SANDIA REPORT

SAND99–2985
Unlimited Release
Printed November 1999

Moment-Based Probability Modelling and Extreme Response Estimation
The FITS Routine, Version 1.2

Lance Manuel, Tina Kashef, and Steven R. Winterstein

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of
Energy under Contract DE-AC04-94AL85000.

Sandia National Laboratories

This report was printed exactly as received from Stanford University.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: http://www.ntis.gov/ordering.htm

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01

SAND99-2985
Unlimited Release
Printed November 1999

Moment-Based Probability Modelling and Extreme Response Estimation The FITS Routine

Version 1.2

Lance Manuel, Tina Kashef and Steven R. Winterstein
Civil Engineering Department
Stanford University
Stanford, CA 94305

ABSTRACT

This report documents the use of the FITS routine, which provides automated fits of various analytical, commonly used probability models from input data. It is intended to complement the previously distributed FITTING routine documented in RMS Report 14 (Winterstein et al., 1994), which implements relatively complex four-moment distribution models whose parameters are fit with numerical optimization routines. Although these four-moment fits can be quite useful and faithful to the observed data, their complexity can make them difficult to automate within standard fitting algorithms. In contrast, FITS provides more robust (lower moment) fits of simpler, more conventional distribution forms. For each database of interest, the routine estimates the distribution of annual maximum response based on the data values and the duration, T, over which they were recorded. To focus on the upper tails of interest, the user can also supply an arbitrary lower-bound threshold, χ_{low}, above which a shifted distribution model—exponential or Weibull—is fit. (In estimating the annual maximum response, the program automatically adjusts for the decreasing rate of response events as the threshold χ_{low} is raised.)

Version History

Version 1.0: The first version of the FITS routine. This version was documented in an earlier report (Stanford RMS Report 19; Winterstein, 1995)

Version 1.1: This version added the "quadratic Weibull" distribution to the original library. This distribution is fitted to the first three moments of a data set, and has been found especially useful in modelling fatigue loads on wind turbine blades (see Stanford RMS Report 31; Kashef and Winterstein, 1998).

Version 1.2: The current version of the FITS software supports more compact formats for the input data. In addition to the original format which required raw data one entry per response "event," two other database formats are now accepted by the software. Data can be input in a binned or histogram format wherein bin centers and the number of occurrences in each bin are read as input. Another alternative is to input the first four moments (mean, standard deviation, skewness, and kurtosis) of the data and the number of events for the duration of the database.

Contents

1 Distribution Fitting: General Methodology

This report documents the usage of the routine FITS, which provides automated fits of various analytical, commonly used probability models to input data. This routine is intended to complement the previously distributed routine, FITTING, documented in RMS Report 14 (Winterstein et al, 1994). The FITTING routine implements relatively complex, four-moment distribution models, whose parameters are fit with numerical optimization routines. While these four-moment fits can be quite useful and faithful to the observed data, their complexity can make them difficult to automate within standard fitting algorithms.

In contrast, the routine FITS is intended to provide more robust (lower moment) fits of simpler, more conventional distribution forms. For each database of interest, the routine estimates the distribution of annual maximum response, based on the data values and the duration, T, over which they were recorded. It can also estimate the annual maximum response distribution due to multiple types of response "events," intended to reflect different statistical populations. Examples include waves or winds from different directions, or response events due to hurricanes, eddy currents, or combined hurricane-current events.

Specifically, FITS first estimates the cumulative distribution functions for each individual response, $F_i(x)$=P[Outcome $< x$ in database i], from the data in that database. It then predicts the corresponding distribution of maximum response X_{max}, over a target duration of time T, by assuming that events from each database occur in independent, Poisson fashion[1]:

[1] In this document we use the notation G_{ann} to represent the probability of exceedence of the response over a given duration of time. However, this target duration need not neccessarily be one year. The exceedence probability will be calculated for unit time, the units being consistent with those used in the input file.

$$P[\text{ Max response in time } T > x] = G_{ann}(x) = 1 - \exp[-\sum_i \nu_i T G_i(x)] \quad (1)$$

in which

$$G_i(x) = 1 - F_i(x) \quad (2)$$

Here ν_i is the mean rate of events per unit time in database i, estimated from the observed number (possibly over a user-defined threshold; see below).

1.1 Available Distribution Types

Specific distributions currently included in `FITS` to estimate $F_i(x)$ include the following, as catalogued by the distribution index `IDIST`:

IDIST=1: Normal Distribution

IDIST=2: Lognormal Distribution

IDIST=3: Exponential Distribution

IDIST=4: Weibull Distribution

IDIST=5: Gumbel Distribution

IDIST=6: Shifted Exponential Distribution

IDIST=7: Shifted Weibull Distribution

IDIST=8: Quadratic Weibull Distribution

IDIST=9: Shifted Quadratic Weibull Distribution

The distributions `IDIST`=1 through 5 and 8 are all fit to statistical moments of all available data. The single-parameter exponential preserves only the mean m_x of the data, while the normal, lognormal, Weibull, and Gumbel preserve both the mean and standard deviation σ_x estimated from the data.

The quadratic Weibull preserves the first three moments of the data (mean, standard deviation, and skewness). It is therefore the most general among the distributions included here. Note that this quadratic Weibull model was not included in the original version of `FITS` (Winterstein, 1995). Its inclusion in a subsequent version of `FITS` (Kashef and Winterstein, 1998) as well as in the current version reflects its successful applications, particularly in modelling fatigue loads on wind turbine blades (Lange and Winterstein, 1996). At the same time, unlike the four-moment models from `FITTING` (Winterstein et al, 1994), the parameter fitting of the quadratic Weibull does not require numerical optimization. It also avoids the tendency of four-moment models toward overfitting, when applied to positive variables such as load peaks or ranges. This is demonstrated here through an additional example in a later section, which applies `FITS` to predict extremes from a data set of observed wind turbine blade loads. Results offer a comparison between the Weibull and quadratic Weibull models, as well as with the four-moment "cubic Weibull' model available from the `FITTING` routine.

Most of the one-sided distributions above (exponential, Weibull, and quadratic Weibull) are also generalized here by imposing an arbitrary shift (IDIST=6, 7, and 9). These permit a user-defined lower threshold x_{low}, ignore data below x_{low}, and fit standard exponential/Weibull/quadratic Weibull models to $x - x_{low}$ based on observed moments. These are perhaps the most relevant distributions when modelling local peaks, Y, which generally have a broadly skewed distribution away from a well-defined lower bound. (In estimating the annual maximum response, the program automatically adjusts for the decreasing rate of response events as the threshold x_{low} is raised.)

The result aims to provide the user with a suite of smooth probability models, to be fit throughout the body of the available data. It does not directly address various special topics of data fitting; e.g., selective tail fitting, fitting bimodal models to hybrid data, etc. Some of these issues can be addressed, in a limited way, through the use here of the *shifted* models (IDIST=6, 7, and 9). In this way the user can focus the distribution modelling resources on the extreme response levels of interest.

More specific tail-fitting procedures have not been given here, because optimal use of these tends to be rather problem-specific. In the same vein our extremal models are limited here to so-called "Type I" behavior, leading to (shifted) exponential distributions of peaks over a given threshold and to Gumbel distributions of annual maxima. Type II and III distributions are ill-suited to our moment-fits, due to potential moment divergence (Type II) or to the difficulty in predicting truncated distributions (Type III) from moment information.

1.2 Software Limitations

Several parameters have been assigned maximum values in the routine FITS. These include the upper limits

- NFMAX, the maximum number of files (databases) to be fit, has been set to 10.
- NMAX, the maximum number of data per database, has been set to 32768.

Both of these limits have been set in PARAMETER statements in the main driver program to FITS. These are rather arbitrarily selected limits, and can be reset by the user without fundamental consequence.

2 Distribution Fitting: Routines

The fitting algorithm calls the following set of subroutines:

CALMOM: Estimates the mean m_x, standard deviation σ_x, skewness α_3 and kurtosis α_4 from an input set of data. These are based on unbiased estimates of the cumulants $k_1{=}m_x$, $k_2{=}\sigma_x^2$, $k_3{=}\alpha_3\sigma_x^3$, and $k_4{=}(\alpha_4{-}3)\sigma_x^4$. If the user includes an optional lower limit x_{low}, moments of the shifted variable $(x - x_{low})^+{=}\max(0, x - x_{low})$ are estimated.

DISPAR: Based on the sample moments estimated in CALMOM, DISPAR seeks a consistent set of distribution parameters. The interpretation of these parameters depends on the distribution type selected by the user. Appendix A includes a complete listing of the distribution functions and their parameters.

GETCDF: For the user-defined distribution type with the distribution parameters from DISPAR, this routines estimates the cumulative distribution function value, $F(x)$=P[Outcome $< x$] for given input x value.

FRACTL: For the user-defined distribution type with the distribution parameters from DISPAR, this routines estimates the fractile x corresponding to a specified input value of the probability $p{=}F(x)$=P[Outcome $< x$].

QDMOM: Uses Gaussian quadrature to estimate the first four moments of the theoretical fitted distribution. These can be compared with the sample moments from the data, as given by CALMOM, to verify the accuracy of the fitted model—and in the case of the higher moments not used in the original fitting, to test its goodness of fit.

The routines GETCDF and FRACTL, which supply general distribution functions and their inverses, may also be useful in other stand-alone applications; e.g., to create a distribution library for standard FORM/SORM or simulation analyses (Madsen et al, 1986), or for use with new Inverse FORM algorithms (Ude and Winterstein, 1996).

3 Input Format and Wave Height Example

3.1 Data Input

The file(s) containing data are read in free format, one datum per line. Non-numerical input are taken as comments and ignored. The first numerical value found is taken to be the duration of the database[2]. Remaining values (1 per line) are interpreted as data, and data are read until the end of the file is encountered.

We illustrate the input here through a simple example, involving significant wave height data. These are in fact 19 annual maximum H_S values, estimated by hindcast in a Southern North Sea location (Winterstein and Haver, 1991). These data are stored in appropriate format in the file `gumbel.dat`.

The contents of `gumbel.dat` are listed below. The first line reflects that 19 years are covered, and the remaining 19 lines contain the actual maximum H_S encountered in each of these 19 years. (In this input file the data are given in descending order; this is not required by the program.)

```
19.
9.66
9.44
9.18
9.17
8.85
8.79
8.60
8.58
8.54
8.49
```

[2]The units for the duration are chosen by the user. The exceedence probability in unit time G_{ann} calculated by the program will use the same units. For example, in order to obtain annual exceedence probabilities the duration must be input in units of years.

```
8.09
8.08
8.06
7.47
7.42
7.41
7.31
7.16
6.92
```

Contents of gumbel.dat.

3.2 Runtime Input: Batch Mode

We seek to invoke FITS under the following conditions:

1. Results to be written to a file named gumbel1.out. (The distinction between lower- and upper-case letters in filenames is honored by Unix, and ignored in DOS.)
2. Distribution results are to be written for x (wave height) values ranging from XMIN=5.0 to XMAX=20.0 m, at increments of DX=0.5 m.
3. The extreme in target duration T=1 (yr) is to be fit. (The T value here conforms to that defined in Eq. 1).
4. Each database will be in raw data format (other options for reading databases in binned/histogram format or as moments are described in Section 6).
5. There is only NFILES=1 database to be fit.
6. The wave height data are stored in a file named gumbel.dat.
7. The user desires to fit a Gumbel distribution (IDIST=5) to these data.

In this example (and for any other in which only 1 database is to be fit), FITS requires 7 lines of input, corresponding to the information given in the 7 items above. In this case the 7 input lines are as follows:

```
gumbel1.out             ; Input line 1: File where output is to be written
5.0  20.0  0.5          ; Input line 2: xmin, xmax, dx for writing dist results
1                       ; Input line 3: target lifetime for calculating Pc
0                       ; Input line 4: iprint = 0,1,2 (raw,histograms,moments)
1                       ; Input line 5: number of databases (input files)
gumbel.dat              ; Input line 6: name of 1st database (input file)
5                       ; Input line 7: idist = dist type (idist=5 -> Gumbel)
```

(the ";" and following information is not read by FITS; these are given here to remind the user of the input definition). These 7 lines have been stored in a file named gumbel1.in. Thus, to execute FITS in a batch mode, the user can type (in either Unix or DOS environment) the following command:

```
fits < gumbel1.in
```

This will create (or overwrite) a corresponding file gumbel1.out, whose contents is discussed in the next section. While FITS is executing, the user will see prompts for terminal inputs; these can be simply ignored in this batch mode operation.

3.3 Runtime Input: Interactive Mode

If the user only types "FITS" without specifying an input file, he will be prompted for each input (the same 7 lines in this case) with interactive explanations. For example, before requesting the distribution index IDIST the program lists all available distribution types and associated IDIST values. Inputs with invalid formats are ignored. This interactive mode may be particularly useful for first-time users. (The text with input prompts is written to the logical unit IOERR, which is set to 0 in the driver program for FITS. The user can reset this if necessary.)

The following is a screen dump of the terminal input prompts and user's response. Lines beginning with ">" are input prompts generated by the program. Other lines are the user's response; in this case there are precisely 7 lines of response, as in the batch mode input file given previously.

```
>
> ** ENTER FILENAME WHERE OUTPUT WILL BE WRITTEN **
>
>        ENTER OUTPUT FILENAME:
gumbel1.out
>
> ** ENTER XMIN  = MIN X VALUE AT WHICH TO OUTPUT CDF
>          XMAX  = MAX X VALUE AT WHICH TO OUTPUT CDF
>            DX  = INTERVAL OF X VALS WHERE CDF IS OUTPUT
>    ALL THREE VALUES ON SAME LINE; E.G.
>
>         0.5 10.0 0.5
>
>    GIVES OUTPUT AT 20 X VALUES FROM 0.5 TO 10.0
>
>        ENTER XMIN,XMAX,DX:
5.0  20.0  0.5
>
> ** ENTER TARGET=TARGET LIFETIME FOR CALCULATING
>    PROBABILITY OF EXCEEDENCE
```

```
>
>       ENTER TARGET:
1
>
> ** ENTER IPRINT=INTEGER VALUE (FLAG) FOR TYPE OF
>    INPUT DATA FILE FORMAT
>    IPRINT = 0 [Raw Data], 1 [Binned Data], 2 [Moments]
>
>       ENTER IPRINT:
0
>
> ** ENTER NFILES=NUMBER OF DATA SET FILES TO BE FIT
>    CURRENT OPTIONS: NFILES= 1  THROUGH  10
>
>       ENTER NFILES:
1
>
>          INPUT FILE NUMBER:   1
> ** ENTER FILENAME WHERE DATA ARE STORED,
>
>       ENTER INPUT FILENAME:
gumbel.dat
>
> ** ENTER IDIST =INDEX OF DISTRIBUTION TYPE TO BE FIT
>    CURRENT OPTIONS:
>                        IDIST =  1 ...  NORMAL
>                        IDIST =  2 ...  LOGNORMAL
>                        IDIST =  3 ...  EXPONENTIAL
>                        IDIST =  4 ...  WEIBULL
>                        IDIST =  5 ...  GUMBEL
>                        IDIST =  6 ...  SHIFTED EXPONENTIAL
>                        IDIST =  7 ...  SHIFTED WEIBULL
>                        IDIST =  8 ...  QUADRATIC WEIBULL
>                        IDIST =  9 ...  SHIFTED QUADRATIC W
>
>       ENTER IDIST:
5
```

4 Output Format and Extreme Wave Height Example

The first output section of FITS provides summary statistics for each of the data files considered. These include (1) sample moments from the data, (2) predicted moments from the fitted distribution, and (3) underlying distribution parameters. Comparison of (1) and (2) can serve to verify the fit of low-order moments, and the agreement of higher moments (e.g., skewness and kurtosis) not used in the fitting can offer a rough goodness-of-fit test.

The second section of output gives distribution estimates for $G_i(x)$, the probability that a future outcome exceeds x as estimated from datafile i. It also reports the distribution of annual maxima[3], $G_{ann}(x)$, as estimated from Eq. 1. Note that all preceding output lines begin with "#", which is interpreted as a comment within the public-domain gnuplot plotting package. Thus the output file can be plotted directly with gnuplot.

The contents of the gumbel1.out output file are shown at the end of this chapter. The output confirms that, as intended, the fitted Gumbel model preserves the mean m_x=8.275m and standard deviation σ_x=0.816m estimated from the data. The Gumbel model tends to overestimate the higher moments, however; its predicted skewness 1.140 and kurtosis 5.40 varies notably from the values -0.053 and 1.91 estimated from the data.

This suggests that the Gumbel model may somewhat overestimate the chance of large wave heights in this case. This has also been observed previously for this data set (Winterstein and Haver, 1991), where a cubic distortion of the Gumbel model was introduced to capture this trend. Here we seek to model this effect by tail-fitting, within this population of annual maxima. Specifically, we also apply FITS to the same data with a shifted exponential model (IDIST=6). Two cases are considered, corresponding to lower-bounds of x_{low}=8.0m and 8.5m. The input files show the additional input line required for shifted distributions, while the output (at least for x_{low}=8.5m) shows a somewhat reduced 100-year wave height h (where

[3]See footnote in Chapter 1

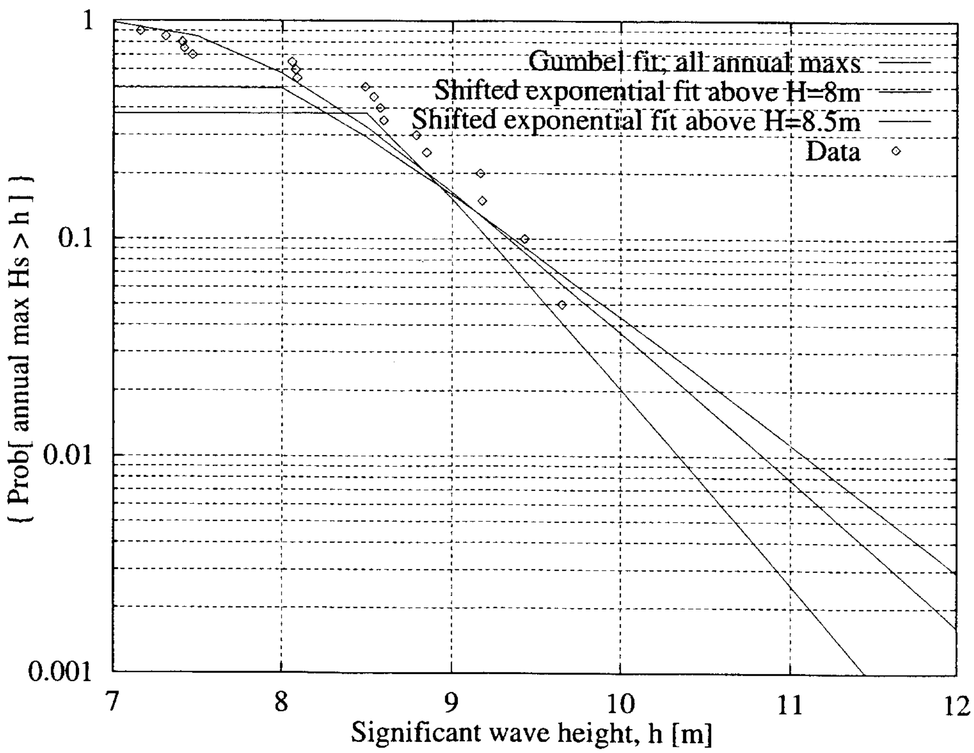

Figure 1: Estimating Annual Max H_S distribution.

P[annual max $H_s > h$] = 0.01) from 10.8m in the Gumbel case to 10.4m. These distribution fits are shown in Figure 1. They suggest that these shifted models can begin to capture trends in distribution tails that are not available in moment-fits to the entire population.

A listing of `gumbel1.out` follows:

```
#
# RESULTS FOR FILE NUMBER:      1
#             FILE NAME  : gumbel.dat
#    DURATION OF DATABASE:    19.00
#      DIST TYPE SELECTED:   GUMBEL
#     NUMBER OF DATA USED:     19
#
#
#    MOMENTS FROM SAMPLE DATA    ( MEAN, SIGMA, SKEWNESS, KURTOSIS)
#   0.8275E+01  0.8186E+00 -0.5282E-01  0.1905E+01
#
#
#    MOMENTS FROM FITTED DIST    ( MEAN, SIGMA, SKEWNESS, KURTOSIS)
#   0.8275E+01  0.8186E+00  0.1140E+01  0.5400E+01
#
#
#    DISTRIBUTION PARAMETERS     (SEE DOCUMENTATION FOR DEFINITION)
#   0.8275E+01  0.8186E+00  0.1567E+01  0.7906E+01  0.0000E+00
#
#
# ** FITTED DISTRIBUTION RESULTS **
#
# P1 = Prob {Outcome > x} in database 1,
# P2 = Prob {Outcome > x} in database 2,
# ...
# Pc = Prob {Ann max > x} including all databases
#
#      X          P1   ....      Pc
  0.5000E+01  0.1000E+01  0.6321E+00
```

0.5500E+01	0.1000E+01	0.6321E+00
0.6000E+01	0.1000E+01	0.6321E+00
0.6500E+01	0.9999E+00	0.6321E+00
0.7000E+01	0.9840E+00	0.6262E+00
0.7500E+01	0.8489E+00	0.5721E+00
0.8000E+01	0.5783E+00	0.4392E+00
0.8500E+01	0.3260E+00	0.2782E+00
0.9000E+01	0.1649E+00	0.1520E+00
0.9500E+01	0.7905E-01	0.7600E-01
0.1000E+02	0.3692E-01	0.3625E-01
0.1050E+02	0.1704E-01	0.1690E-01
0.1100E+02	0.7822E-02	0.7791E-02
0.1150E+02	0.3581E-02	0.3575E-02
0.1200E+02	0.1638E-02	0.1636E-02
0.1250E+02	0.7486E-03	0.7483E-03
0.1300E+02	0.3421E-03	0.3420E-03
0.1350E+02	0.1563E-03	0.1563E-03
0.1400E+02	0.7141E-04	0.7141E-04
0.1450E+02	0.3263E-04	0.3263E-04
0.1500E+02	0.1491E-04	0.1491E-04
0.1550E+02	0.6810E-05	0.6810E-05
0.1600E+02	0.3111E-05	0.3111E-05
0.1650E+02	0.1421E-05	0.1421E-05
0.1700E+02	0.6494E-06	0.6494E-06
0.1750E+02	0.2967E-06	0.2967E-06
0.1800E+02	0.1356E-06	0.1356E-06
0.1850E+02	0.6193E-07	0.6193E-07
0.1900E+02	0.2830E-07	0.2829E-07
0.1950E+02	0.1293E-07	0.1293E-07
0.2000E+02	0.5906E-08	0.5906E-08

5 Wind Turbine Load Example

5.1 The Data

For rotating machine components such as wind turbine blades, fatigue is generally a source of concern. While FITS is most directly aimed at frequency rates of extreme overloads, its statistical characterization of load occurrence rates can also be used to estimate statistics of damage and hence fatigue life (Winterstein and Lange, 1996). In particular, the quadratic Weibull model, which is unique to this version of FITS, has been found promising in wind turbine load characterization for LRFD design against fatigue (Lange and Winterstein, 1996).

We consider here a data set of observed bending stresses, measured on a specific horizontal-axis wind turbine (HAWT). Only the "flapwise" bending mode (out of the plane of rotation) is considered here. The data represent a total of roughly 4 hours duration, constructed from pooling various shorter intervals during which wind conditions were similar. Further, since fatigue is the prime concern, the stress time histories were rainflow-counted to achieve a set of stress range amplitudes, which—in histogram form—comprise the basis of our data set.

Figure 2 shows such a histogram of these "normalized" stress amplitudes. (The units of these data are rather artificial, hence results should be interpreted in only a relative way across different cases and models.) As is typical of such cases, there are many thousands of small-amplitude, high-frequency cycles.

Note, however, that fatigue damage is commonly assumed to be proportional to some power of S—e.g., S^b where b is on the order of 3–6 for typical metals. Hence, these small-amplitude cycles are of little fatigue consequence. To focus modelling attention and data resources, we are therefore led to consider only stresses above a fixed threshold, σ_0. Figures 3 and 4 show such histograms for σ_0=54.2 and 65.1, respectively. While a general trend of exponential-like decay remains, focusing on higher thresholds permits greater

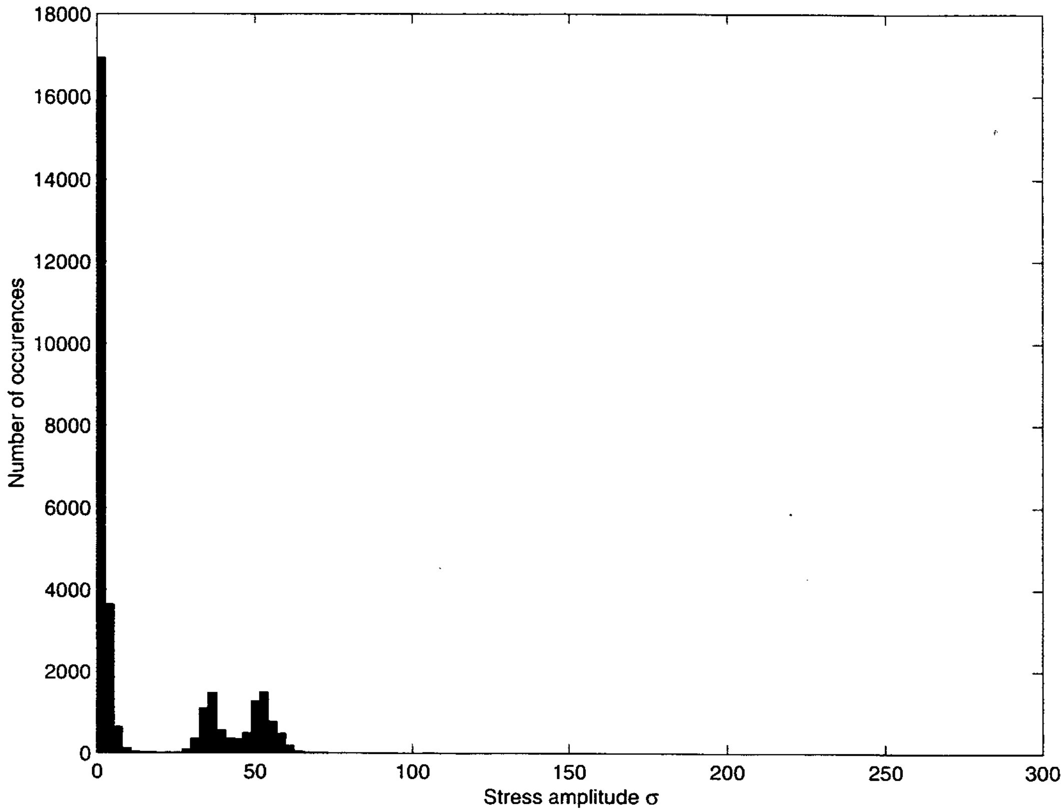

Figure 2: Histogram of rainflow-counted stress amplitudes, all values included.

scrutiny of the rare, high-amplitude stresses that govern fatigue damage. At the same time, the higher thresholds leave us with an increasingly sparse data set, and hence considerable statistical uncertainty. We therefore investigate here whether a more general probabilistic model, such as the quadratic or cubic Weibull, seems to fit the data over a wider range than a simpler Weibull model with fewer parameters. If so, these more general models can be applied with a lower stress threshold, and hence retain a larger fraction of the original data set.

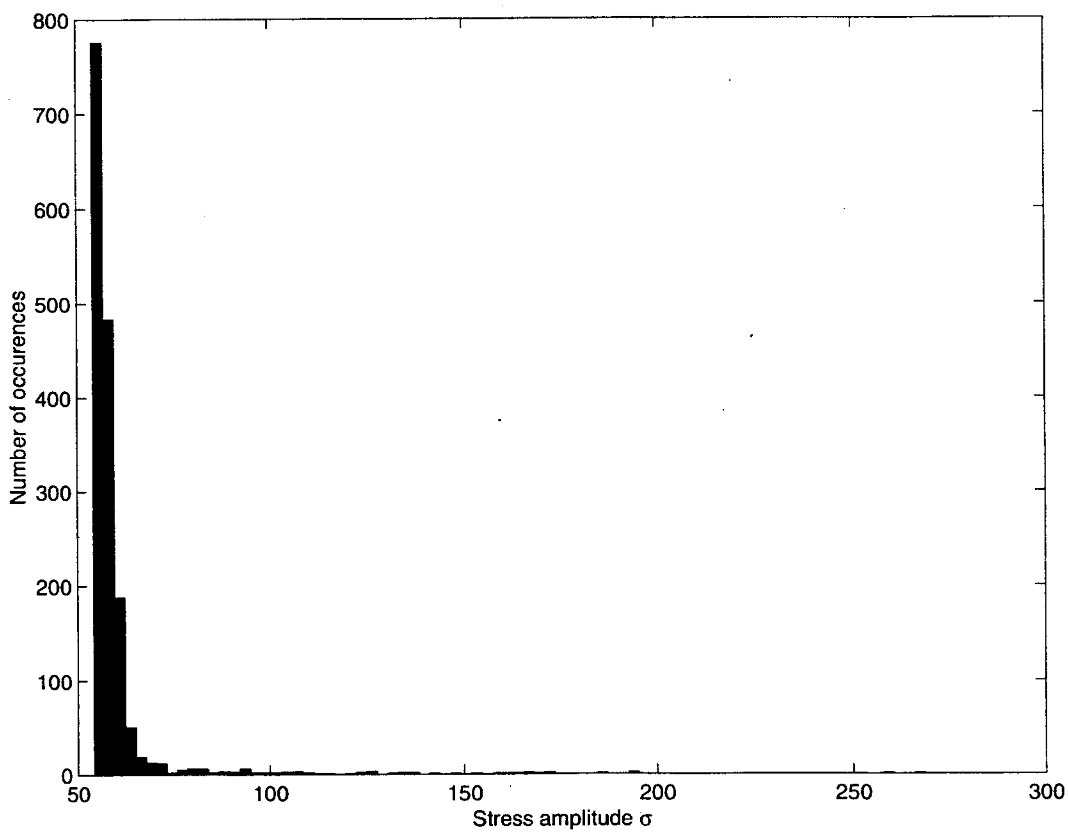

Figure 3: Histogram of rainflow-counted stress amplitudes, values above 54.2 only.

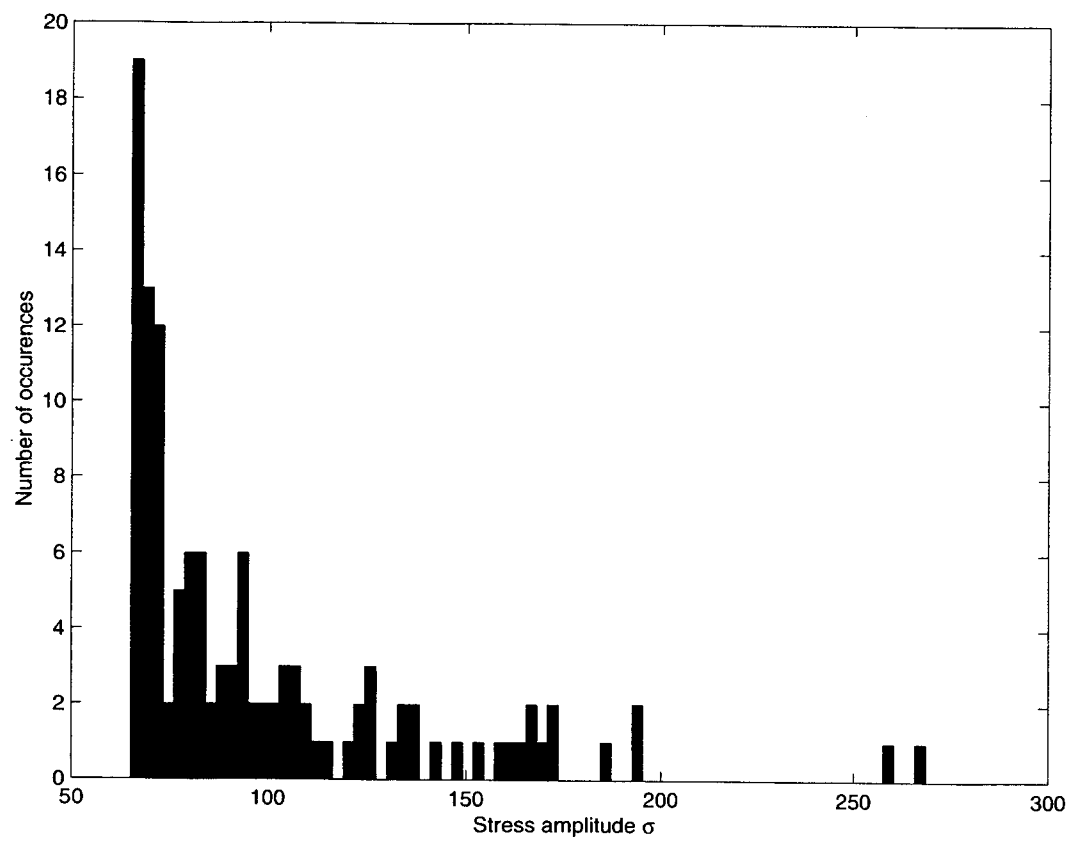

Figure 4: Histogram of rainflow-counted stress amplitudes, values above 65.1 only.

5.2 Sample Input File

We will show here a variety of results based on this data set. In this example, FITS requires 8 lines of input to analyze this single database of wind loads. A typical input file is listed below:

```
quad48.816            ; Input line 1: File where output is to be written
50 400 2.712          ; Input line 2: xmin, xmax, dx for writing dist results
1                     ; Input line 3: target lifetime for calculating Pc
0                     ; Input line 4: iprint = 0,1,2 (raw,histograms,moments)
1                     ; Input line 5: number of databases (input files)
wind.dat              ; Input line 6: name of 1st database (input file)
9                     ; Input line 7: idist = dist type (idist=5 -> Gumbel)
48.816                ; Input line 8: only for shifted distributions: shift
```

(the ";" and following information is not read by FITS; these are given here to remind the user of the input definition).

In our distribution diskette containing FITS, these 8 lines have been stored in a file named quad.in, together with the file wind.dat containing the wind data that produced the histogram in Figure 2. To execute FITS in a batch mode, the user can type (in either Unix or DOS environment) the following command:

```
fits < quad.in
```

Note that this particular input would request a quadratic Weibull distribution (idist=9), and seek to model a T=1 hour maximum.

5.3 Results I: Comparing Weibull and Quadratic Weibull Models

The previous wave height example demonstrated that even a fairly simple, one-parameter distribution model (the exponential)—which clearly does not apply over the entire range of the data set—may be satisfactory once a lower-threshold has been imposed.

In the same way, we may expect that a relatively simple model—here, the standard, two-parameter Weibull mode—would also behave acceptably, at least for a sufficiently high threshold. Figure 5 shows that for thresholds σ_0 at or above 54.2, the Weibull models do appear fairly satisfactory. For the lower threshold of σ_0=48.8, however, results deteriorate significantly. For example, the 10-hour load (hourly p=10^{-1}) appears to be underestimated by roughly one-third.

Figure 6 shows the same results as Figure 5, except that all results shown are now based on the quadratic Weibull model. The use of this more general model is shown to produce a notably more consistent fit across the range of data, even at the lowest threshold level σ_0=48.8. This supports the view, suggested above, that one may usefully extend these more general models farther back into the body of the loads distribution, and hence use a larger fraction of the data when estimating their parameters.

5.4 Results II: Comparing Quadratic Weibull and Cubic Weibull Models

Finally, we compare the quadratic Weibull models in `FITS`, as used to obtain the previous results, with similar results from the four-moment, cubic Weibull model implemented in the routine `FITTING`.

As its name implies, the cubic Weibull model has a distribution function which, when plotted on ordinary Weibull scale, appears as a cubic polynomial. In the same way, the quadratic Weibull appears as a quadratic curve

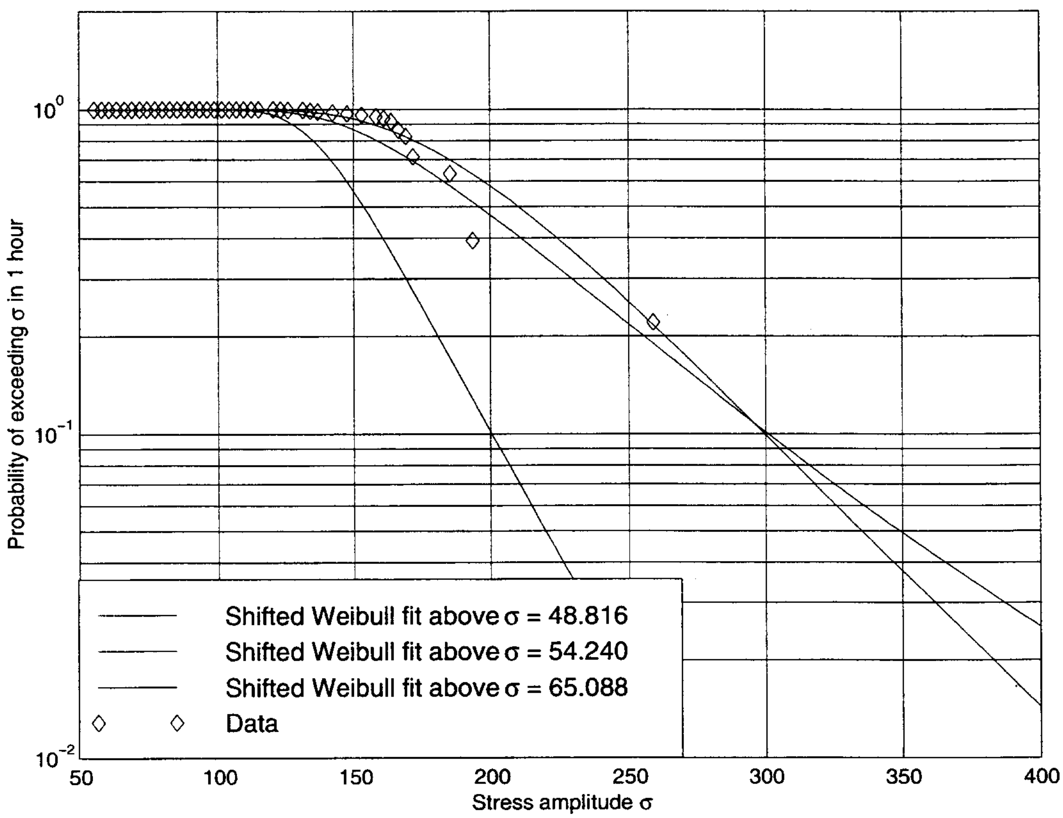

Figure 5: Fitting standard Weibull distributions, with various imposed shifts.

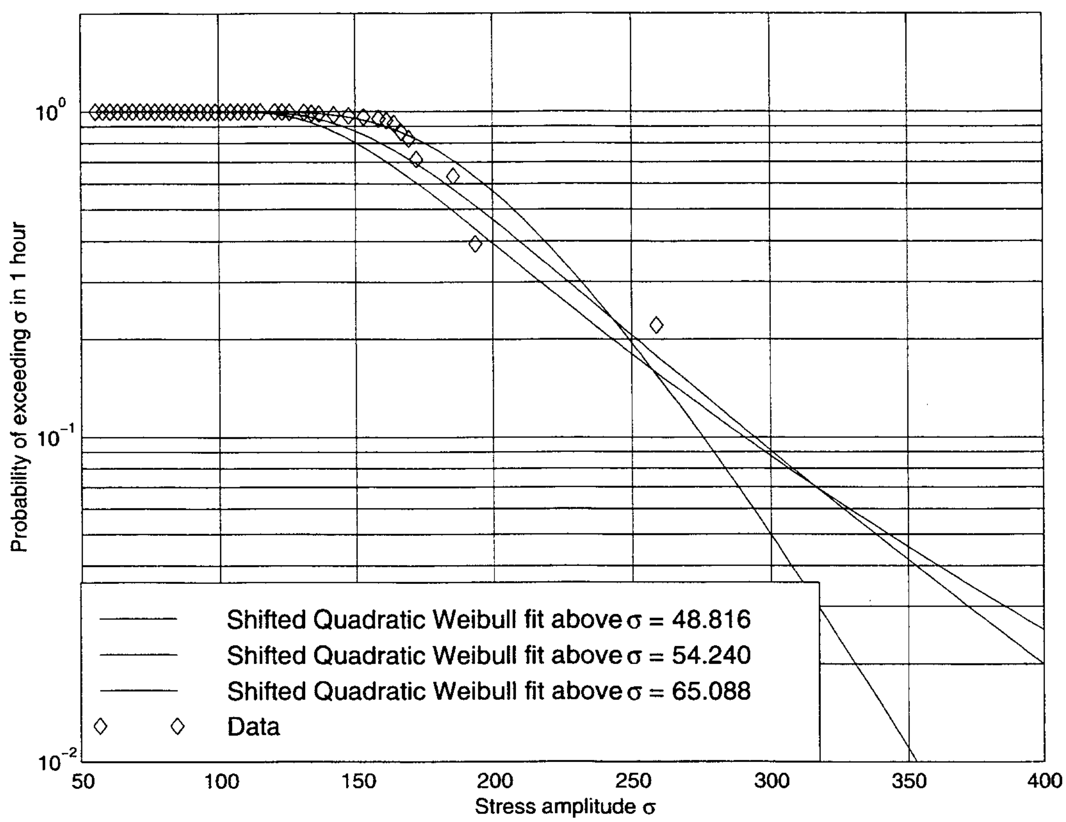

Figure 6: Fitting quadratic Weibull distributions, with various imposed shifts. (Note lesser sensitivity to shift value compared with standard Weibull fit.)

on this Weibull scale. Thus, unlike the quadratic Weibull model, the cubic Weibull distribution function may display a double curvature on Weibull scale. This may sometimes be a desirable capability, e.g., for two-sided phenomena whose left- and right-tail behaviors may differ. For variables such as load amplitudes, which have only a single (upper) tail, this flexibility may lead to overfitting.

Figure 7 shows that for this data set, the cubic Weibull model can display such double curvature. While one may debate whether the data show any such feature, it is fairly clear that extrapolating this highly nonlinear, doubly curved distribution function beyond the range of the data is potentially misleading. The quadratic Weibull appears to follow the trend of the data nearly as well over the range of the data, and by its nature permits a smoother, less tail-sensitive extrapolation beyond the highest data value.

As may be expected, Figure 8 shows that if we choose a sufficiently high threshold, the remaining data appear fairly homogeneous. Thus, in this case the cubic Weibull returns a rather similar result to the quadratic Weibull. (Still in this case, though, note the stronger curvature of the cubic Weibull beyond the last data point.) Indeed, even the standard Weibull may well suffice if only this upper tail is retained for the fit. These examples suggest, however, that the quadratic Weibull model permits a rather useful compromise: it is sufficiently flexible to model a wide range of the data set, yet is sufficiently constrained to retain fairly smooth, well-behaved extrapolation beyond the range of the observed data.

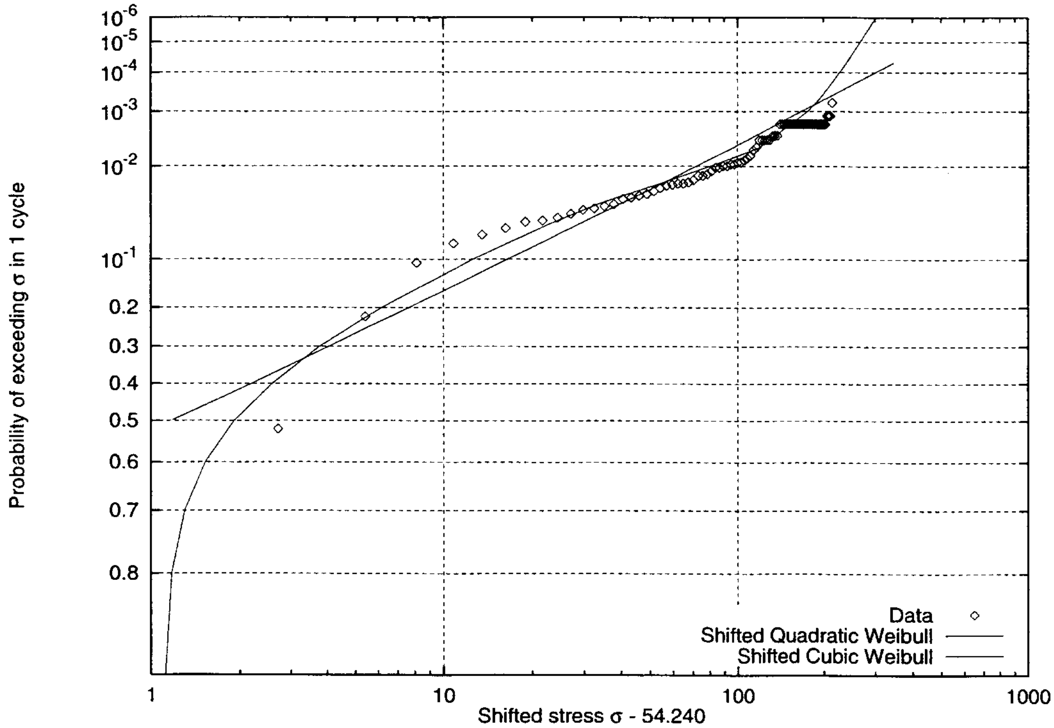

Figure 7: Quadratic vs. cubic Weibull blade load distributions, fit to values above 54.2 only. (Note the double curvature of cubic Weibull model on this Weibull scale plot.)

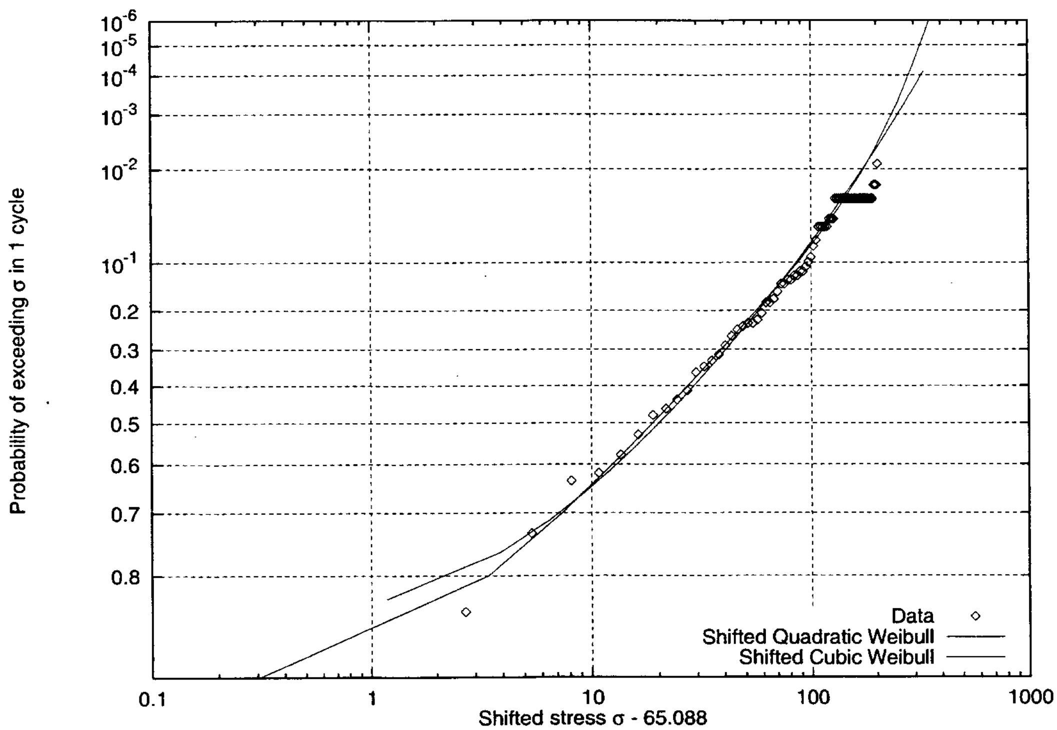

Figure 8: Quadratic vs. cubic Weibull blade load distributions, fit to values above 65.1 only.

6 Additional Input Format Capabilities

6.1 Background

FITS Version 1.1 required that the data sets for fitting be provided in raw data format. For example, in the case of wind applications in fatigue (see Section 5), this meant that large files containing each rainflow-counted stress amplitude needed to be provided as input. Often, such data can be obtained as output from simulation programs or from measurements in more compact forms. A histogram describing the distribution of load amplitudes of different sizes is one such compact form. Another is to describe the database only by the moments of the response/load process along with an indication of the number of events that occurred in the duration covered by the database.

6.2 New Input Capabilities

In the current release of `FITS` the user is permitted to provide data in one of three formats depending on the value of `IPRINT` specified on input.

The following values of `IPRINT` are permissible:

1. `IPRINT`=0 implies that raw data are read on input.
2. `IPRINT`=1 implies that binned/histogram data are read on input.
3. `IPRINT`=2 implies that moments are read on input.

6.3 Illustration of Input Format IPRINT=1

In order to demonstrate the use of the second option for the input format we will use the wind turbine example presented in Section 5. However, the data will be input in histogram format rather than as raw data. The input file for this example is listed below:

```
quad2.out          ; Input line 1: File where output is to be written
50 400 2.712       ; Input line 2: xmin, xmax, dx for writing dist results
1                  ; Input line 3: target lifetime for calculating Pc
1                  ; Input line 4: iprint = 0,1,2 (raw,histograms,moments)
1                  ; Input line 5: number of databases (input files)
wind.hgm           ; Input line 6: name of 1st database (input file)
9                  ; Input line 7: idist = dist type (idist=5 -> Gumbel)
48.816             ; Input line 8: only for shifted distributions: shift
```

In our distribution diskette containing **FITS** these 8 lines have been stored in the file named **quad2.in**. The file **wind.hgm** contains the wind data in histogram format. To execute **FITS** in a batch mode, the user can type (in either Unix or DOS environment) the following command:

```
fits < quad2.in
```

A listing of the file **wind.hgm** is shown below:

```
4.0
47.460  495
50.172 1263
52.884 1481
55.596  776
58.308  483
61.020  188
63.732   51
66.444   19
69.156   13
71.868   12
74.580    2
```

77.292	5
80.004	6
82.716	6
85.428	2
88.140	3
90.852	3
93.564	6
96.276	2
98.988	2
101.70	2
104.41	3
107.12	3
109.83	2
112.54	1
115.26	1
117.97	0
120.68	1
123.39	2
126.10	3
128.82	0
131.53	1
134.24	2
136.95	2
139.66	0
142.38	1
145.09	0
147.80	1
150.51	0
153.22	1
155.94	0
158.65	1
161.36	1
164.07	1
166.78	2
169.50	1
172.21	2
174.92	0

177.63	0
180.34	0
183.06	0
185.77	1
188.48	0
191.19	0
193.90	2
196.62	0
199.33	0
202.04	0
204.75	0
207.46	0
210.18	0
212.89	0
215.60	0
218.31	0
221.02	0
223.74	0
226.45	0
229.16	0
231.87	0
234.58	0
237.30	0
240.01	0
242.72	0
245.43	0
248.14	0
250.86	0
253.57	0
256.28	0
258.99	1
261.70	0
264.42	0
267.13	1

The output of the program with the above input files is identical to that presented in Section 5 for the Shifted Weibull fit with a stress threshold of 48.8. In that case, however, the input file consisted of the raw data, i.e., all the stress amplitudes, and the `IPRINT` option was set to 0.

6.4 Illustration of Input Format IPRINT=2

In order to demonstrate the use of the third option for the input format we will use the wave height example presented in Section 4. In this case, the statistical moments of the annual maximum H_S values are input instead of the raw data.

In this example, `FITS` requires 7 lines of input as follows:

```
gumbel2.out             ; Input line 1: File where output is to be written
5.0  20.0  0.5          ; Input line 2: xmin, xmax, dx for writing dist results
1                       ; Input line 3: target lifetime for calculating Pc
2                       ; Input line 4: iprint = 0,1,2 (raw,histograms,moments)
1                       ; Input line 5: number of databases (input files)
gumbel.mom              ; Input line 6: name of 1st database (input file)
5                       ; Input line 7: idist = dist type (idist=5 -> Gumbel)
```

In the distribution diskette these 7 lines have been stored in the file named `gumbel2.in`. The file `gumbel.mom` contains the moments (mean, standard deviation, skewness, and kurtosis) based on the 19 annual maximum H_S values.

To execute `FITS` in batch mode, the user must type (in either Unix or DOS environment) the following command:

```
fits < gumbel2.in
```

A listing of the file `gumbel.mom` is shown below:

```
19.  19
   0.8275E+01  0.8186E+00 -0.5282E-01  0.1905E+01
```

Note that the in `gumbel.mom`, the first line contains two values. The first value, i.e., 19., refers to the duration (19 years) covered by the database. The second value, i.e., 19, refers to the number of data that were available in the database from which the moments were estimated. (Note that the two numbers on the first line of `gumbel.mom` are the same in this illustration because the database contained only annual maxima. If maxima were collected every three months and were available in a database, the first value would still be 19. but the second value would be 76.)

The second line in `gumbel.mom` contains the first four moments based on the 19 annual maximum H_S values.

7 References

Kashef, T. and S.R. Winterstein (1998). *Moment-based probability modelling and extreme response estimation: The FITS routine,* Rept. RMS–31, Reliability of Marine Structures Program, Dept. of Civ. & Environ. Eng., Stanford University.

Lange, C.H. and S.R. Winterstein (1996). Fatigue design of wind turbine blades: load and resistance factors from limited data. *Proc., Wind Energy 1996,* ASME, SED **17**, 93–101.

Madsen, H.O., S. Krenk, and N.C. Lind (1986). *Methods of structural safety,* Prentice-Hall, Inc., New Jersey.

Ude, T.C. and S.R. Winterstein (1996). *IFORM: An inverse-FORM routine to estimate response levels with specified return periods,* Rept. TN–3, Reliability of Marine Structures Program, Civil Eng. Dept., Stanford University.

Winterstein, S.R. (1995). *The FITS routine: Fitting distributions to multiple databases and estimating combined extremes,* Rept. RMS–19, Reliability of Marine Structures Program, Dept. of Civ. & Environ. Eng., Stanford University.

Winterstein, S.R., C.H. Lange, and S. Kumar (1994). *FITTING: A subroutine to fit four moment probability distributions to data,* Rept. RMS–14, Reliability of Marine Structures Program, Civil Eng. Dept., Stanford University.

Winterstein, S.R. and C.H. Lange (1996). Load models for fatigue reliability from limited data. *Journal of Solar Energy,* Trans. ASME, **118**(1), 64–68.

Winterstein, S.R. and S. Haver (1991). Statistical uncertainties in wave heights and combined loads on offshore structures. *J. Offshore Mech. Arctic Eng.,* ASME, **114**(10), 1772–1790.

A Specific Distribution Types and Functional Forms

Our purpose here is twofold. First, we seek to define the precise forms of probability distributions available in FITS. Second, we wish to indicate precisely how the distribution parameters are defined, and reported in FITS through the output of the subroutine DISPAR. This latter information may be useful if the user seeks to perform additional, off-line calculations based on the fitted distributions.

The following basic distribution types are currently available within FITS.

IDIST=1: **Normal Distribution.** The cumulative distribution function (CDF) is given by

$$P[\text{ Outcome } \leq x] = F(x) = \Phi(\frac{x - m_x}{\sigma_x}) \tag{3}$$

in terms of the standard normal CDF

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2) du \tag{4}$$

In this case the subroutine DISPAR returns the parameters DPARM(1)=m_x and DPARM(2)=σ_x.

IDIST=2: **Lognormal Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = \Phi(\frac{\ln(x) - m_{\ln x}}{\sigma_{\ln x}}); \quad x \geq 0 \tag{5}$$

with $\Phi(x)$ again given by Eq. 4. In this case the subroutine DISPAR returns the parameters DPARM(1)=m_x, DPARM(2)=σ_x, DPARM(3)=$m_{\ln x}$, and DPARM(4)=$\sigma_{\ln x}$.

IDIST=3: **Exponential Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = 1 - \exp[-(\frac{x}{m_x})]; \quad x \geq 0 \tag{6}$$

In this case the subroutine DISPAR returns the parameters DPARM(1)=m_x and DPARM(2)=σ_x, and no additional parameters.

`IDIST`=4: **Weibull Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = 1 - \exp[-(\frac{x}{\beta})^{\alpha}]\,; \quad x \geq 0 \qquad (7)$$

In this case the subroutine `DISPAR` returns the parameters DPARM(1)=m_x, DPARM(2)=σ_x, DPARM(3)=α, and DPARM(4)=β.

`IDIST`=5: **Gumbel Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = \exp\{-\exp[-\alpha(x-u)]\} \qquad (8)$$

In this case the subroutine `DISPAR` returns the parameters DPARM(1)=m_x, DPARM(2)=σ_x, DPARM(3)=α, and DPARM(4)=u.

`IDIST`=6: **Shifted Exponential Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = 1 - \exp[-(\frac{x - x_{low}}{m_x})]\,; \quad x \geq x_{low} \qquad (9)$$

As when `IDIST`=3, in this case the subroutine `DISPAR` returns the parameters DPARM(1)=m_x and DPARM(2)=σ_x, and no additional parameters.

`IDIST`=7: **Shifted Weibull Distribution.** The CDF is given by

$$P[\text{ Outcome } \leq x] = F(x) = 1 - \exp[-(\frac{x - x_{low}}{\beta})^{\alpha}]\,; \quad x \geq x_{low} \qquad (10)$$

As when `IDIST`=4, in this case the subroutine `DISPAR` returns the parameters DPARM(1)=m_x, DPARM(2)=σ_x, DPARM(3)=α, and DPARM(4)=β.

`IDIST`=8: **Quadratic Weibull Distribution.** The quadratic Weibull model relates the physical variable X to a Weibull variable, W, with the same coefficient of variation as that of X. If the skewness of the data exceeds that of W, a quadratic term is added to W to broaden its probability distribution:

$$X = x_{min} + \kappa[W + \epsilon W^2] \qquad (11)$$

If the skewness of the data is instead less than that of W, the roles of W and X in Eq. 11 are interchanged:

$$W = \left(\frac{X - x_{min}}{\kappa}\right) + \epsilon\left(\frac{X - x_{min}}{\kappa}\right)^2 \qquad (12)$$

(This quadratic equation is readily inverted to yield an explicit result for X in terms of W.) In either case, the CDF of W is analogous to Eq. 7:

$$P[\text{ W } \leq w] = F(w) = 1 - \exp[-(\frac{w}{\beta_W})^{\alpha_W}]; \quad w \geq 0 \qquad (13)$$

In this case the subroutine **DISPAR** returns the parameters DPARM(1)=x_{min}, DPARM(2)=κ, DPARM(3)=β_W, DPARM(4)=$1/\alpha_W$. and DPARM(5)=$\pm\epsilon$. (The actual ϵ value used in Eqs. 11 or 12 is always positive; a negative value of ϵ indicates that the program uses Eq. 12 rather than Eq. 11.)

IDIST=9: **Shifted Quadratic Weibull Distribution.** This model shifts the quadratic Weibull model by a user-defined lower limit, x_{low}. A quadratic Weibull model is then fit to $X - x_{low}$; i.e.,

$$X = x_{low} + x_{min} + \kappa[W + \epsilon W^2] \qquad (14)$$

or

$$W = \left(\frac{X - x_{low} - x_{min}}{\kappa}\right) + \epsilon\left(\frac{X - x_{low} - x_{min}}{\kappa}\right)^2 \qquad (15)$$

Note the distinction between x_{low} and x_{min}: x_{low} is the intended lower bound given by the user, and x_{min} is an additional shift the program computes (in order to preserve the mean of the data). The actual lowest value X can attain is therefore $x_{low} + x_{min}$.

DISTRIBUTION

H. Ashley
Dept. of Aeronautics and
Astronautics Mechanical Engr.
Stanford University
Stanford, CA 94305

B. Bell
Zond Systems Inc.
13000 Jameson Road
P.O. Box 1910
Tehachapi, CA 93561

C. P. Butterfield
NREL
1617 Cole Boulevard
Golden, CO 80401

G. Bywaters
Northern Power Technology Co.
Box 999
Waitsfield,VT 05673

J. Cadogan
U.S. Department of Energy
Office of Photovoltaic & Wind
Technology
Energy Efficiency & Renewable Energy
EE-11
1000 Independence Avenue SW
Washington, DC 20585

D. Cairns
Montana State University
Mechanical & Industrial Engineering Dept.
Bozeman, MT 59717

S. Calvert
U.S. Department of Energy
Office of Photovoltaic & Wind
Technology
Energy Efficiency & Renewable Energy
EE-11
1000 Independence Avenue SW
Washington, DC 20585

J. Chapman
OEM Development Corp.
840 Summer St.
Boston, MA 02127-1533

R. N. Clark
USDA
Agricultural Research Service
P.O. Drawer 10
Bushland, TX 79012

J. Cohen
Princeton Economic Research, Inc.
1700 Rockville Pike
Suite 550
Rockville, MD 20852

C. Coleman
Northern Power Technology Co.
Box 999
Waitsfield, VT 05673

K. J. Deering
The Wind Turbine Company
515 116th Avenue NE
No. 263
Bellevue, WA 98004

A. J. Eggers, Jr.
RANN, Inc.
744 San Antonio Road, Ste. 26
Palo Alto, CA 94303

P. R. Goldman
Acting Deputy Director
Office of Photovoltaic and
Wind Technology
Energy Efficiency & Renewable
Energy, EE-11
U.S. Department of Energy
1000 Independence Avenue
Washington, DC 20585

G. Gregorek
Aeronautical & Astronautical Dept.
Ohio State University
2300 West Case Road
Columbus, OH 43220

C. Hansen
Windward Engineering
4661 Holly Lane
Salt Lake City, UT 84117

S. Hock
Wind Energy Program
NREL
1617 Cole Boulevard
Golden, CO 80401

K. Jackson
Dynamic Design
123 C Street
Davis, CA 95616

G. James
University of Houston
Dept. of Mechanical Engineering
4800 Calhoun
Houston, TX 77204-4792

A. Lucero
Librarian
National Atomic Museum
Albuquerque, NM 87185

R. Lynette
Springtyme Co.
212 Jamestown Beach Lane
Sequim, WA 98382

D. Malcolm
Kamzin Technology Inc.
425 Pontius Avenue North
Suite 150
Seattle, WA 98109

J. F. Mandell
Montana State University
302 Cableigh Hall
Bozeman, MT 59717

T. McCoy
RLA
425 Pontius Avenue North
Suite 150
Seattle, WA 98109

R. N. Meroney
Dept. of Civil Engineering
Colorado State University
Fort Collins, CO 80521

P. Migliore
NREL
1617 Cole Boulevard
Golden, CO 80401

A. Mikhail
Zond Systems, Inc.
13000 Jameson Road
P.O. Box 1910
Tehachapi, CA 93561

E. Moroz
Zond Systems, Inc.
13000 Jameson Road
P.O. Box 1910
Tehachapi, CA 93561

W. Musial
NREL
1617 Cole Boulevard
Golden, CO 80401

NWTC Library (5)
NREL
1617 Cole Boulevard
Golden, CO 80401

V. Nelson
Department of Physics
West Texas State University
P.O. Box 248
Canyon, TX 79016

R. G. Rajagopalan
Aerospace Engineering Department
Iowa State University
404 Town Engineering Bldg.
Ames, IA 50011

Michael Robinson
NREL
1617 Cole Boulevard
Golden, CO 80401

D. Sanchez
U.S. Dept. of Energy
Albuquerque Operations Office
P.O. Box 5400
Albuquerque, NM 87185

L. Schienbein
CWT Technologies, Inc.
4006 S. Morain Loop
Kennewick, WA 99337

R. Sherwin
Atlantic Orient
PO Box 1097
Norwich, VT 05055

Brian Smith
NREL
1617 Cole Boulevard
Golden, CO 80401

K. Starcher
AEI
West Texas State University
P.O. Box 248
Canyon, TX 79016

F. S. Stoddard
Dynamic Design-Atlantic Office
P.O. Box 1373
Amherst, MA 01004

A. Swift
University of Texas at El Paso
320 Kent Ave.
El Paso, TX 79922

R. W. Thresher
NREL
1617 Cole Boulevard
Golden, CO 80401

W. A. Vachon
W. A. Vachon & Associates
P.O. Box 149
Manchester, MA 01944

B. Vick
USDA, Agricultural Research Service
P.O. Drawer 10
Bushland, TX 79012

L. Wendell
2728 Enterprise Dr.
Richland, WA 99352

R. E. Wilson
Mechanical Engineering Dept.
Oregon State University
Corvallis, OR 97331

S. R. Winterstein (10)
Civil Engineering Department
Stanford University
Stanford, CA 94305

M. Zuteck
MDZ Consulting
931 Grove Street
Kemah, TX 77565

M.S. 0555	B. Hansche, 9122	
M.S. 0557	T. J. Baca, 9119	
M.S. 0557	T. G. Carne, 9119	
M.S. 0557	T. L. Paez, 9119	
M.S. 0708	H. M. Dodd, 6214 (25)	
M.S. 0708	T. D. Ashwill, 6214	
M.S. 0708	D. E. Berg, 6214	
M.S. 0708	P. L. Jones 6214	
M.S. 0708	D. L. Laird, 6214	
M.S. 0708	M. A. Rumsey, 6214	
M.S. 0708	H. J. Sutherland, 6214	
M.S. 0708	P. S. Veers, 6214	(10)
M.S. 0847	D. W. Lobitz, 9125	
M.S. 0847	D. R. Martinez, 9124	
M.S. 0612	Review & Approval Desk, 4912 For DOE/OSTI	
M.S. 0899	Technical Library, 4916	(2)
M.S. 9018	Central Technical Files, 8940-2	