \cos\frac{4\pi\tau}{365}$$

and a_0, a'_1, a'_2, b'_1, b'_2 are given in the table.

	a'_0	a'_1	a'_2	b'_1	b'_2
FAMITA:	2.26	0.93	0.23	0.33	—0.25
UTSIRA:	1.75	0.57	0.17	0.17	—0.24
AMI:	2.22	0.82	0.04	0.32	—0.11
MIKE:	2.34	0.66	—0.01	0.09	0.07

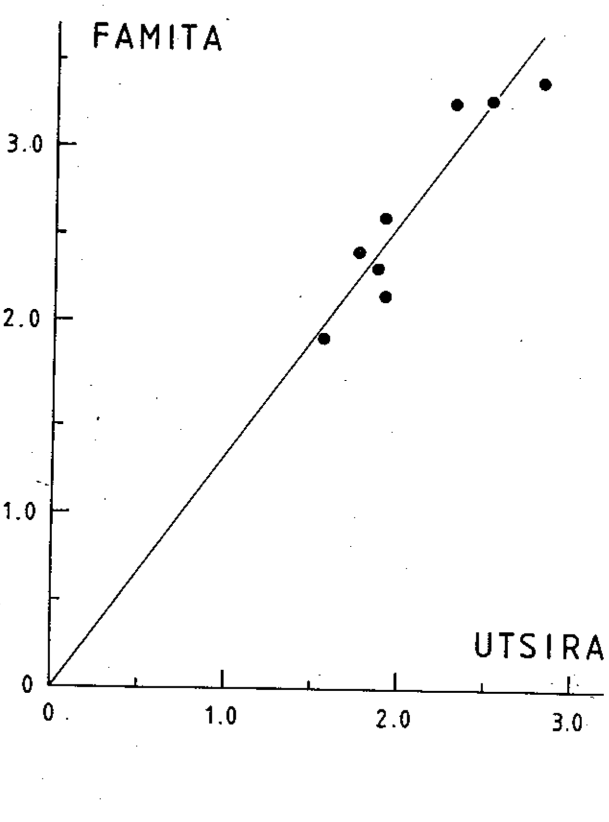
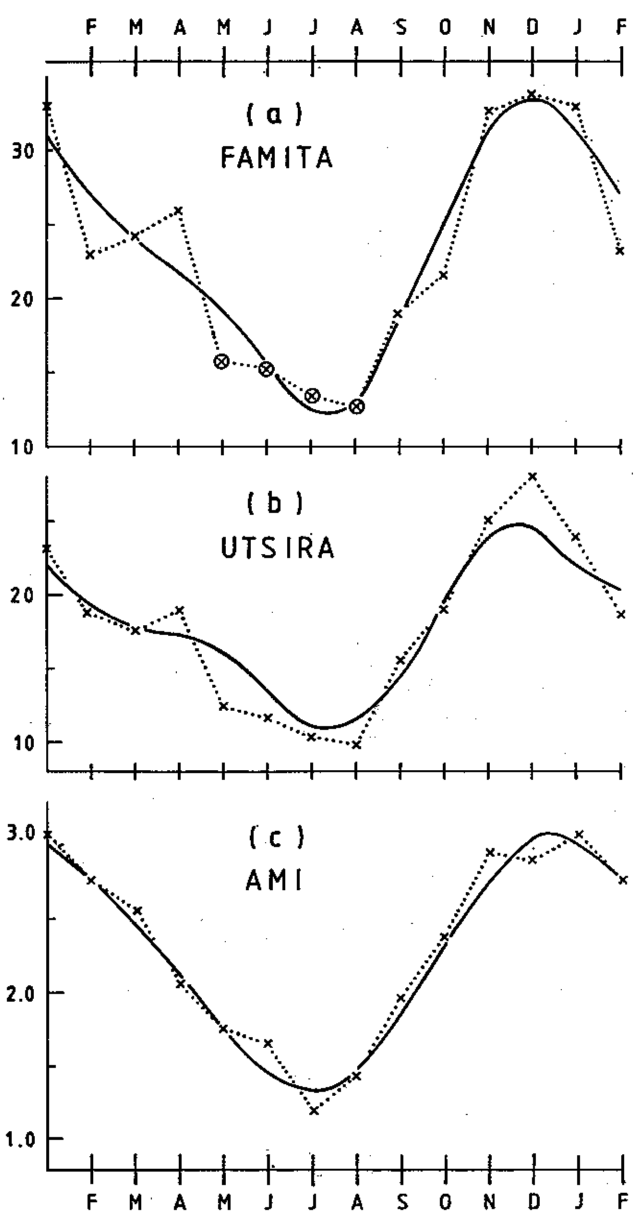


Fig. 5. Simultaneous values in metres of the monthly sample means of significant wave-heights at UTSIRA and FAMITA for the months September through April.



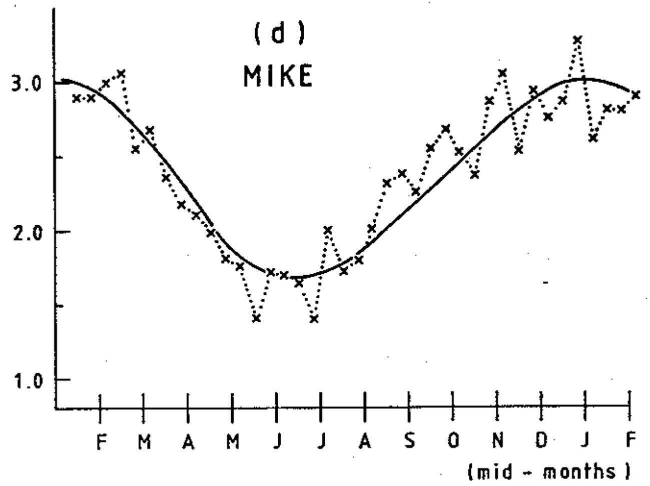


Fig. 6. Estimates of first moments of significant wave-height. Crosses on dotted curves represent estimates from monthly sample means, valid approximately for mid-month days, whereas the full lines represent estimates for arbitrary days of the year obtained from a Fourier development of the mid-month estimates in an annual and semi-annual component in addition to the annual average.

We observe that the frequencies of computed heights at all four stations are markedly too high below approximately 1 m, and markedly too low between 1 m and 4–5 m. Above 4–5 m the frequencies of computed wave-heights are too high at Ekofisk, Statfjord and Tromsøflaket, whereas there are no signs of systematic differences between the two distributions at MIKE/Mike.

We have previously arrived at the result that the cumulative frequency distributions of the observed significant wave-heights at different times of the year differ significantly only by scale factors at the h-axis. This result does not apply to the computed waves. This may for instance be seen by considering the variation of the ratio $\bar{h}/h_{s \cdot d}$ over the year. Except for random variations this should be a constant if the mentioned transformation rule is correct. Figure 8 a–b shows the variation of $\bar{h}$ and $h_{s \cdot d}$ at MIKE for the observed and computed waves obtained from the

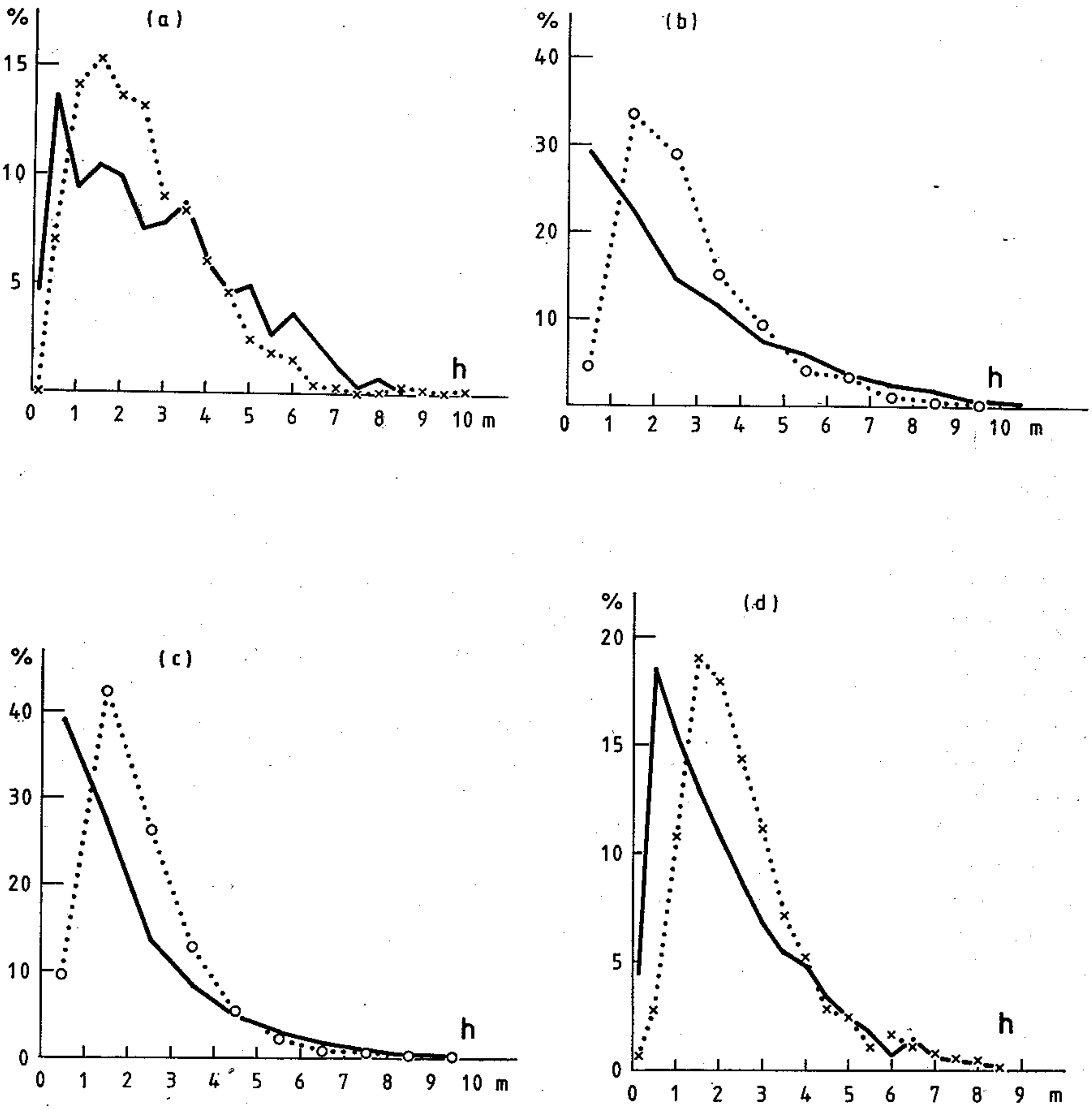


Fig. 7. All-year interval frequencies of significant wave-height for observed (dotted lines) and computed (full lines) at gridpoints nearby. Fig. (a): FAMITA/Ekofisk. (b) BRENT/Statfjord. (c) AMI/Tromsøflaket. (d) MIKE/Mike. Interval length 0.5 m in (a) and (d); 1.0 m in (b) and (c).

observation period October 3, 1972–September 30, 1977, which clearly demonstrates that $\bar{h}/h_{s\cdot d}$ is not a constant over the year for the computed waves, by contrast to what is the case for the observed wave-heights.

This fact, of course, complicates the correction rules we must apply in order to make the statistics of the computed wave heights acceptable.

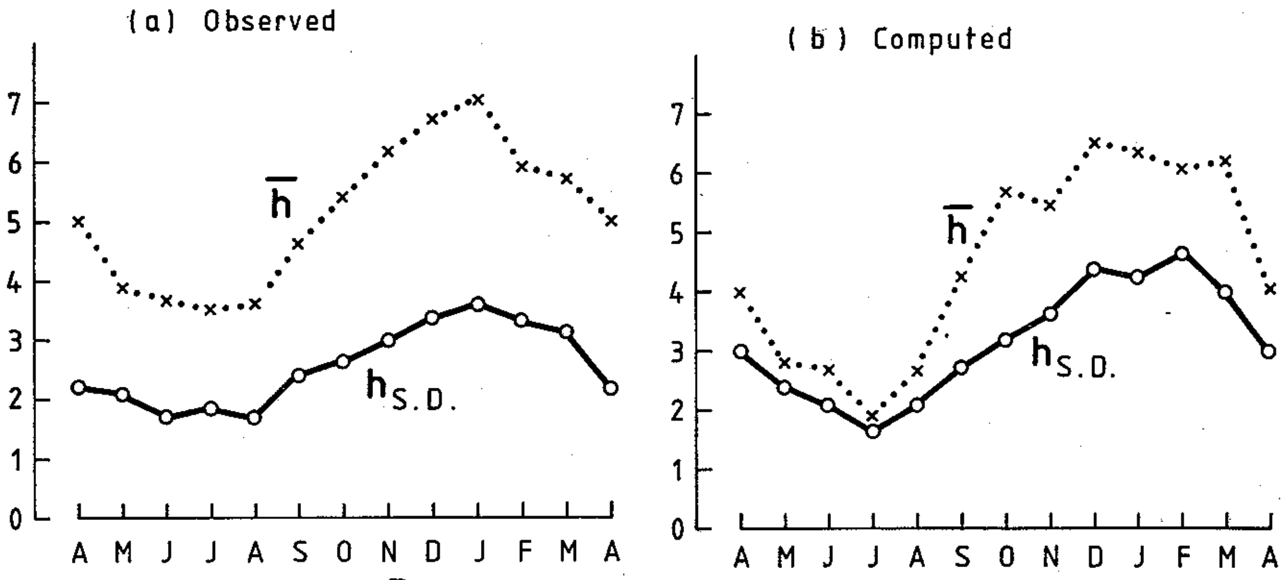


Fig. 8. Monthly sample means $\bar{h}$ and standard deviations $h_{s.d.}$ at MIKE. Fig. 8 (a): Observed waves. Fig. 8 (b): Computed waves nearby.

1.11 *A correction procedure for the frequency distributions of computed significant wave-heights*

Although it may not necessarily be the best correction method, we shall present in this section a short-cut to *one* method for obtaining a correction to the frequency distributions of computed wave-heights. We then at first summarize one result from the above comparisons, namely that the interval frequencies agree in height intervals as given below:

Ekofisk – FAMITA:	3.75 m–4.25 m
Mike – MIKE:	4.25 m–4.75 m
Statfjord – BRENT:	4.50 m–5.50 m
Tromsøflaket – AMI:	4.25 m–5.25 m

To see if this holds throughout the year, we have calculated the differences between interval frequencies of computed and observed significant wave heights at MIKE for each month separately. The results which are given in the table below show no systematic difference over the year in the interval 4.25–4.75 m.

Table 7

Difference between computed and observed significant wave-height interval frequencies at MIKE and nearby gridpoint. Unit 1/1000.

J	F	M	A	M	J	J	A	S	O	N	D	Interval
6	0	13	7	65	100	123	45	40	32	13	6	0.00–0.25 m
6	86	77	200	258	173	303	284	173	150	93	77	0.25–0.75 m
26	107	39	60	—45	93	39	26	—7	124	80	19	0.75–1.25 m
26	—36	—65	20	—45	—147	—168	—213	—107	—46	40	13	1.25–1.75 m
58	—121	—32	—40	—116	—153	—187	—58	13	—137	—80	—6	1.75–2.25 m
—65	—29	—32	—113	—39	—60	—58	—58	—40	—65	—53	—71	2.25–2.75 m
6	—43	—39	—100	—90	7	—26	—39	—73	—72	—53	0	2.75–3.25 m
—71	—36	3	—13	—13	—7	26	6	—20	—40	—7	—39	3.25–3.75 m
19	—43	0	—27	13	—13	—32	6	—7	39	7	—13	3.75–4.25 m
13	14	39	0	19	—7	—19	0	27	7	—7	—26	4.25–4.75 m
—13	36	—19	—13	6	13	0	—6	—13	7	0	6	4.75–5.25 m
—13	14	13	7	0	0	0	6	7	—7	7	19	5.25–5.75 m
—13	—21	—13	—7	—13	0	0	0	—7	—7	—33	—6	5.75–6.25 m
32	7	—13	13	0	0	0	0	0	—13	20	—6	6.25–6.75 m

This result will be applied to the other stations as well for the intervals above.

We shall denote the monthly interval frequencies of the computed wave-heights by $f_C(m)$:

$$f_C(m) = F_{comp}(h' + \Delta h;m) - F_{comp}(h' - \Delta h;m) \qquad (40)$$

The corresponding model frequencies, $f_M(m)$, are

$$f_M(m) = G(h' + \Delta h;\alpha,\beta(m)) - G(h' - \Delta h;\alpha,\beta(m)) \approx 2\,\Delta h\, G' =$$

$$= \frac{2\Delta h}{\beta(m)\alpha!} e^{-\frac{h'}{\beta(m)}} \left(\frac{h'}{\beta(m)}\right)^{\alpha} \qquad (41)$$

Putting $f_C = f_M$ in the interval where they agree gives

$$\frac{2\Delta h}{\beta(m)\alpha!} e^{-\frac{h'}{\beta(m)}} \left(\frac{h'}{\beta(m)}\right)^{\alpha} = f_C(m) \qquad (42)$$

At stations where α is known, or may be found, this equation may be used to find $\beta(m)$ from knowledge of $f_C(m)$.

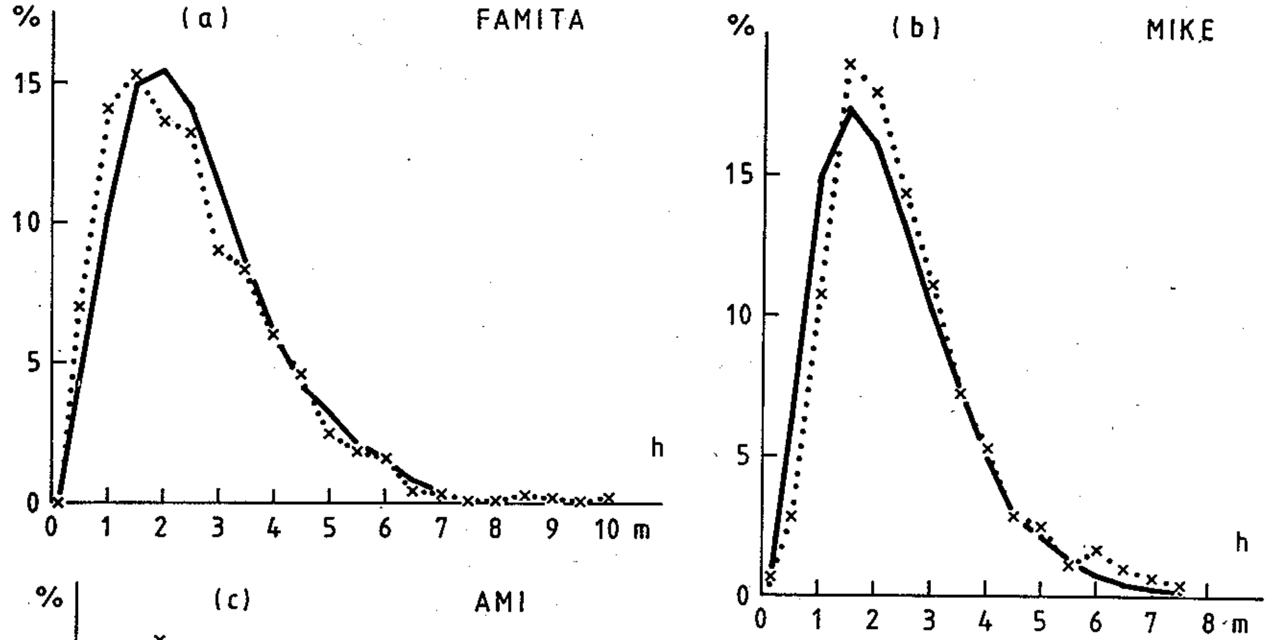
Below in Table 8 we give first the results for Ekofisk, Mike and Tromsøflaket, where previously α has been estimated, based on interval frequencies $f_C(m)$ obtained for the period October 3, 1972–September 30, 1977. We have added a column for $\bar{h}$ derived from $\bar{h} = (\alpha + 1)\beta'$.

Table 8

f_c: Observed monthly percentage frequencies of computed sign. wave heights in intervals $h' - \Delta h, h' + \Delta h$.
β': Corresponding estimates of the parameter β in a Gamma distribution with numerical values of the parameter α as given in the table.
$\bar{h}$: Corresponding estimates in metres of the first moments using the formula $\mu = (\alpha + 1)\beta$.

	EKOFISK $\alpha = 2.5$ $h' = 4.0$ m $\Delta h = 0.25$ m			MIKE $\alpha = 3.5$ $h' = 4.5$ m $\Delta h = 0.25$ m			TROMSØFLAKET $\alpha = 3.0$ $h' = 4.75$ m $\Delta h = 0.50$ m		
	f_C	β'	$\bar{h}$	f_C	β'	$\bar{h}$	f_C	β'	$\bar{h}$
J	7.90	0.87	3.05	5.60	0.645	2.90	6.37	0.64	2.56
F	5.10	0.675	2.36	5.80	0.655	2.95	6.96	0.66	2.64
M	5.80	0.72	2.52	5.70	0.650	2.93	6.70	0.65	2.60
A	5.10	0.675	2.36	3.05	0.535	2.41	3.33	0.54	2.16
M	2.40	0.54	1.89	2.70	0.520	2.34	0.48	0.40	1.60
J	1.85	0.43	1.51	0.60	0.300	1.35	0.17	0.36	1.44
J	0.95	0.44	1.54	0.30	0.250	1.13	0.73	0.42	1.68
A	1.05	0.45	1.58	0.65	0.310	1.40	0.89	0.435	1.74
S	4.00	0.62	2.17	2.65	0.515	2.32	3.67	0.56	2.24
O	4.15	0.625	2.19	5.01	0.620	2.79	4.66	0.59	2.36
N	5.90	0.725	2.54	5.25	0.630	2.84	9.25	0.73	2.92
D	8.55	0.95	3.33	4.45	0.595	2.68	8.39	0.70	2.80
	%		m	%		m	%		m

With the monthly values of β thus found together with the used values of α, we can, by substitution into the gamma distribution $G(h;\alpha,\beta(m))$, find the monthly corrected frequencies within arbitrary limits of h for the computed wave-heights. After this has been accomplished, these monthly frequencies are added and averaged for October–March at Ekofisk, and for January–December for Mike and



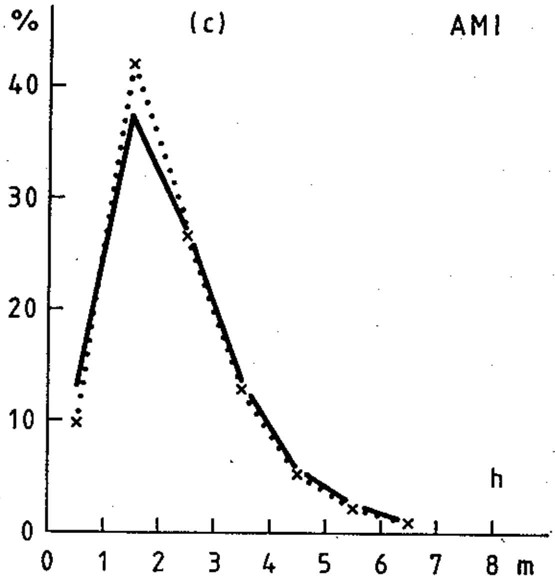


Fig. 9. This figure corresponds to Fig. 7 (a), (c), and (d) after the frequencies for the computed waves have been corrected.

Tromsøflaket. The results are illustrated in Figure 9 a–c by the heavy drawn lines.

By contrast to the non-corrected frequencies of the computed waves, which were illustrated in Figure 7, we see that the corrected frequencies now compare well with those of the observed waves, notwithstanding the difference in periods at FAMITA and AMI for the computed and observed waves.

At Statfjord the parameter α has not previously been estimated. Instead we have adopted the value α which, after the $\beta(m)$'s have been determined from the computed wave-frequencies in the interval 4.5–5.5 m obtained from the period October 3, 1972–September 30, 1977, gives the best agreement with the instrument-measured frequency distribution in the period December 1, 1975 to December 31, 1977, when averaged over this 25-month period. This has resulted in the approximate value $\alpha = 3$.

The Table for f_C, β' and $\bar{h}$ is as follows:

STATFJORD

$\alpha = 3$; $h' = 5$ m; $\Delta h = 0.5$ m

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
f_C	7.25	6.05	3.95	2.50	0.75	0.50	0.95	0.95	3.65	3.45	3.65	5.55
β'	1.04	0.89	0.74	0.64	0.50	0.47	0.53	0.53	0.72	0.70	0.72	0.85
$\bar{h}$	4.16	3.56	2.96	2.56	2.00	1.88	2.12	2.12	2.88	2.80	2.88	3.40

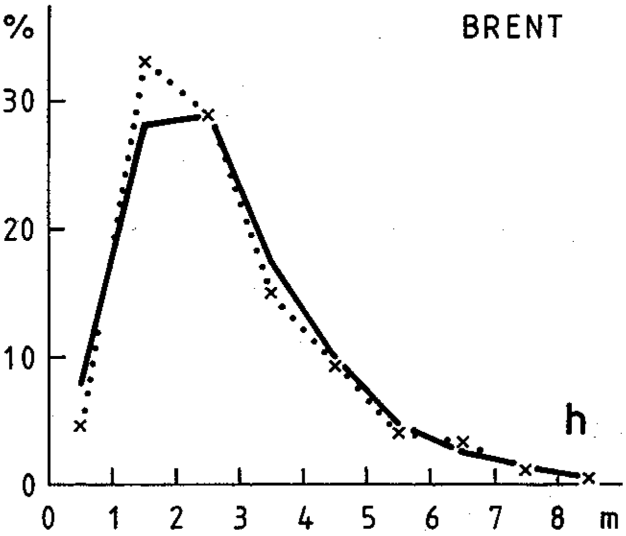


Fig. 10. Figure 7 (b) after correction of the frequencies of the computed waves at Statfjord.

Figure 10 illustrates the comparison between the corrected frequency distribution of computed waves and that of instrument measured waves for the periods mentioned above at BRENT and the nearby gridpoint Statfjord.

PART II

2.1 Markov *models for successive events*

A: Let $T_1, T_2, \ldots, T_M, \ldots, T_{MM}$ denote a sequence of time sections of which we assume

$$(t \in T_M) > (t \in T_{M-1}) \qquad (1)$$

Starting year of T_1 is arbitrary, starting day τ is fixed. (2)

B: Suppose rules have been decided upon according to which the function values of a time-dependent variable V, whatever they should be, within each time section may be classified to belong to one and only one of II categories. For convenience the number II of categories is considered constant.

C: Suppose that the categories within each time section are arranged in a certain order, labelled by the integers

$$x \in X \equiv (1,2,\ldots I,\ldots II) \qquad (3)$$

To each single result of a sequence of MM categories there corresponds then an MM-dimensional number $\vec{x} = (x_1,x_2,\ldots x_M, \ldots x_{MM})$ and vice versa. For N years there correspond N results, from which we may find the observed relative frequencies f_{obs} defined by

$$f_{obs}(\vec{x}) = \frac{n(\vec{x})}{N} \qquad (4)$$

where $n(\vec{x})$ is the number of times a given result $\vec{x}$ has occurred in the sample of size N.

Altogether there are II^{MM} different possible results. This is our sample space S:

$$S \equiv X^{MM} = \{\vec{x} \,|\, x_M \in X\} \qquad (5)$$
$$M = 1,2,\ldots, MM$$

We suppose now that to any event A belonging to S there is connected a probability Pr(A). When dealing with particular events of interest, they will often be defined by a criterion c for selecting those results $\vec{x}$ belonging to A. We may write this symbolically by

$$A = A(c) = \{\vec{x}|c\} \qquad (6)$$

or

$$A = A(c) = \bigcup_{\vec{x}|c} \{\vec{x}\} \qquad (7)$$

where the last expression is the *union* of all single results $\vec{x}$ which satisfy the selection criterion. If, as a particular case, the criterion is

c: x_M has specified arbitrary value x, the other components of $\vec{x}$ being unspecified

we write the corresponding A(c):

$A(c) = X_M$, or $A(c) = (x_M = x)$

If x_M and $x_{M'}$ both have specified values, the others not, then A(c) is the *intersection* of X_M and $X_{M'}$ which shall be written as

$$A(c) = X_M X_{M'} \text{ or}$$
$$A(c) = (x_M = x)(x_{M'} = x')$$

which generalizes to an arbitrary number of components having specified values.

A criterion c of special interest is a condition on all or some of the first M components. In this case we shall write:

$$A(c) = A(c,M) \equiv \{\vec{x} \mid \text{some or all of the first M components satisfy a given condition}\}$$

Let us now consider an event which is the intersection of A(c,M), X_{M+1} and X_{M+2}. The probability of this event may then be written as:

$$Pr[A(c,M)X_{M+1}X_{M+2}] = Pr[A(c,M)X_{M+1}]Pr(X_{M+2}|A(c,M)X_{M+1}) \quad (8)$$

where the last factor is the *conditional* probability that X_{M+2} has a specified value when the event $A(c,M)X_{M+1}$ is given.

The *first order* Markov model may now be formulated as:

$$Pr(X_{M+2}|A(c,M)X_{M+1}) = Pr(X_{M+2}|X_{M+1}) \quad (9)$$

Using this in (8), we obtain

$$Pr[A(c,M)X_{M+1}X_{M+2}] = Pr[A(c,M)X_{M+1})]Pr(X_{M+2}|X_{M+1}) \quad (10)$$

In the *second order* Markov model the equations corresponding to (9), (10) are:

$$Pr(X_{M+3}|A(c,M)X_{M+1}X_{M+2}) = Pr(X_{M+3}|X_{M+1}X_{M+2}) \quad (11)$$

$$Pr[A(c,M)X_{M+1}X_{M+2}X_{M+3}] = Pr[A(c,M)X_{M+1}X_{M+2}]Pr(X_{M+3}|X_{M+1}X_{M+2}) \quad (12)$$

2.2 *Multinomial distributions in Markov models*

Consider now a condition c on the first M components given by

$$\begin{aligned} c: x &= 1 && \text{occurs } r_1 \text{ times} \\ x &= 2 && \text{occurs } r_2 \text{ times} \\ &\text{-----------} && \\ x &= II - 1 && \text{occurs } r_{II-1} \text{ times} \end{aligned} \quad (13)$$

Obviously, we must have

$$r_1, r_2, \ldots, r_{II-1} \leq M;$$
$$r_1 + r_2 + \ldots, + r_{II-1} \leq M \quad (14)$$

We may symbolize this condition by

$$c: (r_1, r_2, \ldots, r_{II-1}, M) \quad (15)$$

An event A can always be written:

$$A = AS \quad (16)$$

Writing S as the union

$$S = \bigcup_{I=1}^{II} (x_M = I)$$

and substituting this together with the A fulfilling condition (15), we obtain

$$A(r_1, r_2, \ldots, r_{II-1}, M) = \bigcup_{I=1}^{II} A(r_1, r_2, \ldots, r_{II-1}, M)(x_M = I) \quad (17)$$

Now the two conditions

$$c: (r_1, r_2, \ldots, r_J, \ldots, r_{II-1}, M + 1) \text{ and } x_{M+1} = J$$
$$c: (r_1, r_2, \ldots, r_J - 1, \ldots, r_{II-1}, M) \text{ and } x_{M+1} = J$$

are equivalent conditions and define therefore identical events:

$$A(r_1,r_2,\ldots, r_{II-1},M + 1)(x_{M+1} = J) = A(r_1,r_2,\ldots, r_J - 1,\ldots, r_{II-1},M)(x_{M+1} = J) \tag{18}$$

Substituting here from (17) we obtain

$$A(r_1,r_2,\ldots, r_{II-1},M + 1)(x_{M+1} = J) = \bigcup_{I=1}^{II} A(r_1,\ldots, r_J - 1,\ldots, r_{II-1},M)(x_M = I)(x_{M+1} = J) \tag{19}$$

Here the events on the right-hand side are disjoint. Using the result (10) valid in a first-order Markov model, we therefore get for the corresponding probabilities

$$\Pr[A(r_1,r_2,\ldots, r_{II-1},M + 1)(x_{M+1} = J)] = \sum_{I=1}^{II} \Pr[A(r_1,\ldots, r_J - 1,\ldots r_{II-1},M)(x_M = I)]\Pr(x_{M+1} = J|x_M = I) \tag{20}$$

and correspondingly from (17)

$$\Pr[A(r_1,r_2,\ldots r_{II-1},M + 1)] = \sum_{J=1}^{II} \Pr[A(r_1,r_2,\ldots r_{II-1},M + 1)(x_{M+1} = J)] \tag{21}$$

(20), (21) is a recurrence scheme to obtain the multinomial probability distribution in a first order Markov model.

We introduce functions defined by

$$F(r_1,r_2,\ldots, r_{II-1},M,I) \equiv \Pr[A(r_1,r_2,\ldots, r_{II-1},M)(x_M = I)]$$
$$G(r_1,r_2,\ldots, r_{II-1},M) \equiv \Pr[A(r_1,r_2,\ldots, r_{II-1},M)]$$
$$PP(I,J,M) \equiv \Pr(x_{M+1} = J|x_M = I)$$

Then (20), (21) become

$$F(r_1,r_2,\ldots, r_{II-1},M + 1, J) = \sum_{I=1}^{II} F(r_1,\ldots r_J - 1,\ldots, r_{II-1},M,I)PP(I,J,M)$$
$$G(r_1,r_2,\ldots,r_{II-1},M) = \sum_{J=1}^{II} F(r_1,r_2,\ldots, r_{II-1},M,J) \tag{22}$$

As input we need PP(I,J,M) and the initial field of F. If P(I,1) is defined as

$$P(I,1) = \Pr(x_1 = I) \tag{23}$$

we easily find that the initial field F is determined from

$$F = 0 \text{ for } M = 1 \text{ except that}$$
$$F(0,\ldots,r_J = 1,\ldots 0,M = 1,J) = P(J,1)$$
$$J = 1,2,\ldots,II - 1 \tag{24}$$
$$F(0,0,\ldots,0, M = 1, J = II) = P(II,1)$$

If we use a *second-order* Markov model we get, instead of (22), (23), the following recurrence scheme for deriving the multinomial probability distribution:

$$F(r_1,r_2,\ldots,r_{II-1},M+1,J,K) =$$

$$\sum_{I=1}^{II} F(r_1,r_2,\ldots,r_K-1,\ldots,M,I,J)PPP(I,J,K,M-1) \qquad (25)$$

$$G(r_1,r_2,\ldots,r_{II-1},M) = \sum_{J,K=1}^{II} F(r_1,r_2,\ldots,r_{II-1},M,J,K)$$

where

$$PPP(I,J,K,M) \equiv \Pr[x_{M+2}=K|(x_M=I)(x_{M+1}=J)]$$

In addition we need the initial field for F, i.e. for $M = 2$. The scheme below for the case where $II = 3$ will demonstrate how to construct the initial field in the general case.

$II = 3;\ M = 2$ (26)

I	J	r_1	r_2	F
1	1	2	0	$P(1,M=1)PP(1,1,M=1)$
1	2	1	1	$P(1,M=1)PP(1,2,M=1)$
1	3	1	0	$P(1,M=1)PP(1,3,M=1)$
2	1	1	1	$P(2,M=1)PP(2,1,M=1)$
2	2	0	2	$P(2,M=1)PP(2,2,M=1)$
2	3	0	1	$P(2,M=1)PP(2,3,M=1)$
3	1	1	0	$P(3,M=1)PP(3,1,M=1)$
3	2	0	1	$P(3,M=1)PP(3,2,M=1)$
3	3	0	0	$P(3,M=1)PP(3,3,M=1)$

Otherwise $F = 0$.

2.3 *Probability distributions of sums in first-order Markov models*

Let now

$$c: (q,M) \qquad (27)$$

denote the condition

$$x_1 + x_2 + \ldots + x_M = q \qquad (28)$$

Obviously

$$q \in Q \equiv \{M, M+1, \ldots, II \cdot M\} \qquad (29)$$

As in the previous section we can, corresponding to equation (17), write:

$$A(q,M) = \bigcup_{I=1}^{II} A(q,M)(x_M = I) \qquad (30)$$

Furthermore

$c: (q, M+1)$ and $(x_{M+1} = J)$
$c: (q-J,M)$ and $(x_{M+1} = J)$

are identical conditions defining the identical events

$$A(q,M+1)(x_{M+1}=J) = A(q-J,M)(x_{M+1}=J) \qquad (31)$$

Substituting here on the right-hand side from (30), we obtain

$$A(q,M+1)(x_{M+1}=J) = \bigcup_{I=1}^{II} A(q-J,M)(x_M=I)(x_{M+1}=J) \qquad (32)$$

Noting that the events on the right-hand side of (32), (30) are disjoint, we obtain the equations for the corresponding probabilities, using the function notation for the probabilities:

$$F(q,M+1,J) = \sum_{I=1}^{II} F(q-J,M,I)PP(I,J,M) \qquad (33)$$

$$G(q,M+1) = \sum_{J=1}^{II} F(q,M+1,J)$$

with the initial field of F given by

$$F = 0 \text{ except that}$$
$$F(J,M=1,J) = P(J,1) \qquad (34)$$

2.4 *Probability distributions of extremes in first-order Markov models*

Let us consider the event V_{Max} defined by

$$V_{Max} = \text{Maximum } V \text{ at } M \text{ time points } t_1, t_2, \ldots, t_M \qquad (35)$$

Then

$$Pr[V_{Max} \le v] = Pr[(V(t_1) \le v)(V(t_2) \le v) \ldots (V(t_M) \le v)] \qquad (36)$$

With the labelling

$$x = 1 \longleftrightarrow V(t_{M'}) \le v$$
$$M' = 1,2,\ldots,M$$
$$x = 2 \longleftrightarrow V(t_{M'}) > v$$

(36) may be written

$$Pr[V_{Max} \le v] = Pr[(x_1 = 1)(x_2 = 1) \ldots (x_M = 1)] \qquad (37)$$

According to (10) in a *first-order* Markov model (37) may be written

$$Pr[V_{Max} \le v] = Pr(x_1 = 1)(x_2 = 1) \ldots (x_{M-1} = 1)] \, Pr(x_M = 1 | x_{M-1} = 1) \qquad (38)$$

and by induction

$$Pr[V_{Max} \le v] = Pr(x_1 = 1) \prod_{M'=1}^{M-1} Pr(x_{M'+1} = 1 | x_{M'} = 1) \qquad (39)$$

With $F_x(v,M)$ defined by

$$F_x(v,M) = Pr[V_{Max} \le v]$$

and the earlier notations for $Pr(x_1 = 1)$ and $Pr(x_{M'+1} = 1 | x_{M'} = 1)$ equation (36) may be written

$$F_x(v,M) = P(1,1) \prod_{M'=1}^{M-1} PP(1,1,M') \qquad (40)$$

2.5 *Probability distributions of duration of 'runs'*

In this case the chosen number of categories is two, labelled by $x = 1$ and $x = 2$. A 'run' of length D of category 1 which starts at t_1 is defined as the intersection of events

$$A(D) = (x_o = 2)(x_1 = 1) \ldots (x_D = 1)(x_{D+1} = 2) \text{ given that } x_o = 2 \text{ and } x_1 = 1.$$

Its probability is given by the conditional probability

$$Pr[A(D)] = \frac{Pr[(x_o = 2)(x_1 = 1) \cdots (x_D = 1)(x_{D+1} = 2)]}{Pr[(x_o = 2)(x_1 = 1)]}$$

Using here the general multiplication law in the numerator and denominator, we get

$$\Pr[A(D)] = \frac{\Pr(x_o = 2) \cdot \Pr(x_1 = 1|x_o = 2) \ldots \Pr[x_{D+1} = 2|(x_o = 2)(x_1 = 1) \ldots (x_D = 1)]}{\Pr(x_o = 2) \cdot \Pr(x_1 = 1|x_o = 2)} \quad (41)$$

In the *first-order* Markov model this simplifies to

$$\Pr[A(D)] = \left(\prod_{M=1}^{D-1} \Pr(x_{M+1} = 1|x_M = 1)\right) \Pr(x_{D+1} = 2|x_o = 1)$$

In function notation this becomes:

$$F(D) = \left(\prod_{M=1}^{D-1} PP(1,1,M)\right) PP(1,2,D) \quad (42)$$

In the *second-order* Markov model we get instead

$$F(D) = PP(2,1,1,M = 0) \prod_{M=1}^{D-2} PP(1,1,1,M)\ PP(1,1,2,D-1) \quad (43)$$

For the *expectation* of the variate D we always get

$$E(D) = \sum_{D=1}^{\infty} D \Pr[A(D)] = \frac{1}{PP(1,2)} \quad (44)$$

when the time series is *stationary*. This important rule does not use any Markov model concept, but only that the series is stationary. For its proof we note first that because of the stationarity we may now consider a sample of size N as the results at N successive time points $t_1, t_2, \ldots\ldots, t_M, \ldots t_N$. We now let

$n(1,2)$ = number of events $(x_M = 1)(x_{M+1} = 2)$ = number of runs
$n(1)$ = number of results $x = 1$
$n(D)$ = number of runs of result $x = 1$ of length D in the sample of size N.

Obviously

$$n(1) = \sum_{D=1}^{\infty} Dn(D)$$

Dividing here on both sides with $n(1,2)$ we get

$$\frac{\sum_{D=1}^{\infty} Dn(D)}{n(1,2)} = \frac{1}{\frac{n(1,2)}{n(1)}}$$

which are estimates of $E(D)$ and $\frac{1}{PP(1,2)}$.

Letting n increase indefinitely, these estimates will approach the underlying probabilities, resulting in (44).

PART III

3. *Applications to the study of the wave-climate.*

3.1 *Probability distributions that $H \leq h$ r times at M successive equidistant time points*

In this case the time sections are the time points $t_1, t_2, \ldots, t_M, \ldots, t_{MM}$ with $t_M - t_{M-1} = \Delta t$. The categories may be arranged as

$x = 1$: $H \leq h$; $x = 2$: $H > h$

at any time t_M, where h is an arbitrary fixed wave-height. The probability distribution we are seeking is the multinomial distribution in 2.2 with II = 2. In a second order Markov model this distribution may be found from the recurrence scheme in (25). The output of this scheme is a function G(r,m) defined by

G(r,M) = Pr[H $\leq$ h r times at M successive, equidistant time points]

A more flexible function for the users is the corresponding cumulative distribution GC(r,M) defined by

GC(r,M) = Pr[H $\leq$ h at most r times at M successive, equidistant time points]

We note that

1 − GC(r,M) = Pr[H $\leq$ h *at least* r + 1 times at M successive time points]

and

GC(r,M) = Pr[$H > h$ at least M–r times at M successive time points]

In applying this recurrence scheme to wave-height data, we have so far only considered the case with

Δt = 24 hrs.; MM = 30;
t_1 = first day of a month

Then we may quite safely consider P(I,M), PP(I,J,M), and PPP(I,J,K,M) as constants equal to their mid-month values, which may be estimated from calendar month groups of observations. The corresponding recurrence scheme has been programmed for the NORDIC computing system at the Norwegian Meteorological Institute.

It has been applied to visually estimated wave-heights at MIKE using observations for the period 1954–1972, and for computed heights at five different locations using computations for the period October 3, 1972–September 30, 1977. The results are listed for selected h-values for M = 3,6, 9,12,18,21,24,27,30 days, and available from the Meteorological Institute. The results for MIKE are reported (R. Fjørtoft 1977).

Because of the general interest of this programme, we shall below write it down in a very schematic form, so that it may be easily reprogrammed if so desired. We shall make changes in the notations as follows:

For I,J,K = 1,2 we write

P(I,M) = PI
PP(I,J,M) = PIJ
PPP I,J,K,M = PIJK
F(r,M,I,J) = AIJ(r)
F(r,M + 1,I,J) = BIJ(r)

The recurrence scheme may be written:

For r = −1,0,...,MM, make AIJ, BIJ=0

Then make:

A11(2) = P1 P11; A12(1) = P1 P12
A21(1) = P2 P21; A22(0) = P2 P22

Then for

M = 3,...,MM
r = 0,...,M
make

B11(r) = P111 A11(r − 1) + P211 A21(r − 1)
B12(r) = P112 A11(r) + P212 A21(r)
B21(r) = P121 A12(r − 1) + P221 A22(r − 1)
B22(r) = P122 A12(r) + P222 A22(r)

$$A11(r) = B11(r); A12(r) = B12(r)$$
$$A21(r) = B21(r); A22(r) = B22(r)$$

$$G(r,M) = A11(r) + A12(r) + A21(r) + A22(r)$$

$$GC(r,M) = \sum_{r'=0}^{r} G(r', M)$$

Write GC(r,M) for selected values of M.

3.2 *An alternative method for deriving the probability distribution of the previous section*

The genuine Markov probability model concept applies the same basic ideas to the *probabilities* of the future possible states of a well-defined physical system, as do the physical laws to the states themselves, in that the future development is uniquely determined when, in addition to the probability – respectively the physical laws, the state is given initially. The initial state in a single location, however, does of course not define a physical system. If, however, the past local history back to some time is known, this somewhat reflects the state at some earlier time of a *volume* of the atmosphere-ocean system. Markov models, given such local histories, may therefore be a little more soundly founded than are models in which only the local initial states are taken as known. However, great difficulties are encountered if one endeavours to describe the history in too great details. The example treated below illustrates how one may proceed to overcome some of these difficulties.

We consider now time sections defined by MM successive collections of N time points:

$$T_1 = \{t_1, \ldots, t_N\}, \ldots, T_M, \ldots, T_{MM} = \{t_{(MM-1)N+1}, \ldots, t_{MM \cdot N}\}$$

Within each T the number s of times H is less or equal to a given height h, ranges from s = 0 to s = N.

Taking these as our II categories, labelled by the integers $X = \{1,2,\ldots,I,\ldots,II\}$ such that $s = I - 1$, we obtain that

$$q = x_1 + x_2 + \ldots + x_M = r + M$$

where r is the number of times H is less or equal to h at $M \cdot N$ time points. Accordingly

$$\Pr[H \leq h \ q - M \text{ times at MN time points}] = \Pr[x_1 + x_2 + \ldots + x_M = q] \qquad (1)$$

For the probabilities on the right-hand side of (1) we have derived the recurrence scheme (33), (34) in Part II applicable in a first order Markov model. Using this in the present case, the probability functions P(I,1) and PP(I,J,M) in the recurrence scheme are defined by:

$$P(I,1) = \Pr(x_1 = I) = \Pr[H \leq h \ I - 1 \text{ times in the first N time points}]$$
$$P(I,J,M) = \Pr(x_{M+1} = J | x_M = I) =$$
$$\Pr(H \leq h \ J - 1 \text{ times within } T_{M+1} \text{ when } H \leq h \ I - 1 \text{ times within } T_M)$$

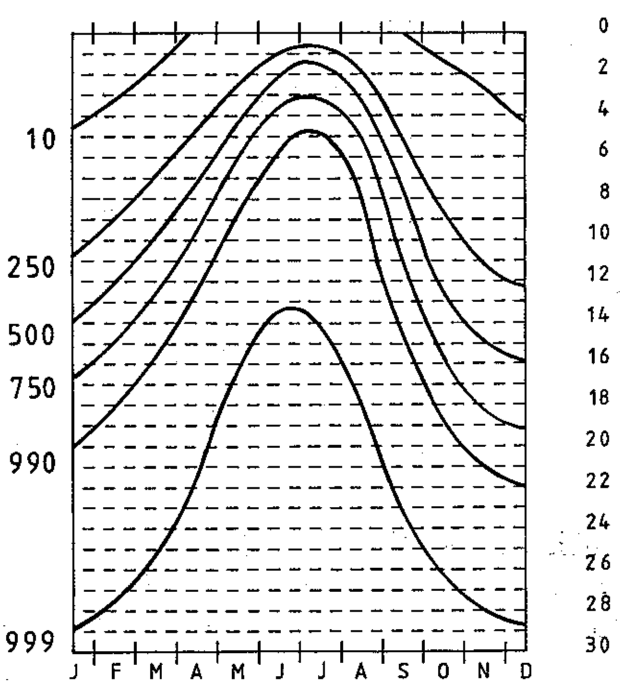


Fig. 11. Selected monthly probabilities in unit 1/1000 at MIKE that significant waveheight exceeds 2.75 m in at most r′ (numbers at vertical axis) out of 30 successive days.

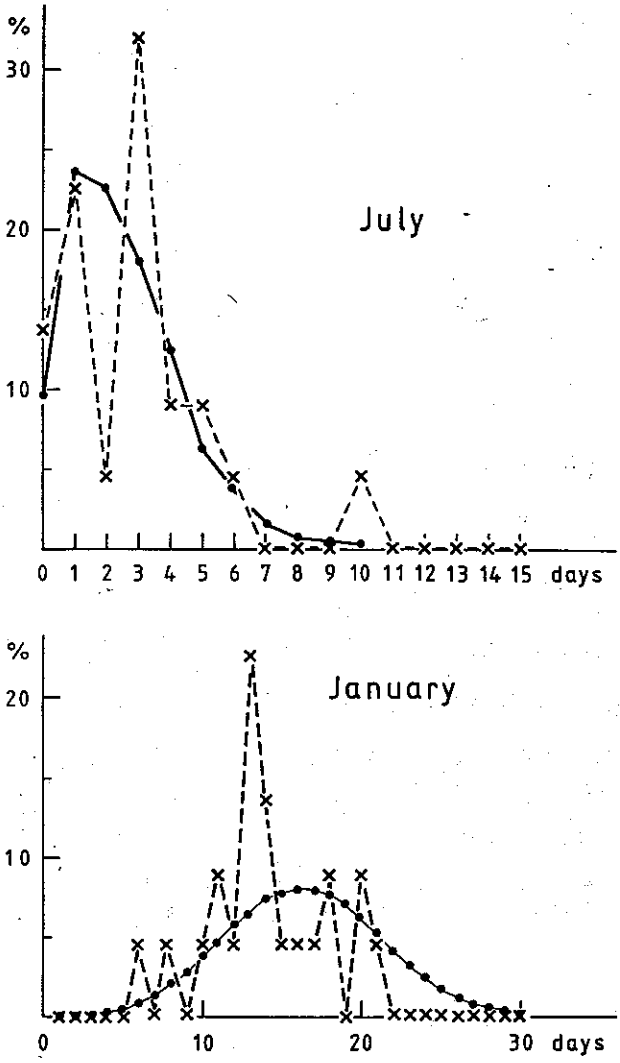


Fig. 12. Probabilities that visually estimated significant wave-height exceeds 2.75 m in a certain number of days out of 30 successive days in July and January.

Crosses: Estimates from direct counting

Dotts: Estimates from 1. order Markov chains applied to successive 6 days intervals.

These probabilities have been estimated for each calendar month using the visually estimated wave-heights at MIKE for the years 1954–1972, taking N = 6 and observations 24 hours apart. The starting days are the first days of the calendar months and M is running up to MM = 15 at most. Figure 11 shows selected percentiles of GC (r′, MN = 30) calendar month when h = 2.75 m. r′ is the integer 30 − r and accordingly

$$GC(r', MN = 30) = Pr[H \geq 2.75 \text{ m} \quad r' \text{ times at most in 30 days}]$$

It can be shown that the derived distributions fit well to observations and represent an improvement of the results obtained from the second order Markov model described in 3.1. Figure 12 illustrates how the model fits observations in July and January.

3.3 *Probability distributions of H mean over M equidistant, successive time points*

The recurrence scheme (33), (34) of part II may be used to derive this distribution. At each time point we consider the II categories

$$0 < H \leq \Delta h, \ \Delta h < H \leq 2\Delta h, \ldots,$$
$$(I - 1)\Delta h < H \leq I\Delta h, \ldots,$$
$$(II - 1)\Delta h < H,$$

labelled by the integers $X = \{1,2,\ldots,I,\ldots,II\}$ such that $x = I$ corresponds to $(I-1)\Delta h < H \leq I\Delta h$. The mean value HAM of the first M values of H may then be written approximately as

$$HAM = \frac{\Delta h}{M}(x_1 + x_2 + \ldots + x_M) - 0.5\,\Delta h = \frac{\Delta h \cdot q}{M} - 0.5\,\Delta h.$$

Hence

$$\Pr\left(HAM = \frac{\Delta h}{M} \cdot q - 0.5\,\Delta h\right) = \Pr(x_1 + x_2 + \ldots + x_M = q)$$

Using the scheme (33), (34) to find the probability distribution for q we have at the same time the probability distribution for $HAM = \frac{\Delta h}{M}q - 0.5\,\Delta h$. This has been accomplished for the normalized variable $h^* = h/\bar{h}(\tau)$ at MIKE, taking $II = 15$ and the corresponding interval length Δh^* as the unit of h^*. The transition probability matrix PP(I,J), I,J = 1,15, is estimated from 22 years of observations for pairs of days 5 times a month. The values are given below together with P(I) in Table 9.

The resulting interval frequencies of H*, H*A2, H*A15 and H*A30 are given in Figure 13 and give an impression of the rate of decrease of variance and approach to normal distribution.

3.4 *Probability distributions of HMAX in periods of arbitrary lengths*

Such distributions may be represented with negligible errors by certain asymptotic distributions if the periods are long enough, but will otherwise be of no help. In such cases we may take advantage of a model distribution for $F(h;\tau)$ and the relevant transition probabilities. In the following we shall briefly outline *one* such method, and next apply it as an example to the station Statfjord.

Table 9

MIKE

$PP(I,J) = \Pr(H^*_{M+1} = J - 0.5 \mid H^*_M = I - 0.5)$. Unit: 1/1000.

J \ I	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	325	47	11	7	7	3	3	2	1	1	2	3	3	3	2
2	245	398	128	80	47	32	26	19	11	11	11	14	11	8	4
3	177	314	340	224	175	158	106	93	75	80	74	69	48	34	19
4	122	131	238	303	231	200	199	176	149	136	133	124	111	85	37
5	76	63	128	158	191	196	209	185	194	170	178	138	138	127	112
6	27	26	83	99	140	158	173	176	149	170	181	138	144	153	187
7	14	10	34	61	93	106	105	111	130	132	130	122	133	136	225
8	5	4	17	30	47	61	63	74	100	109	74	110	111	121	187
9	3	3	9	16	30	32	46	65	78	64	63	89	94	119	112
10	2	2	6	10	17	24	30	44	42	43	52	69	83	85	56
11	1	1	3	6	11	15	18	19	25	32	37	52	56	68	26
12	1	1	2	3	5	6	9	15	19	21	28	34	33	34	19
13	1	0	1	2	3	4	6	9	13	17	19	21	22	17	7
14	1	0	0	1	2	3	4	7	10	11	15	14	11	8	4
15	0	0	0	0	1	2	3	5	4	3	3	3	2	2	2
$P(I) = \Pr(H^* = I - 0.5)$															
	14	72	222	236	162	117	72	41	24	17	10	5	3	2	1

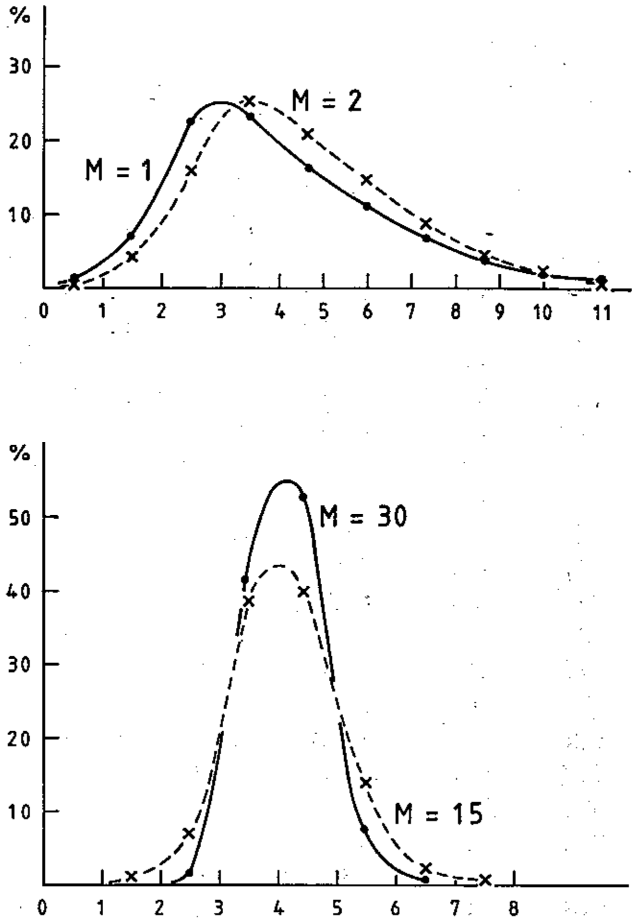


Fig. 13. Probability distributions of normalized significant wave-heights, $h/\bar{h}$, at MIKE averaged for 1, 2, 15, and 30 days. (The values of $h/\bar{h}$ have been multiplied by a certain scale factor).

We consider successive equidistant time points, for instance 24 hours apart, and arrange them in groups each of which contains N successive time points:

$$\begin{aligned} T_1 &= \tau_1, \tau_1 + 1, \ldots, \tau_1 + N - 1 \\ T_2 &= \tau_2, \tau_2 + 1, \ldots, \tau_2 + N - 1 \\ &\text{-----------------} \\ T_M &= \tau_M, \tau_M + 1, \ldots, \tau_M + N - 1 \end{aligned} \quad (2)$$

Let further $Fx(h;M')$ and $Fxx(h; M \cdot N, \tau_1)$ represent cumulative distributions of HMAX in $T_{M'}$ and $T_1, T_2, \ldots, T_M$, respectively. With *two* first-order Markov models these distributions may be written (cf. Part II, eq. (40)):

$$Fx(h;M') = F(h; \tau_{M'}) \prod_{N'=0}^{N-1} (F11(h; \tau_{M'} + N') \quad (3)$$

$$Fxx(h; M \cdot N, \tau_1) = Fx(h;1) \prod_{M'=1}^{M-1} Fx11(h; M') \quad (4)$$

Here F11 and Fx11 are the transition probabilities

$$F11\ (h;\tau) = Pr[H \leq h \quad at \quad \tau+1 \,|\, H \leq h \quad at \quad \tau]$$
$$Fx11\ (h;M') = Pr[HMAX\ in\ T_{M'+1} \leq h \,|\, HMAX\ in\ T_{M'} \leq h]$$

Inserting the inverses $h = H(F;\tau)$ and $H(Fx;M')$ we may write

$$F11 = F11(F;\tau) \quad (5)$$

$$Fx11 = Fx11(Fx;M') \quad (6)$$

It may be shown that these equations may be replaced approximately by

$$F11 = F11(F) \quad (7)$$

$$Fx11 = Fx11(F_x) \quad (8)$$

for larger parts of the year. At least (8) will be approximately true if N is not too large. Besides increasing the significance in the estimates of the functions F11 and Fx11, eqs. (7) and (8) also simplify computations somewhat.

For $F(h;\tau)$ we have the model representation

$$F\ (h;\tau) = G\left(\frac{h}{\bar{h}(\tau)};\ \alpha, \frac{1}{\alpha+1}\right) \quad (9)$$

$$\begin{aligned} \bar{h} = a_o + a_1 \cos\frac{2\pi\,\tau}{365} + b_1 \sin\frac{2\pi\,\tau}{365} \\ + a_2 \cos\frac{4\pi\,\tau}{365} + b_2 \sin\frac{4\pi\,\tau}{365} \end{aligned} \quad (10)$$

Table 10

STATFJORD
Pr (MAX SIGN. H in x days starting October 1 $\leq$ h). Units 1/1000, and metres

h: 6	7	8	9	10	11	12	13	14	x days
890	960	987	996	998.8	999.7	999.90	999.97	999.99	6
814	926	975	992	997.5	999.3	999.79	999.94	999.98	12
740	891	961	987	996.0	998.6	999.65	999.90	999.97	18
668	853	945	981	994.2	998.2	999.48	999.85	999.96	24
594	813	927	974	991.8	997.5	999.24	999.78	999.94	30
523	769	906	966	989	996.6	998.93	999.69	999.90	36
452	721	881	956	985	995.3	998.47	999.54	999.86	42
382	669	851	944	980	993.5	997.86	999.34	999.79	48
316	613	817	928	973	991.1	997.00	999.0	999.69	54
255	555	779	909	965	998	995.79	998.6	999.52	60
200	493	736	885	954	983	994.05	998.0	999.29	66
152	430	689	857	941	977	991.60	997.0	998.9	72
112	367	637	824	924	970	988	995.7	998.4	78
80	308	583	787	904	961	984	994.0	997.7	84
56	254	528	747	880	949	979	991.7	997.5	90
38	204	472	704	854	936	973	988.9	996.3	96
25	162	418	660	826	921	965	985.5	994.8	102
16	127	367	615	797	905	957	981.7	993.2	108
11	99	322	573	767	888	948	977.8	991.3	114
7	77	281	534	738	871	939	973.6	989.5	120
5	60	246	497	711	855	931	970.0	987.8	126
3	47	217	465	686	840	924	966.2	986.2	132
2	38	193	437	663	828	917	963.2	984.9	138
1.4	31	174	413	645	817	912	960.0	983.8	144
1.0	26	159	395	630	809	908	959.0	983.1	150

Eqs. (3), (4), (alternatively (5), (6)), (7), (8), (9) and (10) represent a scheme by which Fxx(h, M · N, τ_1) can be obtained given α, a_o, a_1, b_1, a_2, b_2 and the initial day τ_1.

Using N = 6 and τ_1 = October 1 the table above shows the results for STATFJORD up to M · N MAX = 150 days. Going from maximum *significant* wave-heights to maximum wave-heights, the heights above should be multiplied by a factor $\approx$ 1.9. Also, using 6 hours time resolution instead of the 24 hours used above, the heights should be magnified by a factor $\approx$ 1.15.

ACKNOWLEDGEMENTS

I am indebted to Inge Fjørtoft for carrying out extensive programming of the developed recurrence schemes and testing a number of wave- and meteorological data, including estimates of the parameter transition probabilities. Without his help and interest, this work would hardly have been possible.

REFERENCES

Fjørtoft, R. 1977. *Bidrag til beskrivelse av bølgeklimaet i norskehavet.* Internal Report, The Norwegian Meteorological Institute.

Haug, O. 1968. *A Numerical Method for Prediction of Sea and Swell.* Met. Ann.

Mood, A. M. & Graybill, F. A. 1963. *Introduction to the Theory of Statistics.* p. 410. McGraw-Hill.

Instructions to Authors

GEOPHYSICA NORVEGICA

publishes papers in English. When preparing manuscripts for submission, authors should consult current copies of the journal and follow its style as closely as possible.

MANUSCRIPTS

Manuscript must be typewritten, double spaced throughout, on one side of the paper, with a wide margin. Authors should submit the *original* manuscript (preferably with one copy) to the editor, whose address is shown on page 2 of the cover.

Separate sheets should be used for the following: 1) title page, with the author's name and institution, and, if the title is longer than 40 letters and spaces, a short title not exceeding this limit for use in the running heads; 2) an abstract not exceeding 12 lines (910 letters and spaces) with the name and full postal address underneath of the author to whom communications, proofs, and reprints are to be sent; 3) references; 4) Tables with their headings; 5) legends to Figures.

Brief *Acknowledgements* of grants and other assistance, if any, will be printed at the end of the text.

FIGURES, TABLES AND MATHEMATICAL SYMBOLS

All illustrations are to be considered as Figures. Each graph, drawing, or photograph should be numbered in sequence with arabic numerals, and should be identified on the back by the name of the journal, the author's name, and the Figure number. The top should be indicated. The Figures should be the original drawing. The columns of *Geophysica Norvegica* are 67 mm broad, and the size of the original drawings should be in proportion. Lines must be thick enough to allow for reduction. Letters and numbers should not be less than 2 mm high in the printed illustration. Photographs should be submitted as unmounted glossy enlargements showing good details.

Tables are to be numbered consecutively. Each Table should be typed on a separate sheet, with a descriptive heading that makes the Table self-explanatory.

All Figures and Tables should be referred to in the text by their number. Their approximate position should be indicated in the margin of the manuscript.

All numbered equations and all unnumbered but complicated equations should be typed on separate lines. Equations should be punctuated.

All text material will be set in roman type unless otherwise marked. Hence, all variables and other characters to be set in italic type should be underlined once with a straight line. Vectors and other characters in boldface type should be indicated by underlining with a single wavy line.

No footnotes should be used.

REFERENCES TO LITERATURE

In the text, Brown (1957, p. 9), Brown & White (1961). If more than two authors, Brown et al. (1963). Multiple references: 'As several authors have reported (Brown 1967, Brown & White 1961, Green et al. 1963)', i.e. chronological order, no commas between names and year.

Lists of References are to be unnumbered and in alphabetical order. The international alphabetical order of Scandinavian and German vowels should be observed: Å = AA, Æ and Ä = AE, Ø and Ö = OE, Ü = UE. Indicate 1st, 2nd, 3rd, etc. works by the same author in the same year by a, b, c, etc. (White 1966a). No ditto marks should be used. Titles of journals should be abbreviated according to *World List of Scientific Periodicals.*

Examples:

Cadle, R. D. 1966. p. 83 in *Particles in the Atmosphere and Space.* Reinhold Publishing Corporation, New York.

Craig, R. A. 1965. p. 161 in *The Upper Atmosphere. Meteorology and Physics.* International Geophysics Series, Vol. 8. Academic Press, New York and London.

Eliassen, A. & Kleinschmidt, E. 1957. p. 66 in *Handbuch der Physik.* Vol. 48, Part 2, edited by S. Flügge. Springer-Verlag, Berlin.

Junge, C. 1972. *Quart. J. R. Met. Soc. 98,* 711.

PROOFS

Two copies of the first proof will be sent (page proofs). One copy, duly corrected, should be returned to the editor with the least possible delay. All technical parts of the article, including references, names, figures (numbers, formulae), illustrations, etc. are the responsibility of the authors. Authors will be required to pay for any major alterations they may make.

REPRINTS

Fifty reprints of each article will be supplied free. Additional reprints can be ordered at a charge.

International Journals

Title	Languages	Summary/ Abstract	Issues per volume
BLYTTIA Journal of the Norwegian Botanical Association	Norwegian	English	4
BOREAS An International Journal of Quaternary Geology	English, French and German	English	4
LETHAIA An International Journal of Palaeontology and Stratigraphy	English, French and German	English	4
LITHOS An International Journal of Mineralogy, Petrology and Geochemistry	English, French and German	English	4
NORSK GEOGRAFISK TIDSSKRIFT Norwegian Journal of Geography	English, Norwegian and German	English	4
NORSK GEOLOGISK TIDSSKRIFT Norwegian Journal of Geology	English, French and German		4

Periodicals

Title	Languages
ACTA BOREALIA Journal of Arctic Botany, Geology and Zoology	English
ASTROPHYSICA NORVEGICA Norwegian Journal of Theoretical Astrophysics	English
FOLIA LIMNOLOGICA SCANDINAVIA Scandinavian Journal of Limnology	English
SARSIA Journal of Marine Biology	English

UNIVERSITETSFORLAGET

GEOPHYSICA NORVEGICA · VOL. 32 · NO. 3 · PAGES 45-76 · 1982 UNIVERSITETSFORLAGET