CHAPTER 26

APPLICATION OF LOGNORMAL TRUNCATED DISTRIBUTION TO PREDICTION OF LONG TERM SEA STATE USING VISUAL WAVE HEIGHT DATA.

Angel Gonzàlez[1] Javier Martìnez[2] Rafael Blàzquez[3]

ABSTRACT

The purpose of this paper is to analyze the behaviour of the lognormal distribution function in wave studies, when visual observations are used to predict the long term wave height distribution.

We are focused our attention on the problem of visual biased samples. For this kind of data the lognormal model does not fit accurately the sample leading to wrong extrapolations.

One possible solution to avoid this problem is to consider the sample as truncated. According to this procedure the distribution function is estimated based only on data higher than the point of truncation.

1.-INTRODUCTION

It is well known that the knowledge of wave parameters, mainly the wave height, is indispensable to carry out projects in maritime engineering, that is the reason why at the present most of the developed countries are trying to improve their measuring networks.

The very best state would be to have a great deal of recorded wave data concerning the area of each particular study, but unfortunately the fact of the matter is that in many cases instrumental data either are not available or do not exist at all, and it is necessary to use other sources of wave data.

1-CENTRO DE ESTUDIOS DE CARRETERAS,CEDEX,Madrid,Spain.
2-CENTRO DE E.DE PUERTOS Y COSTAS,CEDEX,Madrid,Spain.
3-INSTITUTO EDUARDO TORROJA,CSIC,Madrid,Spain.

Sources of wave data may be considered under the three groups: instrumental, hindcast and visual data.

Instrumental data usually come from wave buoys records and they are preferred because of their reliability and quality, but among their shortcomings are the reduced area of coverage and the limited availability which is often due to commercial restrictions.

The hindcast wave data can be obtained from wind field analysis. Wave heights are estimated from knowledge of the wind speed and fetch, covering large areas. On the other hand hindcasting methods are not very accurate.

The third source is visual data. Visual observations provide additional information about wave parameters, for instance wave direction, and also cover large sea areas. The reliability of visual data has been fully criticized for different reasons. In this paper we are focused our attention on the problem of visual biased samples.

2.-STATEMENT OF PROBLEM

Visual data usually come from ship reports,that is why an important part of them are concentrating on the main shipping routs.

These data bring up some shortcomings due to the observation itself:

-Wave heights reported in adverse climatic conditions tend to be overestimated by the observer (Jardine,1979).

-More ships sail in good weather conditions consequently samples are biased towards lower wave height values.

The result is that these samples lead to wrong extrapolations of the long term wave height distribution when the observations reported as calms are included in the sample.

3.-METHODOLOGY

3.1.-Fundamentals

Since the lognormal distribution function is by and large considered appropriate for wave studies, Jasper (1956), Darbyshire (1956), Khanna and Andru (1974), the cumulative distribution function (CDF) corresponding to the sample we are dealing with (figure 1) has been plotted in the lognormal probality paper (figure 2).

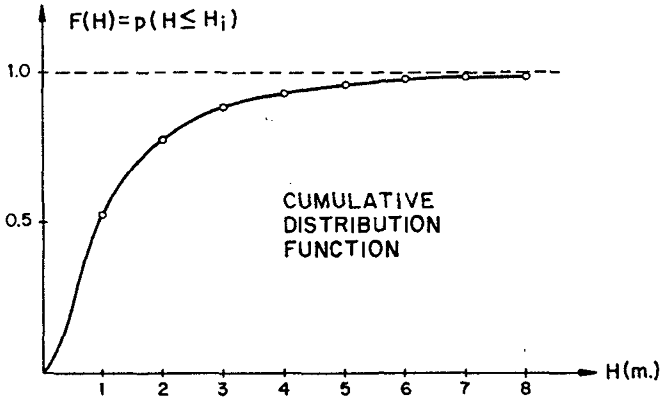


Figure 1

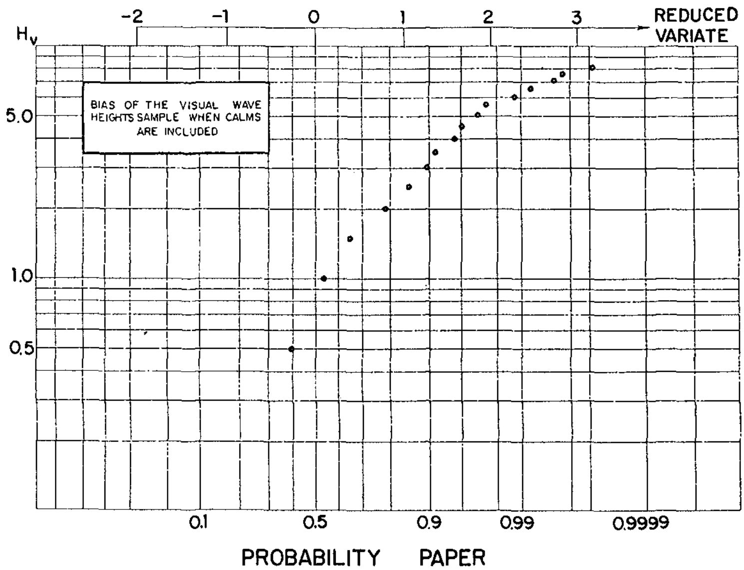


Figure 2

It can be seen that, when the calms are included in the sample, a simple straight line does not fit the points accurately enough.

One possible solution to overcome this problem is to consider the sample as truncated in order to get rid of those data causing a biased sample.

A variable may be such it appears to be lognormal except in that part of the distribution for which the values of the variable either can not occur or are not observed. The distribution of such a variable is said to be incomplete or, more commonly, truncated.

The resulting distribution might then have a shape like figure 3 shows, if the distribution of the original population was lognormal.

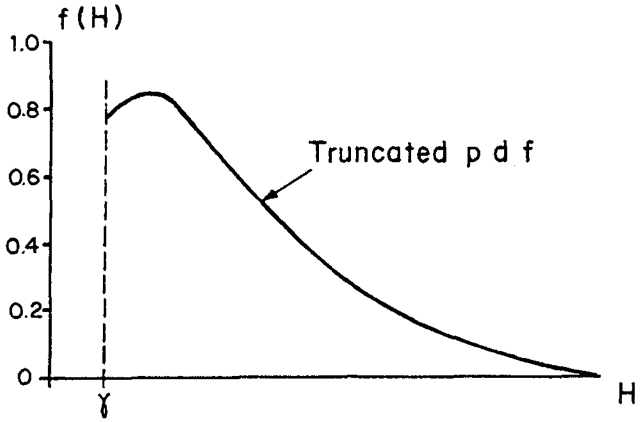


Figure 3

In this case the members of the population below γ have been eliminated. Below γ point there are no values of random variable, and their probalities are removed.

Censored distributions differ slightly in application in that the total population is present but the exact frequencies are known only up to (or beyond) a certain value.

In these cases the distribution is assumed to have the probability below (or above) γ lumped at that γ point.(See figure 4).

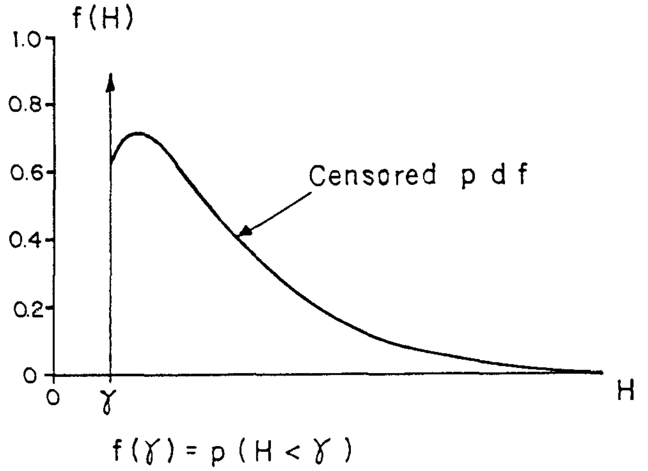


Figure 4

The distribution between truncation and censorship arises from the fact that in the first the available information is confined to the range (γ ,∞) , whereas in the second a limited knowledge of the variable in the range (0, γ) permits consideration of the complete range (0,∞), (Benjamin, 1981).

In practice truncated samples arise with several types of experimental data in which recorded measurements are available over only a partial range of the variable.

In our case the random variable is the visual wave height and the point of truncation does not physically exist. Obviously there are wave heights below the point of truncation and the low wave heights , even calms, occur in Nature.

Strictly speaking we are dealing with a lower bound censored distribution in which probabilities below a certain value are unreliable. However, if the distribution is considered as truncated and we assume that the point of truncation is known, the problem becomes mathematically tractable.

So, a correct value for the point of truncation has to be chosen and the distribution function will be estimated based only on data higher than the point of truncation.

3.2.-Procedural steps

Figure 5 shows the shape of both untruncated and truncated lognormal density functions.

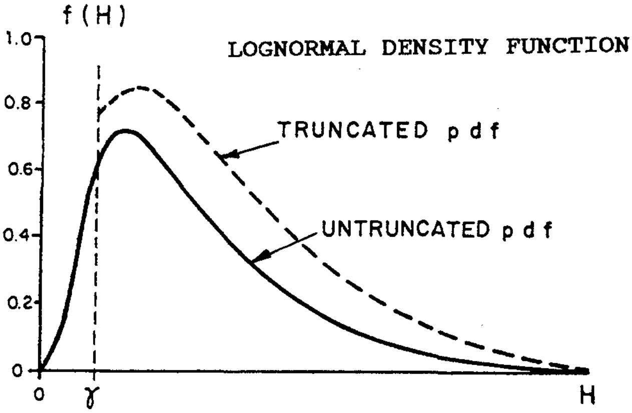


Figure 5

The truncated probability density function only exists above γ point, and obviously its area must equal one, so the PDF is zero up to γ and K times the untruncated probability density function above that point.

Thus, if the original population has a lognormal PDF :

$$f(x) = \frac{1}{x\, B \sqrt{2\pi}}\, e^{-1/2 \left(\frac{\ln x - A}{B} \right)^2}$$

then, the truncated PDF may be written as

$$ft(x) = 0 \qquad , \ x < \gamma$$

$$ft(x) = K\, f(x) \qquad , \ x >= \gamma$$

The probability axiom

$$\int_{\gamma}^{\infty} f_t(x)\, dx = 1$$

leads to

$$K = \frac{1}{1 - F(\gamma)}$$

where $F(\gamma)$ is de CDF value at γ point.

Although in principle it would be possible to apply the method of moments to estimate A and B parameters there seems to be no easy way of solving the resulting equations for the first and second sample moments.

Therefore the method of maximun likelihood is a good choice to obtain the estimators of those parameters.

A technique frequently appropriate in maximun likelihood estimation problems is to find the maximun of the loglikelihood function. If the sample is $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ then the likelihood function of the sample is

$$V = f(x_1)\, f(x_2) \,\ldots\ldots\, f(x_n) =$$

$$= \prod_{i=1}^{n} \frac{\dfrac{1}{x_i B\sqrt{2\pi}}\, e^{-1/2\left(\frac{\ln x_i - A}{B}\right)^2}}{1 - F(\gamma)}$$

and so the loglikelihood function may be written

$$L = \ln V =$$

$$= \sum \ln \left[\frac{1}{x_i B\sqrt{2\pi}}\, e^{-1/2\left(\frac{\ln x_i - A}{B}\right)^2} \frac{1}{1 - F(\gamma)} \right]$$

The derivatives with respect to the parameters when set equal to zero, permit to obtain A and B by means of a function which was tabulated by Hald, (Aitchison and Brown, 1957).

4.-APPLICATION OF PROPOSED METHODOLOGY

The application of the truncation method has been carried out using a sample of visual data from the National Climatic Center (Asheville, Norht Carolina, USA).

The sample concerns the Spanish northeren coast , in the Bay of Biscay near Bilbao, exactly from 43.3° to 44.0° North and from 2.5° to 3.5° West.

Figure 6 shows the exact location where the sample was obtained.

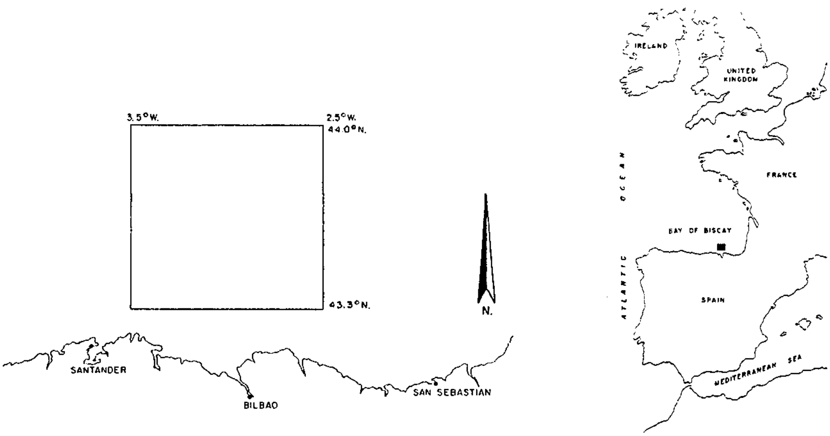
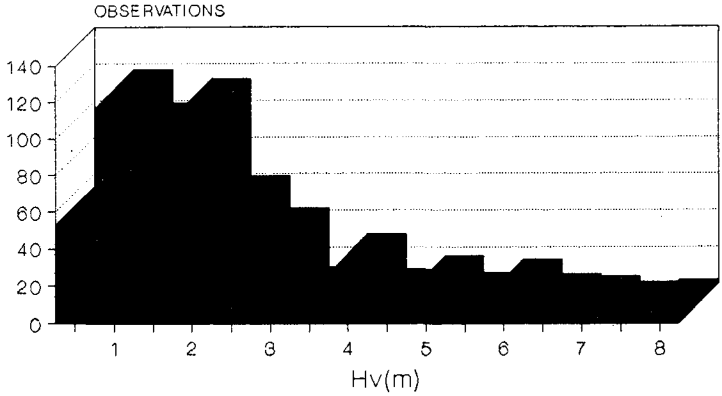


Figure 6

This sample has 572 valid observations and 277 calms. In the histogram (figure 7) the calms are not included and it can be seen how it takes a lognormal distribution shape.



CALMS : 277

VALID OBSERVATIONS : 572

Figure 7

The lowest range of the variable has 53 observations, when the calms are not included, and its probability is almost 0.1, exactly 0.09.

Subsequently the calms have been included in the sample, and the cumulative frequencies have been plotted in the lognormal probability paper (Figure 8).

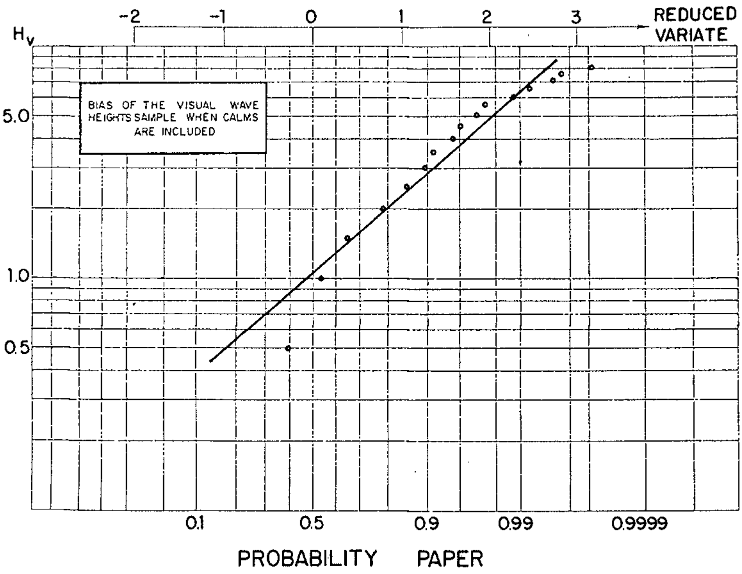


Figure 8

This drawing shows that the probability of the lowest range of the variable is almost 0.4 whereas it hardly was 0.1 when the calms were not included.

At this moment we should consider if the sample is either biased or unbiased. In the second case, there is no problem to include the calms in the lowest range of the variable, but if the sample is biased, as it happens in our present case, a standard fitting of cumulative distribution shows that the straigh line does not fit the data points very accurately (see figure 8).

The methodology has been carried out by choosing two different points of truncation, γ = 0.1 m and γ = 0.5 m.

In the figure 9 we can see how both truncated distributions are very close whereas the untruncated is quite different.

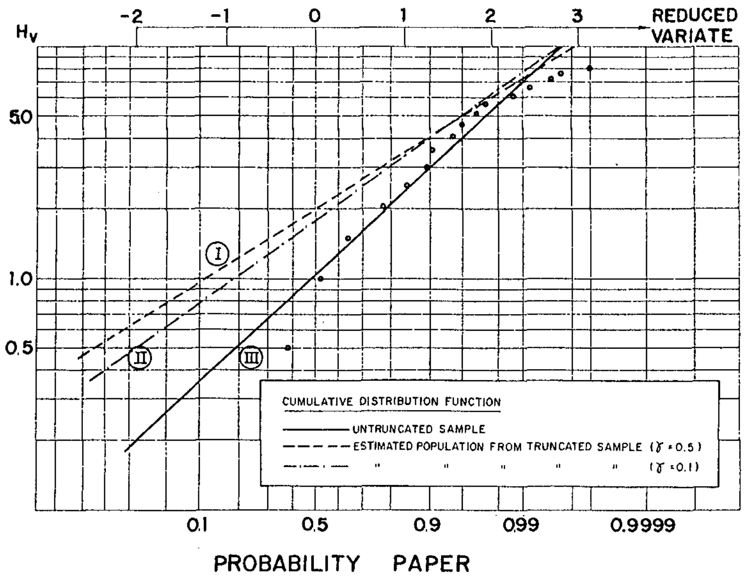


Figure 9

The prediction given by then show how the truncated distributions are in good agreement. Table I represents wave heights versus cumulative probabilities.

TABLE I

Hv(m)	Probability		
	I	II	III
1.5	0.33	0.41	0.67
2.0	0.52	0.59	0.79
2.5	0.67	0.72	0.86
3.0	0.90	0.91	0.92
	good agreement		

5.-GOODNESS OF THE METHOD

In spite of the fact that instrumental data are unbiased, a sample collected by a wave buoy has been truncated at the point γ = 0.5 m in order to test the goodness of the truncated distribution.

So, all the data up to 0.5 m have been removed and the truncation method has been applied.

Thus, we have two populations, the original untruncated sample and the truncated sample in which data below o.5 m have been removed.

The result is that both, truncated and untruncated CDF`s are very close.(See figure 10).

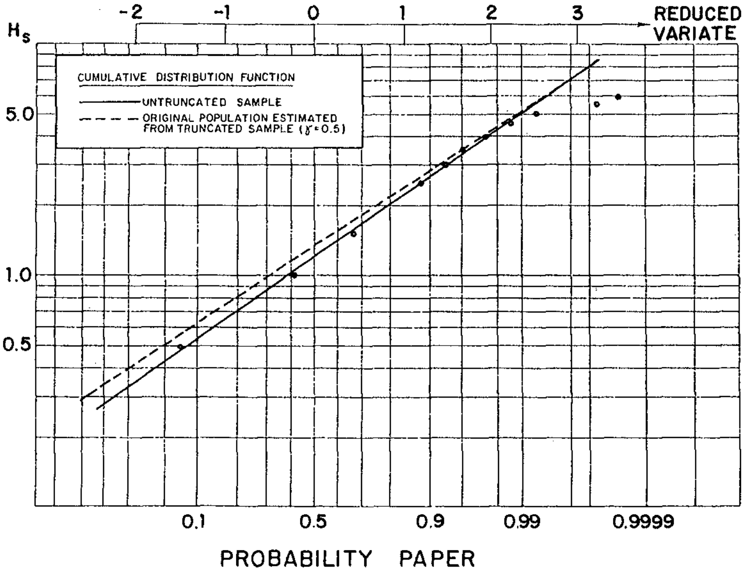


Figure 10

Consequently both of then lead to similar predictions (see Table II).

TABLE II

Hv(m)	Probability	
	Truncated	Untruncated
1.5	0.57	0.64
2.0	0.75	0.80
2.5	0.85	0.88
3.0	0.92	0.93

6.-CONCLUSIONS

1.-In many cases visual wave samples are biased towards lower height values.

2.-The lognormal model, which is usually appropriate for wave studies, does not fit accurately the sample when the calms are included.

3.-For visual wave data the truncation method improves current procedures and avoids wrong extrapolations from observations below the point of truncation.

4.-For instrumental wave data predictions of CDF`s of wave heights of both, truncated and untruncated samples, are in good agreement.

5.-The method reported herein is consistent and permits a more effective use of available visual wave data for coastal and harbour engineering projects.

REFERENCES

AITCHISON, J. and BROWN, J.A.C. (1957), "The lognormal Distribution", Cambridge University Press, London.

BENJAMIN, J.R. and CORNELL, C.A. (1970), "Probability Statistics and Decision for Civil Engineers", Mc Graw-Hill, New York.

DARBYSHIRE,J.(1956),"The Distribution of Wave Heights" ,Dock and Harbour Authority, May.

JARDINE, T.P. (1979), "The Reliability of Visually Observed Wave Heights", Coastal Engineering, Vol.3,pp. 33-38.

JASPER, N.H.(1956), "Statistical Distribution Patterns of the Ocean Waves and of Wave-Induced Ship Stresses and Motions, with Engineering Applications", Soc .Nav. Archit. Mar. Eng. Trans, 64.

HALD, A. (1949), "Maximun Likelihood Estimation of the Parameters of a Normal Distribution which is Truncated at a Known Point", Skand. Aktnar.Tidskr.32,119.

KHANNA, J. and ANDRU, P. (1974), "Lifetime Wave Height Curve for Saint John Deep, Canada", Proc. Ocean Wave Meas. Anal. 1. pp 301-319.