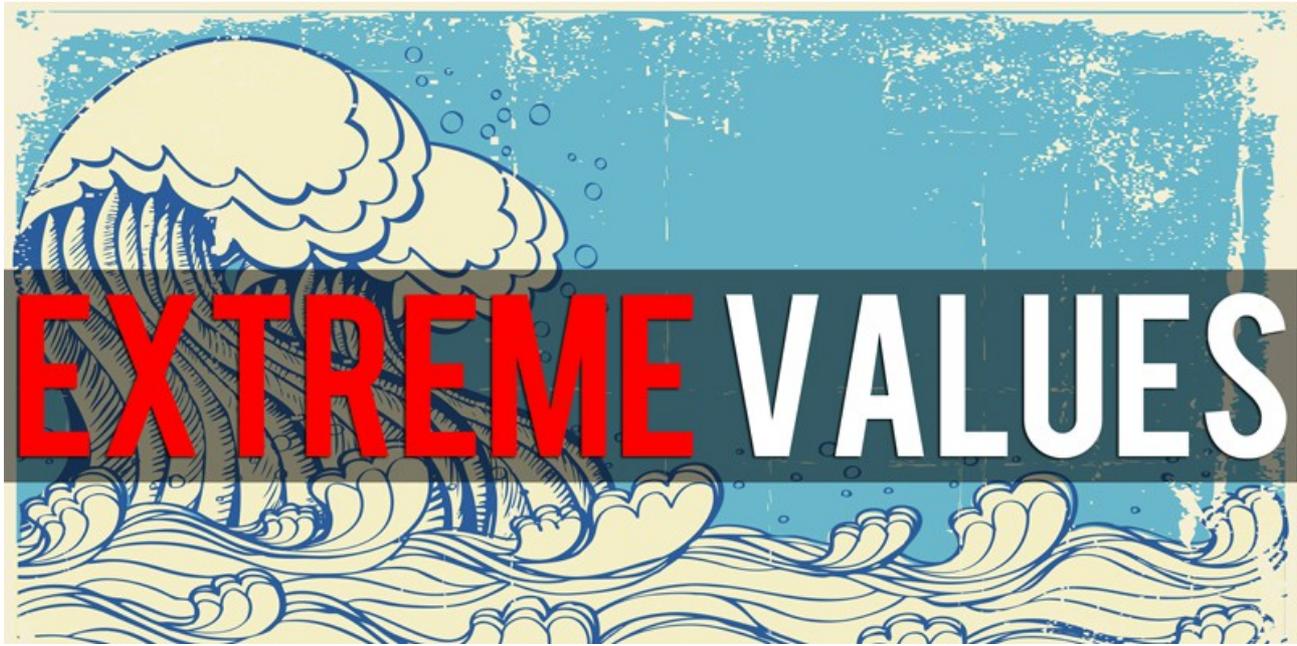


Extreme Values - Part I



by Arnold Doray

The concept of extreme values can be confusing even to the experienced engineer.

Much of this head scratching involves:

1. How extreme values should be interpreted in decision making,
2. How to objectively assess competing estimates of extreme values and
3. How extreme values relate to well-understood concepts like exceedances.

Some of the questions I've typically encountered:

- Can extreme values be estimated from exceedances?
- Is a "risk factor" of 10% the same as an exceedance of 10%?
- Does a "100 year wave" mean that wave height is only experienced just once every 100 years?
- Is the 50-year return value wave the extreme or maximum wave most probably experienced within a 50 year period?
- How should a "10,000 year" extreme value be interpreted?
- Can 10, 20 or 100 year extreme values be reliably estimated from a year of measurement data?
- How can exposure time be factored into extreme value calculations?
- Why are "cyclonic" and "non-cyclonic" winds or waves calculated separately?

In this **3-part article**, I will answer these questions but I really hope to leave you with a practical knowledge of extreme values.

The Lay of the Land

Part 1 : Discusses some basic concepts behind extreme values, as it is applied in mainstream offshore engineering today. This should help clear most of the doubts and misconceptions I've encountered among practicing engineers.

Part 2: Discusses the conceptual underpinnings of extreme value theory. We will then work our way through the "block maxima" and "peaks over threshold" methods since these are commonly used in engineering. I will also discuss the use (and abuse) of the Weibull distribution for estimating extremes. Lastly, I will discuss the largely ignored -- at least in mainstream engineering practice -- but very important idea of the robustness of an extreme value estimate.

Part 3: Discusses extremes along routes, including the concept of "adjusted extremes" pioneered by a certain well-established marine warranty company.

Extreme Value Theory is a highly mathematical subject, but I have tried very hard to avoid any mathematical jargon in these discussions, but rather focus on the concepts. If you know arithmetic and just a little high school algebra, you should be good to go.

A Question of Terms

Extreme values are also known as "return values". We will use these terms interchangeably throughout this article. However, do not confuse "extreme value" to mean "maximum value", they are not the same thing.

What Are Extreme Values?

An extreme value is always defined over a given *return period*, usually measured in years. For example a 10-

year return wave is one that is expected to be equaled or exceeded on average one year out of every 10 years.

Generalizing:

Definition #1: An *X*-year *return value* is the value that is equaled or exceeded on average one year out of every *X* years. *X* is the *return period* in years.

This definition is far more subtle than appears at first glance. In my experience, this subtlety is the source of most problems of interpretation faced by engineers. So, in the rest of this article I will focus on unpacking its meaning.

We'll start by using this definition to answer a question I've often been asked by engineers:

Question 1: Does a "100 year wave" mean that wave height is only experienced just once every 100 years?

The answer is obviously "No". The definition says that extreme value is equaled or exceeded **on average** (note the emphasis!) one year out of every 100 years, not exactly once.

For example, suppose the waves at the Gauge Field are known to have a 10-year extreme wave of 2.5m. One *possible* observation over a 30-year period could be that:

1. Waves equaled or exceeded 2.5m in just one year during the period 1970 to 1979,
2. two separate years from 1980 to 1989 and
3. not at all for 1990 to 1999

The average is $\frac{1+2+0}{3}=1$ and therefore fully compatible with the 10-year return value of 2.5m, since waves equaled or exceeded 2.5m one year out of 10 **on average**.

Fact #1: A "100 year wave" may occur more than once (or never at all) in a 100 year period.

Similar statements apply to other return periods. Later in this article we will revisit this issue to calculate how likely it is to experience a 100-year wave twice (or more) in 100 years.

Question 2: Is the 50-year return value wave the extreme or maximum wave most probably experienced within a 50 year period?

Again, the answer is "No". The definition says that extreme value is **equaled or exceeded** (note the emphasis!) on average one year out of every 50 years. Since the extreme value could be *exceeded*, it might not therefore be the maximum over the 50 years.

Question 3: Can extreme values be estimated from exceedances?

In my experience, this is often two different questions:

1. **Can exceedances be used in a curve-fitting exercise and the tail of that distribution be used to determine the extreme values?** I will discuss this question in detail in **Part 2** of this article, since many companies that measure wind/wave/currents/temperatures over short periods attempt to overcome their limited dataset by using this approach.

2. Can I estimate extreme values directly from exceedance tables?

To answer the second question, let's attempt a simple thought experiment:

"In the land of Abz, there are two coastal towns called Ay and Bi that experience very different weather patterns. The oceanographers at Ay and Bi have each compiled a continuous hourly record of wave heights at their respective ports over a period of 10 years. Upon analysis of the datasets, the oceanographers of Ay note that waves exceeded 2.5m for the entire first year but did not do so for the remaining 9 years. The oceanographers at Bi had a very different result -- the waves exceeded 2.5m 10% of the time for every single one of the 10 years."

Taken over the whole 10 years, the exceedance of 2.5m is 10% at both sites, yet their wave climate are obviously very different. Clearly, the exceedances alone do not tell you the whole story.

What about the 10-year extreme waves? Are they the same for Ay and Bi?

The definition says that 10-year return wave is equaled or exceeded on average **one year out of every 10 years** (note the emphasis!).

In Ay, 2.5m was exceeded only on one year, so it would be reasonable to conclude that the 10-year return wave at Ay is 2.5m. In fact, this is not quite true, since the definition speaks of **averages**, and you cannot accurately estimate an average from just one sample. If the records

at Ay were extended for another 90 years, then you would have 10 samples (10 samples of 10 years), and that would give a better estimate of the 10-year return waves. In **Part 2** we will explore these problems further:

- How to objectively gauge the robustness of a given extreme value calculation and
- How to confidently estimate the 10-year returns from smaller datasets, since very long ones are unlikely to be available.

For now, we can safely conclude that at a rough approximation the 10-year waves at Ay is 2.5m. Note that we did not do this based on the exceedances alone, but by also reasoning on the temporal pattern of waves over the 10 years.

In Bi, the waves exceeded 2.5m every one of the 10-years, so it would be reasonable to conclude that the 10-year return wave at Bi is much higher than 2.5m. Again, this is not quite true, since the definition speaks of **averages**. We will see later on that it is possible, though highly unlikely that waves equal or exceed their 10-year returns ten years in a row! I will give you the tools to calculate exactly how (un)likely this is shortly.

The table below summarizes our findings:

Site	% Exceedance above 2.5m	10-year Return Wave
Ay	10%	Roughly 2.5m
Bi	10%	Likely much higher than 2.5m

Again, I want to emphasize that *the exceedance alone would not allow us to infer anything about the extreme values*. We had to know more about the temporal variation of the waves. This is so since exceedances are taken over the entire period of data collection, while the

"equal or exceeded" part of our definition breaks up that period on a yearly basis.

Thus, absent this temporal information, we would have to make **assumptions** about how the waves varied over time in order to say anything about the extreme values at Ay or Bi. We will look at these assumptions later in **Part 2** of this series.

So, when you see a report or someone inferring a return value from an exceedance value, that should raise a red flag.

The Risk Factor

Given a 10-year return wave of 2.5m, what are the chances of getting say, *five* separate years within a 10 year period when waves equal or exceed 2.5m during those five years? This is of course unlikely, but exactly how unlikely?

To answer this question, we need to use an alternative definition of the "average occurrence":

Definition #2: The average occurrence of an event over a given period is equal to the probability of it occurring at any given year multiplied by the number of years under consideration.

For example, suppose there is a 10% chance that a vessel encounters a wave exceeding 2.5m at the Gauge Field at any one year. Then, over 5 years, there is a $10\%/year \times 5 \text{ years} = 50\%$ chance of waves exceeding that limit.

How about over a period of 10 years? $10\%/year \times 10$

years = 100%. This should be interpreted as there being **on average** a single occurrence of encountering a wave that exceeds 2.5m.

This 10% per year probability is called the *risk factor*, and is often expressed as a fraction:

(eg, $\left(\frac{10}{100}\right)\% = 0.1$) The risk factor frequently

occurs in reports and engineering literature, and is a perennial source of confusion with percentage of exceedance values.

So, let us draw the distinction between a risk factor of (say) 10% and an exceedance of 10%:

1. **A Risk Factor of 10%** means that the event occurs one year out of every 10, on average. In that one year in which it occurs, it could occur more than once, that would not affect the risk factor.

2. **An Exceedance of 10%** means that that event occurs in 10% of the time over 10 years. Whether that is 100% of one year out of 10 years or 10% of every year, is unknown.

Fact #2: The risk factor is **not** related to percentage exceedance values.

Using **Definition #2**, it is easy to see that the risk factor is inversely proportional to return period, since:

An average of 1 Extreme Event = Risk Factor x Return Period.

Fact #3: The risk factor is inversely proportional to the return period.

For example, a 10-year return period is equivalent to a

$$\frac{1}{10} = 0.1 = 10\% \text{ risk factor. Similarly, a 50-year}$$

return period is equivalent to a $\frac{1}{50} = 0.02 = 2\%$ risk factor.

How Likely are Multiple Extreme Events?

We now have all the concepts we need to calculate the chances of zero or multiple extreme events occurring during a given return period. Consider a 2-year return period. We know that:

•By **Fact #3**, we know that the associated risk

$$\text{factor is } \frac{1}{2} = 0.5 = 50\%$$

•The probability of an extreme event **not** occurring in any given year is just 100% - probability of occurrence = $1 - 0.5 = 0.5$.

To calculate the likelihood of multiple extreme events during the 2-year period, we need to enumerate all the possible ways a vessel could experience the extreme event, and calculate the probability of each combination:

Extreme Event in Year 1?	Extreme Event in Year 2?	# Years with Extreme Events	Probability Calculation	Probability
No	No	0	$(1 - 0.5) \times (1 - 0.5) = 0.5 \times 0.5 = 0.25 = 25\%$	25%
Yes	No	1	$0.5 \times (1 - 0.5) = 0.5 \times 0.5 = 0.25 = 25\%$	25%
No	Yes	1	$(1 - 0.5) \times 0.5 = 0.5 \times 0.5 = 0.25 = 25\%$	25%
Yes	Yes	2	$(1 - 0.5) \times (1 - 0.5) = 0.5 \times 0.5 = 0.25 = 25\%$	25%

The Importance of Being Independent

In this calculation, we made an huge assumption that the extreme events are *independent* of each other. This allows us to just multiply probabilities for each year to get the joint probability for the two years. This idea of independence is so critical to most engineering calculations involving extreme events that we will enshrine it in an Assumption:

Assumption #1: Extreme events are assumed to occur independently of each other.

Some examples:

- The **annual** maxima of wind speed over a given location are likely to be independent events.
- The **daily** maxima of air temperature for a given location are **not** independent events. Can you see why?

Most meteorological events become quickly decorrelated (i.e., independent) the further they are separated in time. So, a year's separation used to calculate **annual** maxima ensures that they independent. Today's air temperature on the other hand, is likely affected by yesterday's weather. Thus, daily air temperature maxima are strongly correlated to each other, they are *not* independent.

Coming back to our calculation, we see that:

# Years with Extreme Events	Probability
0	25%
1	25% + 25% = 50%
2	25 %

This table summarizes exactly how likely multiple extreme events are. Note that for a 2-year return period, there is a 1:4 chance of experiencing the extreme event on both years.

*Calculating the Probability of Multiple Extreme Events

This simple calculation can be extended to 10-year or 100-year extreme events, but the tables would be very much longer. Instead, a simpler method is to recognize that these probabilities follow the Binomial distribution, so there is a simple formula to calculate them. The probability P of experiencing k extreme events within a return period of n years at a

risk factor of r is:
$$P(k) = \left(\frac{n!}{k!(n-k)!} \right) r^k (1-r)^{n-k} \text{ and } r = \frac{1}{n}$$

The ! is a factorial. For example, $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5040$. The Excel function is FACT. Eg, $\text{FACT}(7) = 7! = 5040$. For example, to answer our previous question:

Question 4: Given a 10-year return wave of 2.5m, what are the chances of getting say, five separate years within a 10 year period when waves equal or exceed 2.5m during those five years?

We have $k=5$, $n=10$, and $r = \frac{1}{n} = \frac{1}{10} = 0.1$. So,

We have
$$P(5) = \left(\frac{10!}{5!(10-5)!} \right) 0.1^5 (1-0.1)^{10-5} = 0.15\%$$

In other words, there is slightly better than a 1:100 chance of experiencing waves equaling or exceeding 2.5m in 5 years out of the 10. Some other probabilities (for a return period of 100 years)

Return Period (years)	# Years with Extreme Events	Probability (%)
100	0	36.6%
100	1	37.0%
100	2	18.5%

100	3	6.1%
100	4	1.5%
100	5	0.3%

[I have put this into an interactive calculator which you may download here.](#)

Conclusion

This brings us to the end of **Part 1** of this series on Extreme Values. In **Part 2**, we will:

- Explore the conceptual underpinnings of extreme value theory,
- Explain **how** extreme values are usually calculated in mainstream marine engineering,
- Examine the much ignored issue of how to measure the robustness of an extreme value estimate,
- Discuss the use and abuse of the Weibull distribution for estimating extremes.

[Subscribe to KnowledgeVault](#)

If you'd like to get the latest updates of articles, [subscribe](#) to us and we'll send it to you!