LONDON, METEOROLOGICAL OFFICE.

Met.O.3 Technical Note No.9.

A statistical model to simulate sequences of hourly mean wind speeds over the United Kingdom. By SMITH,S.G.

London,Met.Off.,Met.O.3 Tech.Note No.9, 1981,31cm.Pp.12,12 pls. 10 Refs.

An unofficial document - restriction on first page to be observed.

Met O 3 Technical Note No 9

A Statistical Model To Simulate Sequences Of Hourly Mean Wind Speeds Over The United Kingdom

by

S G Smith

Climatological Services Branch (Met O 3)
Meteorological Office
London Road
Bracknell
Berkshire

October 1981

CONTENTS

A Statistical Model To Simulate Sequences Of Hourly Mean Wind Speeds Over the United Kingdom

1. Introduction

The Climatological Services Branch of the Meteorological Office are often asked to provide climatological wind speed statistics for a place or for an area in the United Kingdom. The enquiry can be answered with some confidence if a long period, homogeneous record exists for the location. However this is usually the exception rather than the rule - frequently estimates have to be derived from the nearest station or stations that have a suitable length of record and that can be considered representative of the required area.

It would therefore be an advantage if a statistical model could be produced which would simulate all the important characteristics of a time series of wind speeds and could be used to generate any required length of data. The data would not suffer from problems regularly encountered with actual wind speed time series, such as inhomogeneities, missing values and shortness of record. By a suitable variation of the parameter values of the model it might be hoped to apply it to a wide area of the United Kingdom, including places for which very little data are currently available.

The model described generates continuous sequences of hourly mean speeds, from which daily, monthly etc statistics can easily be derived. This report deals primarily with application of the model to just one location - Birmingham Airport, although its use for other areas is briefly discussed. A subsequent report will describe in greater detail its application to different areas.

2. Literature

Before describing the model it is appropriate to discuss the existing literature on this type of work. Much effort has been directed at modelling speeds for estimating available power from wind energy conversion systems. For this application correct fitting of the highest and lowest speeds is not crucial because wind energy systems are not usually designed to function at relatively low or relatively high speeds. In this study, however, emphasis was placed on an accurate representation of the high speeds because enquiries on these class of speeds form a large proportion of the requests received in the Climatological Services Branch. Often wind speed models are specific to a particular site or a well defined, relatively small area, whereas for the work here it was the intention that the model be applied to a large part of the country.

Bean and Somerville (1979) and Sneyers (1980), amongst others, have modelled the distribution function of wind speeds but not a numerical sequence of values. Vukovich and Clayton (1977) and Cliff et al (1978) describe techniques for estimating speeds at points close to, or that can be considered representative of, a site for which a long period record already exists. Workers that have modelled data at a site in order to generate speeds at that site include Goh and Nathan (1979), Chou and Corotis (1981) and O'Carroll and Williams (1981). The last named use "Box-Jenkins" methods to simulate wind speeds and other parameters for Sullom Voe, in the Shetlands.

3. Data

It was decided initially to fit a model to a wind speed series that was (i) sufficiently long to enable stable estimates of parameter values to be calculated and (ii) from an inland site that was representative of a fairly wide area. Elmdon (Birmingham Airport) was selected; it has no missing hourly mean values over the 19 year period 1961-79, the site is well exposed and can be considered representative of similarly exposed sites in the English Midlands.

In what follows, the term "daily cycle", for example, is used rather loosely to mean a variation of speed having a period of about a day but which is not necessarily regular in either magnitude or period. Figure 1 shows a plot of the speeds at Elmdon for most of 1961. Three distinct cycles or oscillations in the time series were identified:

(i) A daily cycle. This cycle is caused by the effects of solar heating of the land during the daytime, leading to increased turbulence, convection and wind speed in the surface boundary layer. The amplitude of the cycle varies from day to day, depending for example on cloud cover, and the magnitude itself has an annual period, being greater in summer than winter.

(ii) Synoptic cycle. The passage of synoptic features such as depressions and anticyclones in the vicinity of the British Isles leads to fluctuations in the mean speed with a period of a few days. Both the amplitude of the cycle and the frequency of its occurrence are irregular.

(iii) Annual cycle. A seasonal variation in wind speed is present, with mean speeds generally greatest in winter and least in summer. This is largely a reflection of the fact that synoptic depressions in the mid-latitudes of the Northern Hemisphere are usually more vigorous in winter than in summer. The seasonal variation is more readily apparent from figures 2(a), (b), which show the mean monthly speeds and mean monthly extremes (an extreme here is the highest hourly mean for the month). Interesting features of these and other graphs are that values for February and October are less than might be expected from the values for neighbouring months. These results are probably peculiar to the period 1961-79; if a different period had been chosen other months may have given anomalies. To reduce the effects of sampling discrepancies, mean monthly values through the year have been harmonically smoothed by a procedure which is described later.

4. The model

For brevity, only a broad outline of the model will be given and the justification for some of the details will be omitted.

a) Synoptic cycle

In view of the fact that high speeds were of most concern the synoptic cycle (see section 3), which gives rise to the highest speeds, was modelled as a distinct entity. The cycle was considered to be caused primarily by "depression events" (DEs). A DE was defined as a period during which the wind speed at Elmdon exceeded 15 Kn and at some time rose above 20 Kn. However events were excluded if all the following conditions were satisfied:

(i) The maximum speed did not exceed 25 Kn

(ii) The maximum occurred between 1300 and 1500 GMT inclusive

(iii) The DE began after sunrise and ended before sunset on the day in question. Events thus omitted were assumed to occur more through the effects of daytime insolation rather than the proximity of a depression.

Over the 19 year Elmdon record, there were 778 DEs of which 121 were excluded by the above criteria. A breakdown of the remaining 657 is given by year and month in table 1. The following statistics were calculated for each DE and results stored by month:

(i) Maximum speed attained

(ii) Time in hours from the last hour of the event to the first hour of the next event (hereafter referred to as the separation)

(iii) Duration of the event.

The distribution of maximum speeds is shown for January and July in figure 3 and of separations, for given ranges, for the same months in figure 4. The Gamma distribution, which is described by two parameters α and β representing the shape and scale of the distribution respectively, was employed to fit the distributions of maximum speeds and of separations. The Gamma distribution has been widely used in meteorology to fit, for example, rainfall amounts (Gray (1976)) and wind speeds (Sneyers (1980)). Details of the derivation of the estimated monthly values of the parameters, $\hat{\alpha}$ and $\hat{\beta}$, are given in the appendix. $\hat{\alpha}$ was found to vary between 1.16 for January to 2.55 for June. For the separations it ranged from 0.49 for February, March and May to 0.72 for July. In both cases there was little consistent pattern in the month-to-month variation and hence a constant value was assumed for each: $\hat{\alpha} = 1.5$ for the maximum speeds and 0.5 for the separations.

The means of both variables are presented in figures 5(a), (b), together with harmonically smoothed speeds derived from a least squares regression analysis with the first two harmonics of the annual cycle. These smoothed values were used to determine $\hat{\beta}$.

The third variable calculated, the duration of the DE, was plotted against maximum speed, for each month separately. A linear relationship between the variables was evident, although there was a large scatter about the regression lines. There were no systematic variations from month to month for the regression equations and so a single equation was applied to all months, with coefficients determined by the least squares criterion, giving

$$d = 1.38 X_{max} - 21.15 \qquad (1)$$

where d is the duration in hours and X_{max} is the maximum speed (Kn).

This equation is used in the model to produce simulated durations from simulated maximum speeds.

(b) Daily cycle

This cycle was also treated as a separate entity. It was assumed to follow a truncated sine curve as shown:

The magnitude of the cycle, b, was taken to be the difference between the wind speed at 1400 GMT and the speed at sunrise, where sunrise was defined to be the hour nearest to the actual time of sunrise on the 15th day of the month. Results were classified according to the speed at sunrise, namely 0-4, 5-9, 10-14, 15-19 and over 19 Kn. If the magnitude was zero or less, no daily cycle was assumed to have occurred. The cumulative distribution of the magnitude of the cycle was determined for each month and for each category of speed at sunrise. This distribution was used to generate values representing the magnitude of the daily cycle. The distributions for January and July are given in figure 6. These show, for example, that for speeds at sunrise between 0-4 Kn, 56% of days in January had no daily cycle compared to 27% of days in July. Note that results for the 20+ category were combined over all months. The curves reflect the fact that the magnitude of the cycle is generally greater in summer than winter and is inversely proportioned to the speed at the beginning of the day.

c) Annual cycle

It was found that terms explicitly representing the annual cycle were in fact not required since the nature of the terms simulating the DEs produced higher mean speeds in winter than in summer.

d) Non-forcing conditions

It was assumed that for periods not part of a DE or daily cycle the speed was constant except for random hour-to-hour fluctuations. The average speed was therefore calculated for each month for hours when neither DEs nor daily cycles were experienced. Mean monthly and standard deviations of the monthly averages were then calculated and harmonically smoothed. Values are shown in figures 7(a), (b).

5. Generation of speeds

A computer program, written in Fortran, was developed to generate sequences of hourly mean speeds for any number of years. The generated speeds are stored in two parallel series. A "base" series holds the values unmodified; these are in floating point format and some may be less than zero. This series is used to produce one value from previous values, as for example in (2) below. An "output" series, in contrast, holds speeds rounded to the nearest integer and any negative values are set to zero.

For each series generated, values are obtained starting from the first hour of the year, with an initial speed of 10 Kn supplied to the program. The following steps are then executed:

a) Generate a value t_{DE} from the Gamma distribution (section 4(a)) with the magnitude of $\hat{\beta}$ appropriate for January (the starting month) and $\hat{\alpha}$ equal to 0.5. t_{DE} represents the time in hours until the first depression event (DE). Assume for convenience that t_{DE} is greater than 24, so that there is no DE beginning on the first day.

b) From the first hour of the day until sunrise, call time t_s , the speed fluctuates about a mean value $\bar{X}$ such that

$$X_t = \phi X_{t-1} + (1-\phi)\bar{X} + \epsilon_t \qquad (2)$$

where X_t , X_{t-1} are the present and previous hour's speeds respectively
ϕ is a constant which equals 0.93 and corresponds to the first lag auto-correlation
ϵ_t is from a Normal distribution with mean zero and standard deviation 1.25 and represents the random hour-to-hour fluctuations
$\bar{X}$ itself is Normally distributed with parameters as described in 4(d).

c) At time t_s the program generates a random number b corresponding to the magnitude of the daily cycle for that day, where b depends on the month and the speed at t_s . If b equals zero, there is no daily cycle and (2) is repeated until t_s the following day or t_{DE} , whichever occurs first. If b exceeds zero, the generated speeds follow a truncated sine curve with superimposed random fluctuations:

$$X_t = X_s + b \sin\left(\frac{\pi}{l_i+1}(t-t_s)\right) + \epsilon_t \qquad (3)$$

Here l_i is the duration of the cycle (sunrise to sunset)
X_s is the speed at t_s
ϵ_t is distributed as in (2).

d) Between sunset and sunrise the following day, assuming a DE does not begin in this period, one of two steps is carried out: if the speed at sunset is above 10 Kn equation (2) operates. If not, a truncated sine curve is followed, as in (3), with b = -2 and X_s now denoting the speed at sunset.

e) Eventually t will equal t_{DE} . Then the maximum speed c for the DE is generated from the Gamma distribution and the duration derived from (1) with a random component added to simulate the scatter about the linear equation relating the variables. The time of maximum speed, t_{max} , relative to the start and finish of the DE is also allowed to vary, using random values generated from a Normal distribution. If we take the DE to run from time t_{DE} to t_2, generated speeds are given by

$$X_t = 15 + (c - 15) \times \frac{(t - t_{DE})}{t_{max} - t_{DE}} + \epsilon_t \quad : t = t_{DE}, t_{max}$$

$$X_t = 15 + (c - 15) \times \frac{(t_2 - t)}{t_2 - t_{max}} + \epsilon_t \quad : t = t_{max} + 1, t_2 \qquad — (4)$$

where ϵ_t is as defined in (2). The value 15 appears in the expressions because the duration of the DE corresponds to the length of time the speed exceeds 15 Kn. It is seen that the speed rises to, and falls from, the maximum in a linear fashion with random perturbations. Arguably other relationships might be more appropriate. However, Shellard (1975), after plotting profiles of wind speed with time for 21 occasions of strong winds at Lerwick concluded that "it cannot be said there is a typical storm profile".

After generating speeds in this DE, the program supplies another value t_{DE} corresponding to the time of the next event. Then (2) or (3) is applied depending on which hour of the day the DE has ended.

The random numbers used in the program have been produced from the random number generators available in the "NAG" computer package. The program uses about 45 seconds CPU time on the 360/195 (equivalent to approximately 15 units) to generate 100 years of hourly data.

6. Assessment of the model

i) Discussion

Given sufficient time, it would have been possible to devise a model which would have simulated, to a very high degree of accuracy, all the statistical properties of Elmdon data between 1961-79. However such a model would be of limited worth because

a) it would inevitably contain a considerable number of terms and parameters, some of which would have little or no physical interpretation. This would make the model difficult to apply to other areas.

b) the wind speed characteristics of the period 1961-79 will not be repeated exactly in any past or future 19 year period because our climate is not constant.

It was therefore not expected and not intended that the simulated data should precisely reproduce the 1961-79 statistics. The adequacy of the model can only by gauged by studying various criteria and subjectively judging whether the magnitude and frequency of any significant differences between actual and simulated speeds are acceptable.

To test the model, simulated data have been compared with the 1961-79 speeds. A more stringent test would have been to compare the model data with observed speeds from a different period but there were insufficient data to permit this.

ii) Results

Various statistics were derived from 100 years of generated data and compared with corresponding statistics from the 19 years of actual speeds. Mean monthly averages are compared in figure 8(a) and mean monthly extremes in figure 8(b). The 95% confidence limits drawn are equivalent to ± 2 standard errors of the averages or extremes for the actual data. The agreement between simulated and actual values is generally good, although there is a tendency for the simulated extremes to be rather high between February and July and rather low for November to January. A better fit would have been obtained if the actual values had been harmonically smoothed as had been applied to the parameter values in the model.

Table 2a presents statistics for the annual means and annual extremes. To make the comparison for the lowest and highest values valid, the simulated values have been obtained by averaging the lowest and highest from each of five consecutive 19 year periods contained in the 100 year sample. For the annual mean there is excellent agreement in the average and standard deviation but the lowest and highest simulated values are somewhat low. The results for the annual extremes are also encouraging. Table 2b lists the lowest and highest annual extreme speeds obtained in the five 19 year periods from which the means in table 2a were derived. A speed of 54 Kn is thus seen to be the highest speed generated in the 95 year record (in fact it was also the highest in the complete 100 years). Hardman et al (1973) have plotted "once in 50-year" hourly mean speeds for the United Kingdom and the value for the Elmdon area is about 27.5 ms^{-1}, or 53 Kn. Figure 4 of Hardman et al suggests that the "once in 100-year" speed is unlikely to be more than 4 Kn greater than the 50 year extreme. Hence the value of 54 Kn obtained appears realistic. However it is acknowledged that one of the 50 year sub-samples would have had a maximum speed of below 44 Kn, which is rather lower than might be expected.

Figure 9 shows the percentage frequency of speeds occurring in different ranges for four months and for the year, for the actual and simulated (100 years) data. The percentages for the "greater than 33 Kn" category are given at the top of each bar. The results are satisfactory.

Autocorrelations for lags of zero to 48 hours were calculated for Elmdon 1961 data and one year of simulated data. They are plotted in figure 10. The values for the simulated speeds are, in general, rather too high and the peaks at 24 and 48 hours are not as well developed as for the actual data.

Further tests would probably be required before the model could be accepted as satisfactory for simulating Elmdon wind speeds. The results seem reasonable on the basis of the above comparisons, although refinements are probably desirable, particularly with respect to the daily cycle. Also the validity of the highest speeds generated is difficult to determine without a longer record of actual speeds.

7. Application of the model to other areas

As a summary, listed below are the principle features of the model together with the parameters required to simulate them. Some, but not all, of the parameter values vary on a monthly time scale.

(a) Depression events - Two parameters each are needed for the maximum speed, separation and duration.

(b) Daily cycle - The cycle incorporates 65 values, representing cumulative percentages of the magnitude of the cycle exceeding 13 different speeds for 5 ranges of initial speed.

(c) Non-forcing conditions - The mean, variance and lag one auto-correlation coefficient are required.

(d) Random hour-to-hour component. The only non-constant parameter is the variance.

Although 65 values are required for the daily cycle many could be estimated by interpolation without much loss of accuracy and their relative magnitude may not significantly alter for many areas. The length of data required to estimate parameter values differs considerably for the different features - at least 10 years for (a) but one year is probably sufficient for (b) and (d).

The primary "raison d'etre" of the model is to produce speeds from which statistics of strong winds can be derived. This means that
- features (b), (c) and (d) are relatively unimportant and
- if (a) is not simulated adequately the model is of little use.

For these reasons it was decided, for application of the model to other areas, to concentrate on the simulation of the speeds in depression events. Another factor taken into consideration was that anemographs in general operation in the United Kingdom are not designed to measure speeds below about 5 Kn accurately. Anomalies have in fact been found in the frequencies of low speeds recorded at different stations. It was therefore advisable, at least in the early stages, to avoid attempting to simulate these low speeds at different places. Work on the simulation of speeds in depression events at various stations is still continuing (as at October 1981) and will be described in a later report. Problems that have had to be overcome are that the effective heights of the anemometers at different stations are not the same and that there are few stations with long period homogeneous records on which to base estimates of parameter values.

8. Summary and Conclusion

A statistical model has been described which simulates the time series of hourly mean wind speeds for Birmingham Airport and can generate a sequence of values with properties which are in reasonably good agreement with those of actual data. It is intended that the model be adaptable to other areas of the United Kingdom and the possibilities for this have been briefly discussed. A later report will detail these results.

9. References

Bean, S J & Somerville, P N — 1979 — Some models for wind speed. Univ of Central Florida, Orlando, Flo. Sci Report No 4, Contract No F19628-77-C-0080.

Chou, K C & Corotis, R B — 1981 — Simulation of hourly wind speed and array wind power. Solar Energy, 26, pp 199-212.

Cliff, W C Justus, C G & Elderkin, C E — 1978 — Simulation of the hourly wind speeds for randomly dispersed sites. Batelle Pacific NW Lab, PNL 2523.

Goh, T N & Nathan, G K — 1979 — A statistical methodology for study of wind characteristics from a close array of stations. Wind Eng, 3, No 3, pp 197-206.

Gray, Barbara M. — 1976 — Medium term fluctuations of rainfall in south-eastern England. Q.J.R. Met. S., 102, pp 627-638.

Hardman, Carol E, Helliwell, N C & Hopkins, J S — 1973 — Extreme wind speeds over the United Kingdom for periods ending 1971. Met Office, Climatological Services, Climatological Memo No 50A.

O'Carroll, F M & Williams, Patricia M — 1981 — A stochastic model for wind, weather and visibility at Sullom Voe. British Petroleum internal report (for private circulation only).

Shellard, H C — 1975 — Lerwick anemograph records 1957-70 and the offshore industry. Met Mag, 104, pp 189-208.

Sneyers, R — 1980 — Un modèle statistique de la répartition des vents et en direction. Inst Royal Meteorologique de Belgique. Publications, Serie B No 107.

Vukovich, F M & Clayton, C A — 1977 — On a technique to determine wind statistics in remote locations. Res Triangle Inst, RLO-2445-78/1.

10. Appendix

The Gamma distribution has p.d.f.

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

where Γ (x) is the Gamma function. The values of α and β were estimated for each month separately using maximum likelihood estimators viz

$$\hat{\alpha} = \frac{1 + \sqrt{1 + 4A/3}}{4A} \qquad \text{where } A = \ln\left(\frac{\sum x_i}{N}\right) - \frac{\sum \ln x_i}{N}$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}}$$

Here x_i are the individual monthly observations, $\bar{x}$ is the mean and N the number of observations. For the maximum speeds, which are all greater than or equal to 21 kn, 20.5 was subtracted from each value before deriving $\hat{\alpha}$ and $\hat{\beta}$.

Fig 1

Plot of Elmdon hourly mean wind speeds during 1961

ELMDON DATA FOR 1961

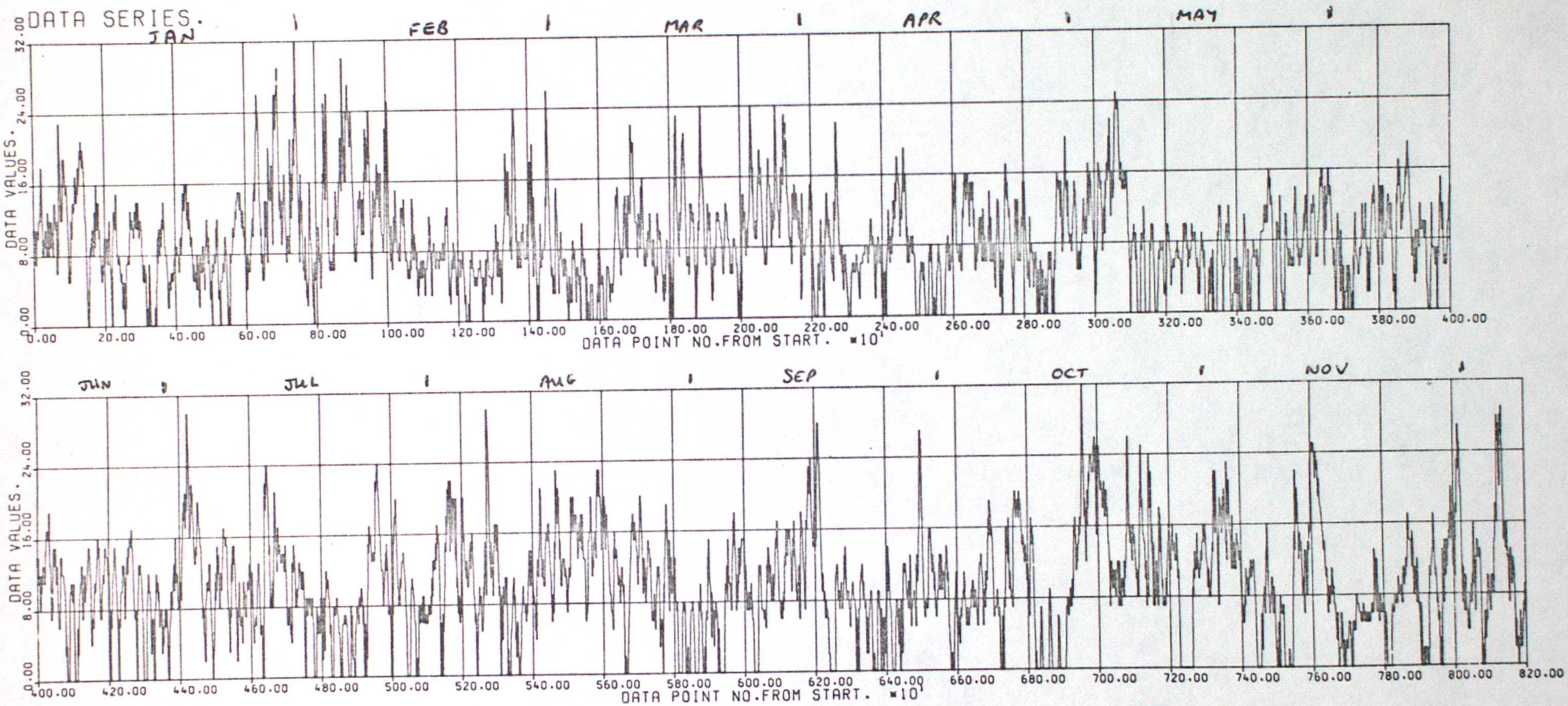


Fig 2a

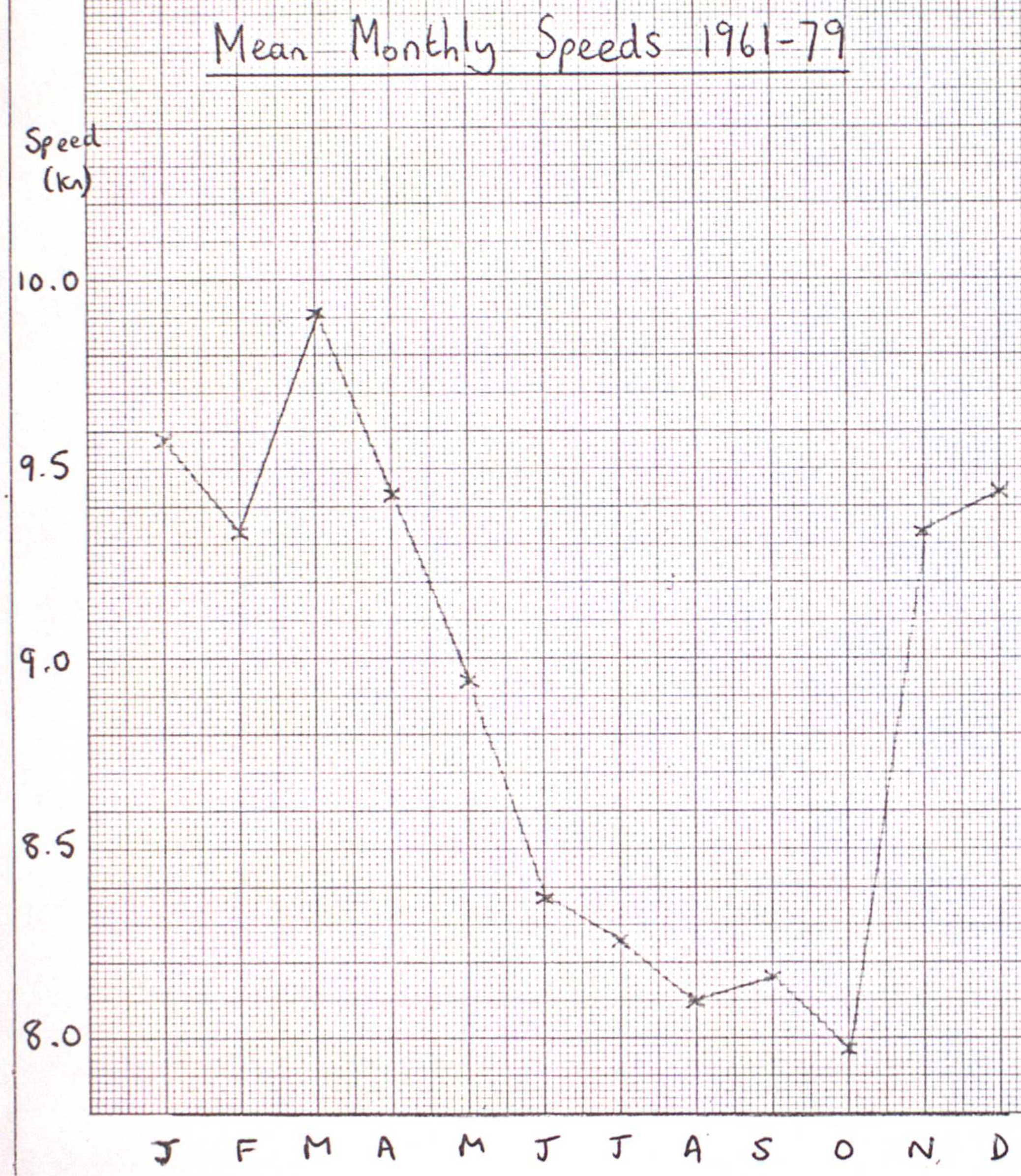
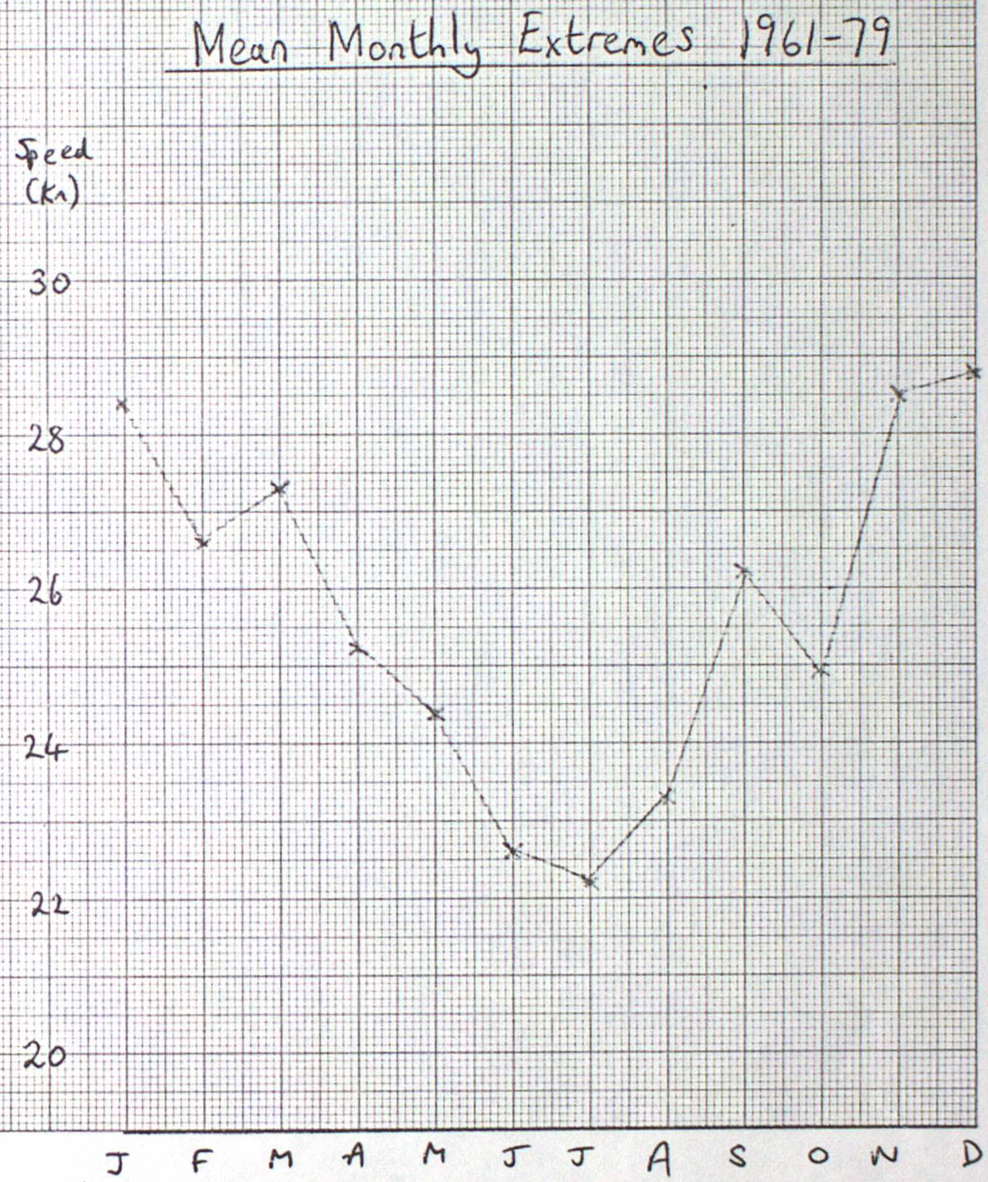


Fig 2b.



Month

Fig 3

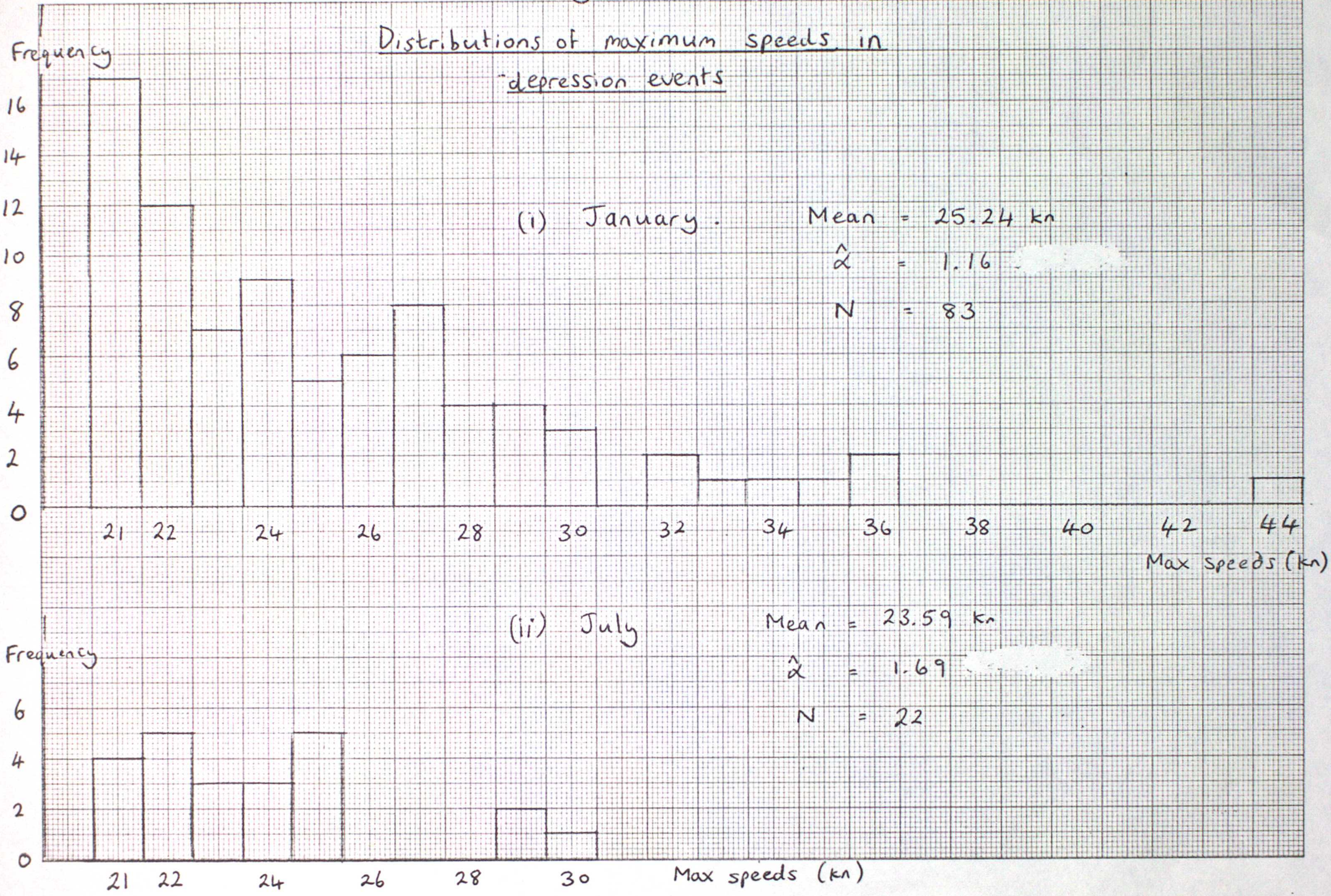
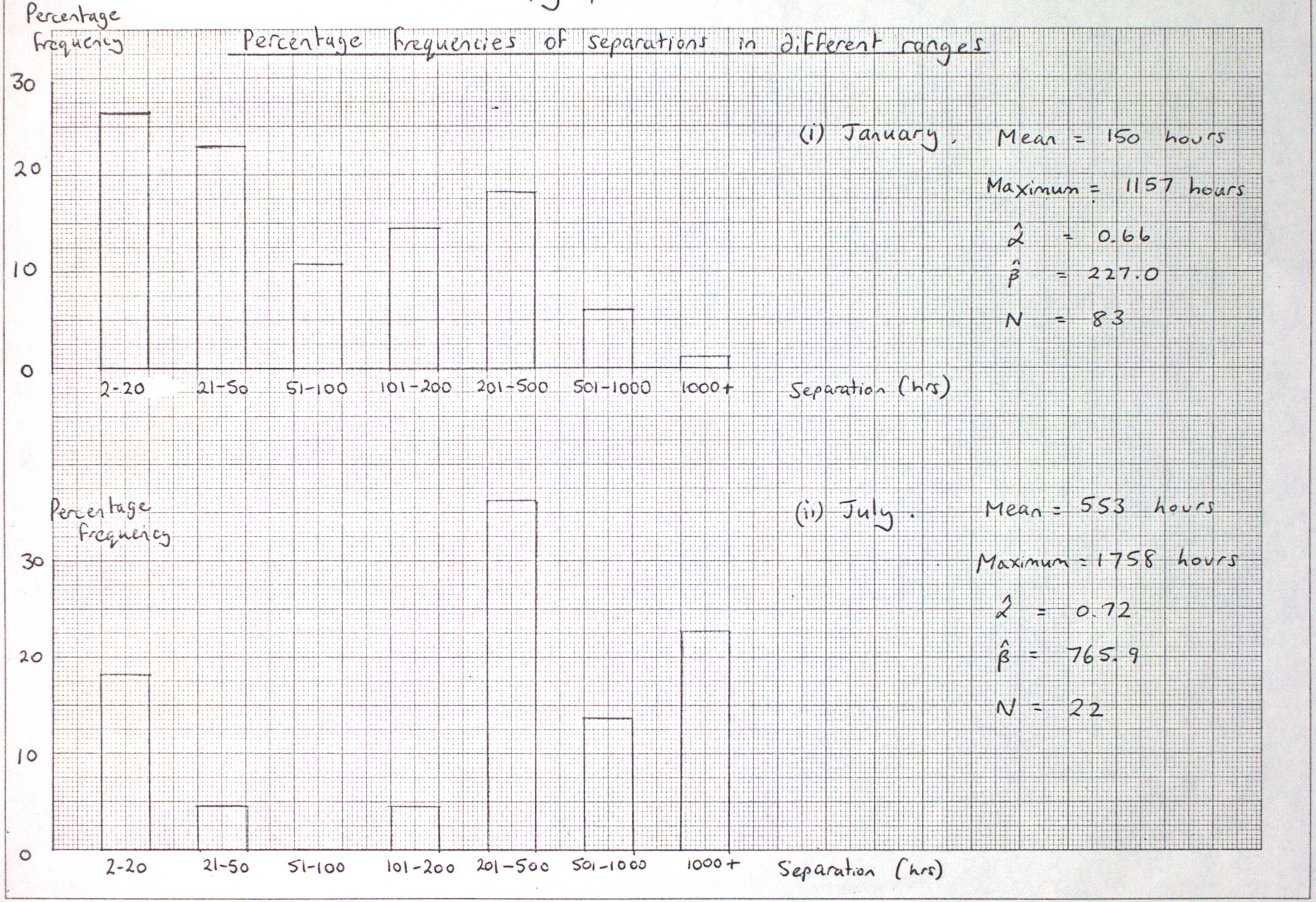

Distributions of maximum speeds in depression events
(i) January.
Mean = 25.24 kn
$\hat{\alpha}$ = 1.16
N = 83
Frequency
0
2
4
6
8
10
12
14
16
21
22
24
26
28
30
32
34
36
38
40
42
44
Max speeds (kn)
(ii) July
Mean = 23.59 kn
$\hat{\alpha}$ = 1.69
N = 22
Frequency
0
2
4
6
21
22
24
26
28
30
Max speeds (kn)

Fig 4



Percentage frequencies of separations in different ranges
Percentage frequency
30
20
10
0
2-20
21-50
51-100
101-200
201-500
501-1000
1000+
Separation (hrs)
(i) January.
Mean = 150 hours
Maximum = 1157 hours
$\hat{\alpha}$ = 0.66
$\hat{\beta}$ = 227.0
N = 83
Percentage frequency
30
20
10
0
2-20
21-50
51-100
101-200
201-500
501-1000
1000+
Separation (hrs)
(ii) July.
Mean = 553 hours
Maximum = 1758 hours
$\hat{\alpha}$ = 0.72
$\hat{\beta}$ = 765.9
N = 22

Fig 5a

Fig 5b

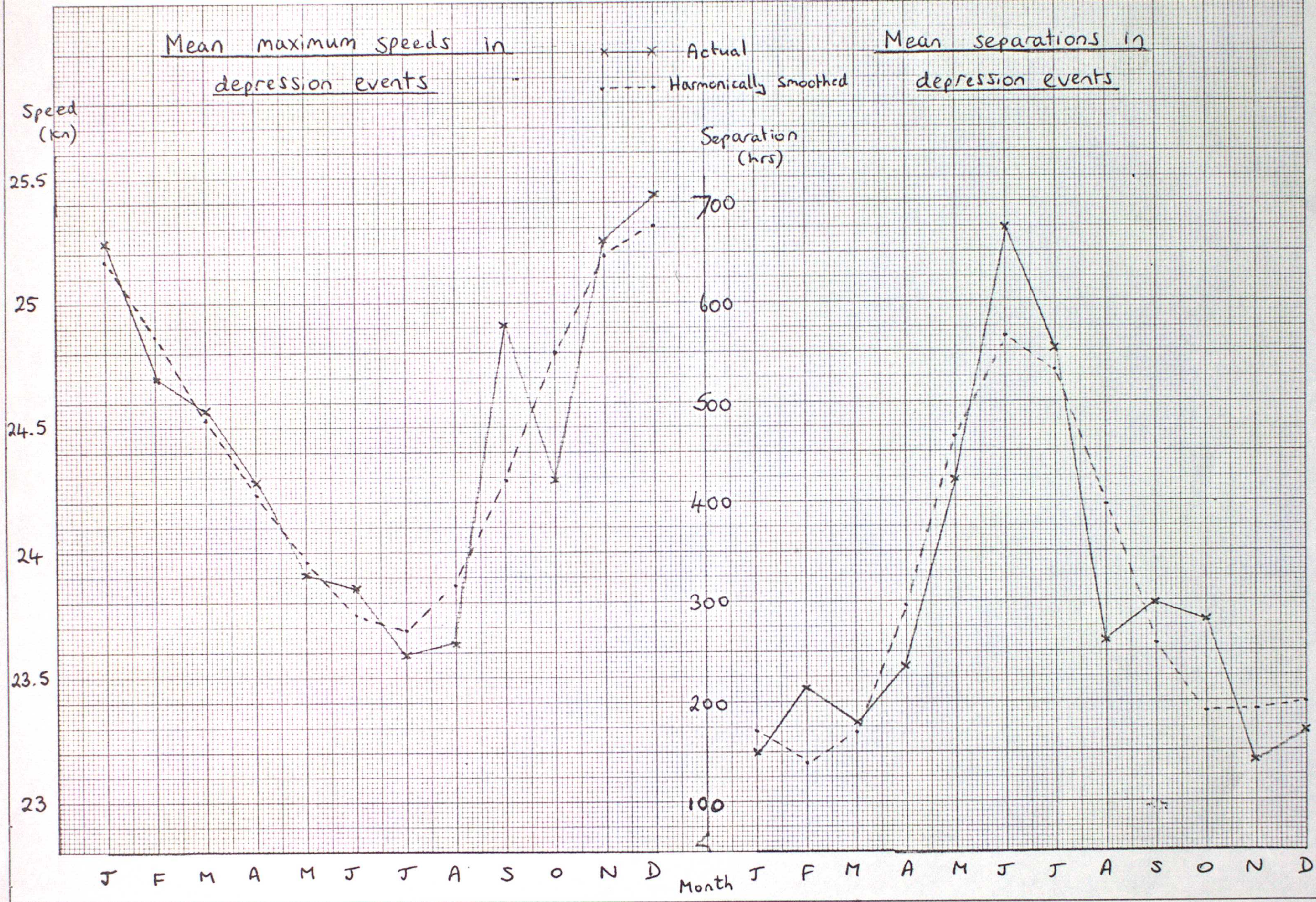
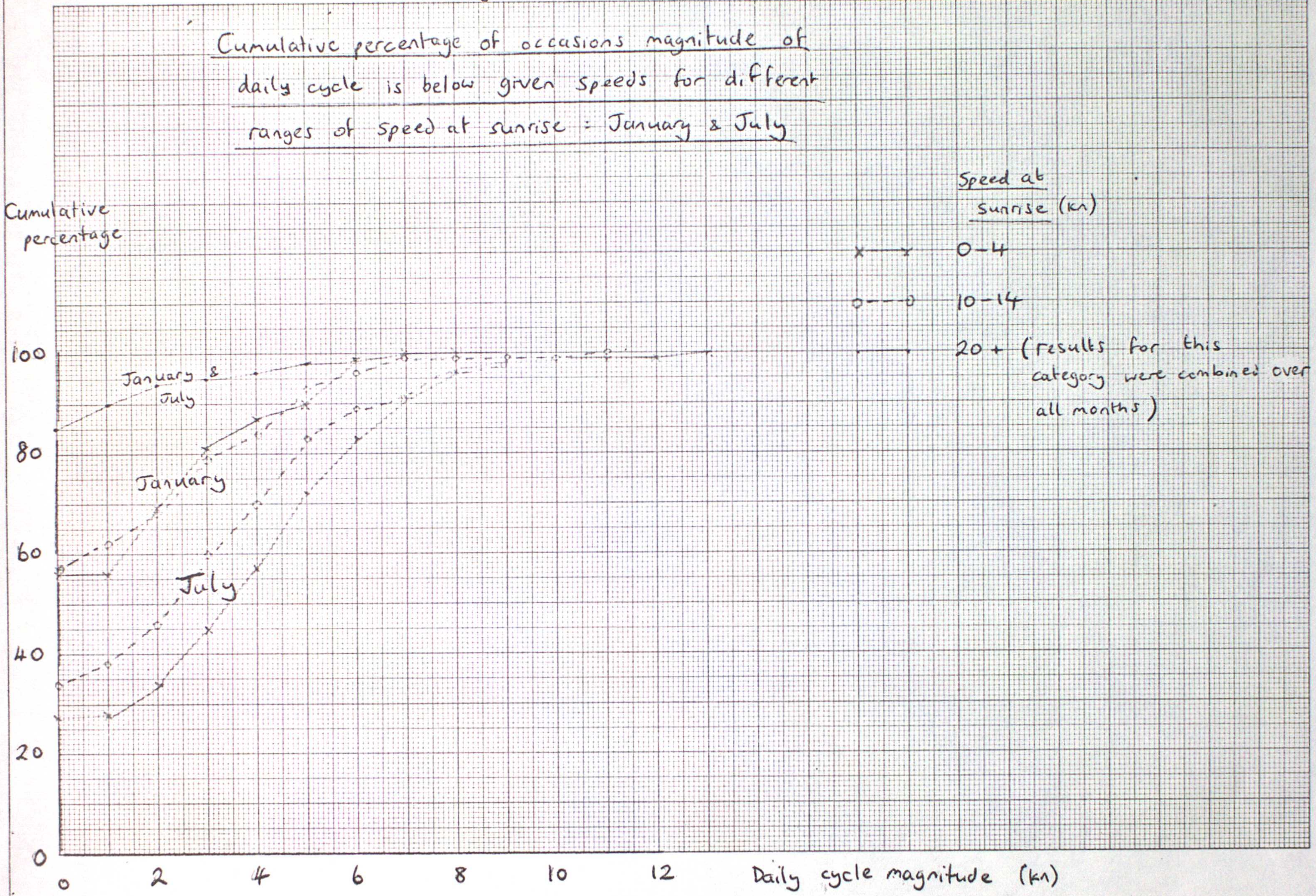

Mean maximum speeds in depression events
Actual
Harmonically smoothed
Mean separations in depression events
Speed (kn)
25.5
25
24.5
24
23.5
23
Separation (hrs)
700
600
500
400
300
200
100
J F M A M J J A S O N D
Month
J F M A M J J A S O N D

Fig 6


Cumulative percentage of occasions magnitude of daily cycle is below given speeds for different ranges of speed at sunrise : January & July
Speed at sunrise (kn)
0-4
10-14
20+ (results for this category were combined over all months)
Cumulative percentage
100
80
60
40
20
0
January & July
January
July
0
2
4
6
8
10
12
Daily cycle magnitude (kn)

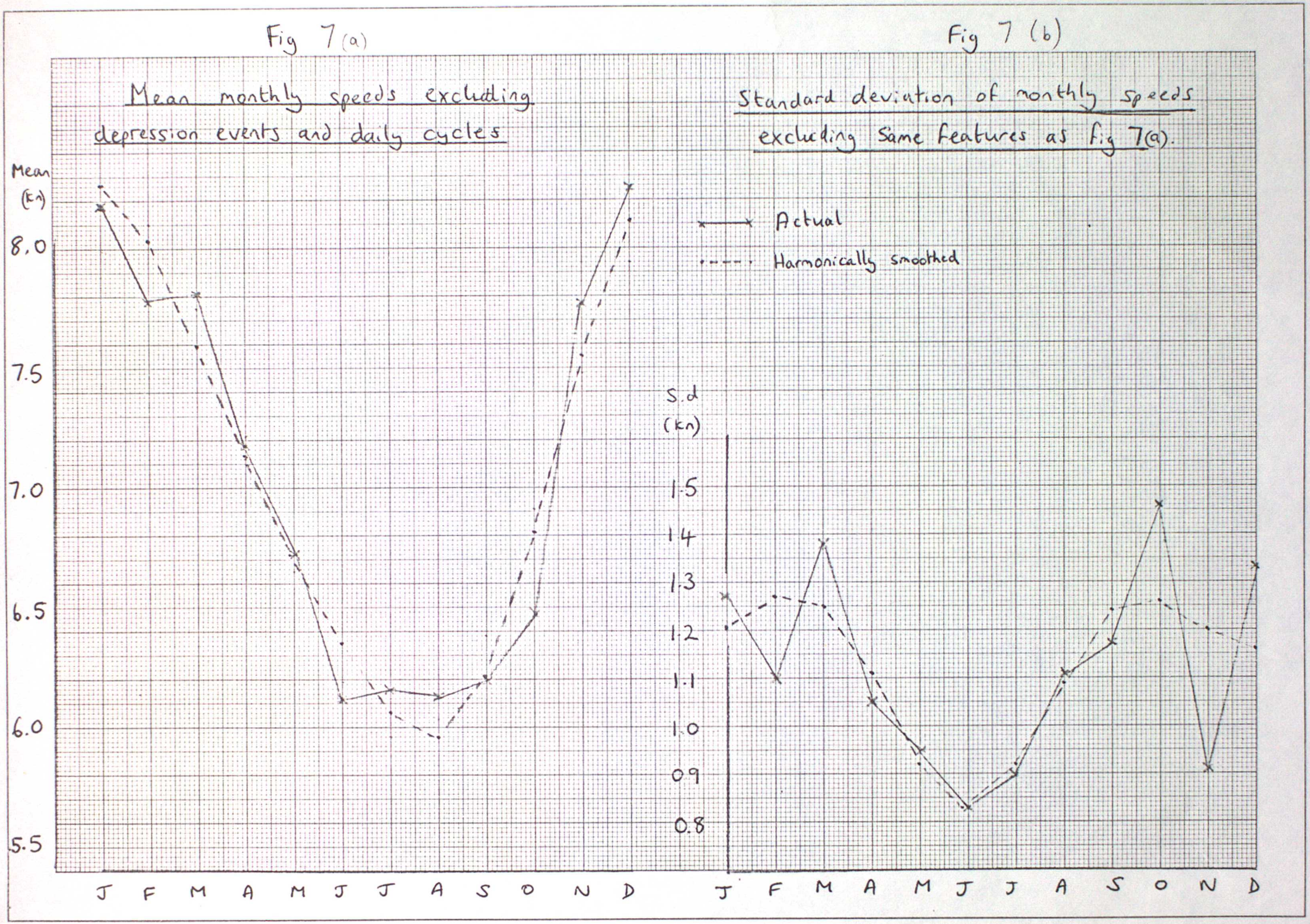


Fig 7(a)
Mean monthly speeds excluding depression events and daily cycles
Mean (kn)
8.0
7.5
7.0
6.5
6.0
5.5
J F M A M J J A S O N D
Fig 7 (b)
Standard deviation of monthly speeds excluding same features as fig 7(a).
Actual
Harmonically smoothed
S.d (kn)
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
J F M A M J J A S O N D

Comparisons between actual and simulated data.

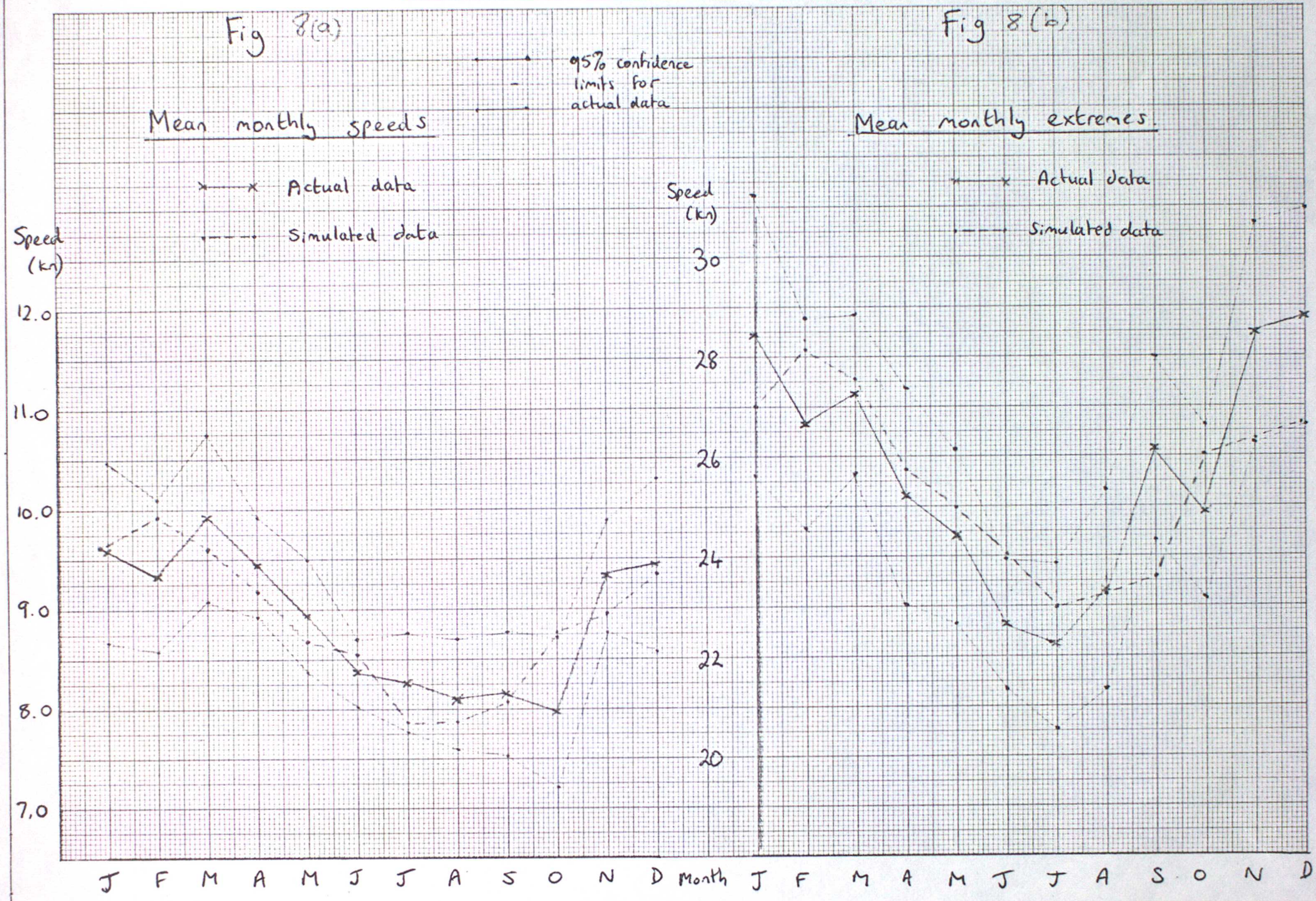


Fig 9

Cumulative percentage frequencies of speeds in specified ranges for actual and simulated data and selected months (speeds in Kn)

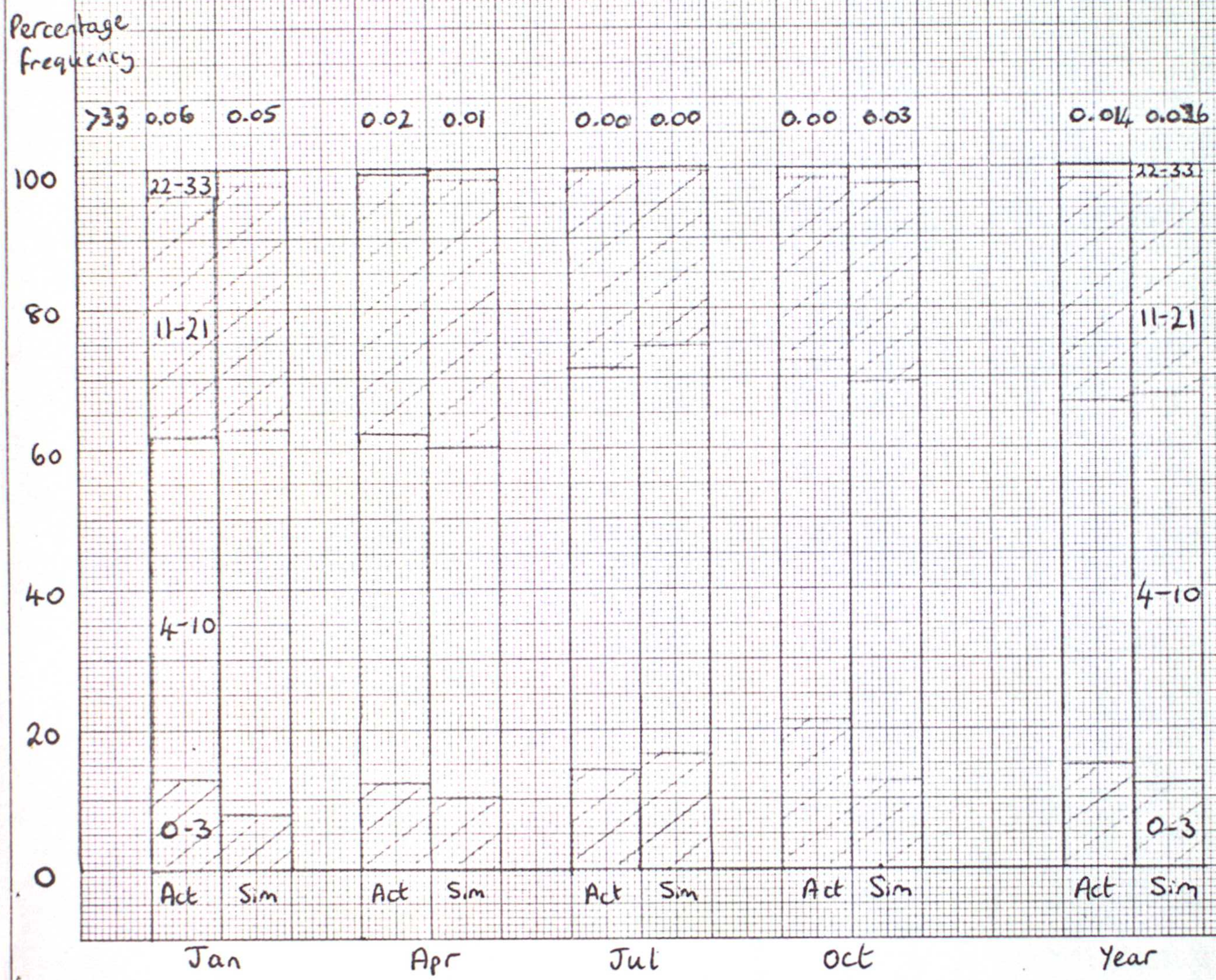


Fig 10

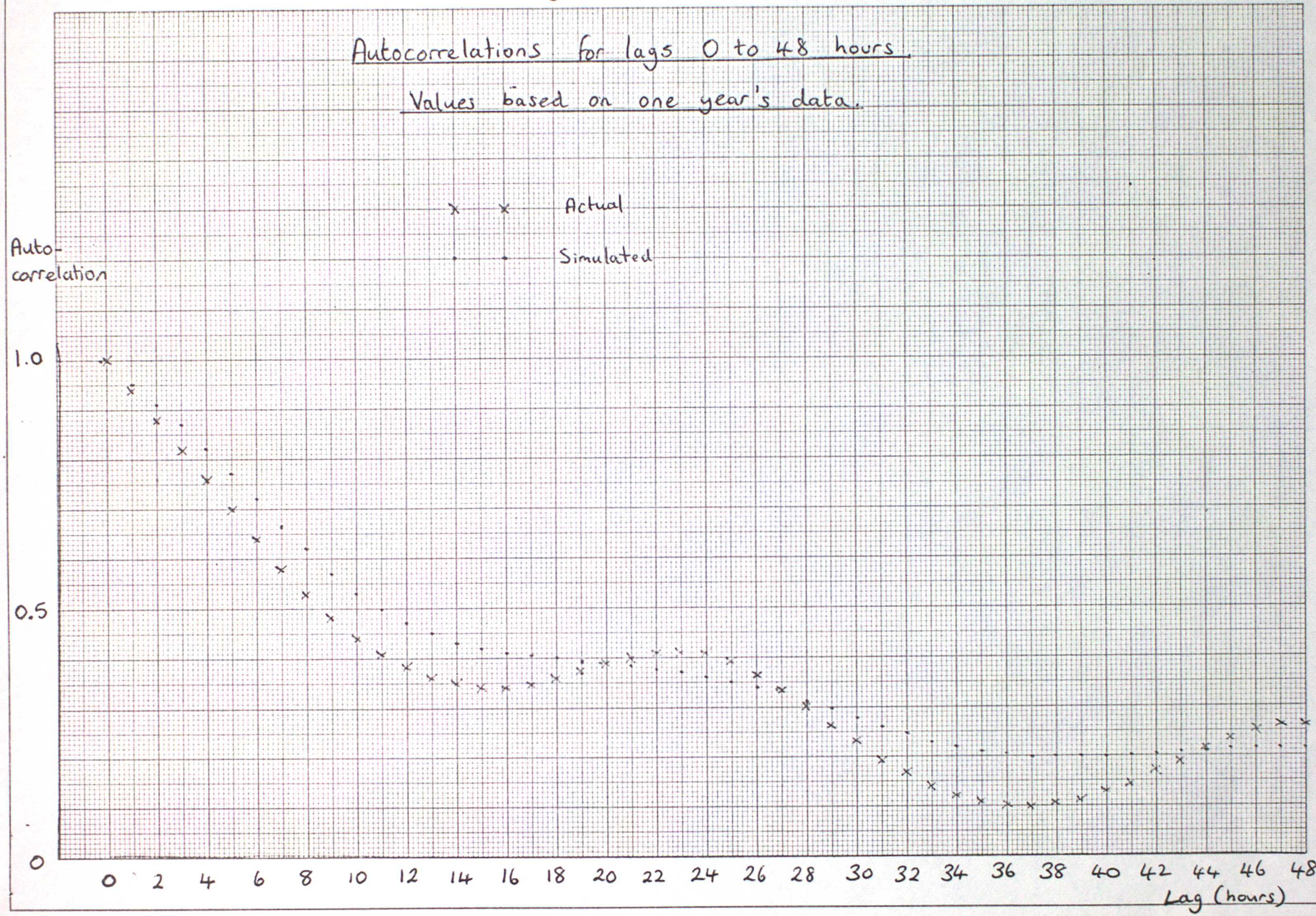

Autocorrelations for lags 0 to 48 hours.
Values based on one year's data.
Actual
Simulated
Auto-correlation
1.0
0.5
0
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
Lag (hours)

Table 1

Frequencies of depression events by year and month for Elmdon data 1961-79

i) Year

	Frequency		Frequency
1961	46	1971	22
1962	48	1972	27
1963	31	1973	36
1964	30	1974	48
1965	39	1975	28
1966	48	1976	21
1967	43	1977	25
1968	34	1978	25
1969	35	1979	29
1970	42		

ii) Month

Jan	83	Jul	22
Feb	69	Aug	35
Mar	84	Sep	35
Apr	61	Oct	42
May	44	Nov	84
Jun	21	Dec	77

Table 2a

Summary statistics for annual speeds

Annual mean	Average	Standard deviation	95% confidence limits for the average Lower	Upper	Lowest	Highest
Actual	8.90	0.55	8.64	9.16	7.97	10.18
Simulated	8.87	0.56	–	–	7.64	9.89
Annual extremes						
Actual	34.00	3.61	32.34	35.66	29	44
Simulated	33.83	4.29	–	–	27.2	44.0

Table 2b

Lowest and Highest annual extremes in each of five 19 year periods for simulated data.

	Period 1	2	3	4	5
Lowest	29	29	27	25	26
Highest	42	41	54	43	40