

Analysis of Tropical Storm Damage using Buffered Probability of Exceedance

Justin R. Davis¹ and Stan Uryasev²

August 10, 2015

RESEARCH REPORT #2014-4

¹justin.r.davis@essie.ufl.edu / 352-294-7810

Engineering School of Sustainable Infrastructure & Environment
University of Florida, Gainesville, FL 32611

²Department of Industrial and Systems Engineering
University of Florida, Gainesville, FL 32611

Abstract

Relying solely on probability of exceedance (POE) or in combination with other statistical measures such as mean or standard deviation does not always give a clear picture of data which have heavy tailed distributions. This paper explains the new concept of buffered probability of exceedance (bPOE). Through averaging of data in the tail, bPOE compactly presents probability as well as loss value of the tail and has the capability to revolutionize the concept of risk-averse engineering. Once defined, bPOE is then demonstrated through its application to normalized tropical storm damage along the Atlantic and Gulf of Mexico coasts of the U. S. from 1900-2013 (average damage \$6.961 billion). For instance, with a \$50 billion threshold, the probability of exceedance (POE) is 3.7% while the bPOE is significantly higher at 9.7%. It can be shown, under general assumptions, that bPOE is more than double POE. E.g., for a 25% probability, the value-at-risk (quantile) determined using POE is \$4.785 billion, while the conditional-value-at-risk (average of damage in excess of the quantile) determined by the bPOE is \$25.176 billion. For a fixed probability, a general formula has been developed for a quantity called “expected excess”, the difference between the damage associated with POE and bPOE multiplied by the bPOE. Formally thought of as the damage above the value-at-risk (quantile), it is, for example, the minimum amount an insurer would need to charge in premiums to “break even”. For example, the \$20.391 difference shown for the 25% probability level, leads to an expected excess of $0.25 * \$20.391 = \5.098 billion. In addition to applying bPOE to an entire dataset, it can also be applied to statistically meaningful subsets. For example, subdividing the data by landfall state, at the 50% probability level, Florida, the state most often hit by tropical storms also has the highest value for expected excess at \$8.111 billion; thus, quantifying the need for insurers to charge Floridians high insurance premiums.

Keywords: Buffered probability of exceedance, conditional-value-at-risk, tropical storm damage, Atlantic basin

Introduction

Tropical storms are widely regarded as one of the most destructive natural hazards. High winds, waves, coastal storm surge and inundation, inland flooding due to torrential rains, and propensity to spawn tornados can cause catastrophic damage and loss-of-life over large expanses. As these storms can occur nearly anywhere in warm low-middle latitude waters, the better understanding of both their causes and effects is a world-wide concern. Although the western Pacific Ocean (where these storms are commonly referred to as “cyclones”) is home to the most active region of tropical storm formation, these storms tend to have relatively infrequent and moderate impacts on the U. S. This is in stark contrast to the Atlantic and Gulf of Mexico coasts of the U.S. where between 1900 and 2013, 242 storms have made landfall along the Atlantic and Gulf of Mexico coasts resulting in damages of \$1.68 trillion (all economic damages reported herein use 2014 constant dollars) [1]. In the Atlantic hurricane basin, particularly large and powerful storms (defined as wind speeds ≥ 74 mph) are referred to as “hurricanes” which owes its word origin in 1555 to the Spanish word *huracán*, the Taino word *hurakán* [2] and ultimately refers to Huracan, a Mayan storm god [3]. Recent examples of notable hurricanes include Hurricane Katrina in 2005 (1,833 deaths / \$108 billion) [4] and Hurricane Sandy in 2012 (147 deaths and \$50 billion) [5].

Ignoring any potential time-varying climate change impacts, on average only two Atlantic basin storms make landfall in the U. S. per year and, on the surface, only induce a relatively modest amount of average damage (\$6.961 billion per storm) [1]. In terms of high risk states, Florida has witnessed the most landfalling storms (on average, less than one per year). Florida’s total gross domestic product (GDP) in 2013 was \$800 billion [6] yet the average landfalling storm caused only \$8.728 billion in damage [1], representing only a small fraction (1%) of its GDP. For the modern day residents of these states, why then does it appear to be much more dangerous than the average would seem to indicate? Why were ancient civilizations so worried, that they prayed to their gods to be spared?

The answer to this apparent discrepancy is in the nature of the events themselves. In a probabilistic sense, these catastrophic events are part of a “heavy-tailed” distribution. That is, while many events result in low-to-moderate amount of damages, a few, every now and then, cause significant damages. Traditional methods of looking at the frequency of events focus on statistics such as Probability of Exceedance (POE). However, much like computing a simple average, POE does not adequately capture the full extent of what is occurring and looking at this statistic alone would give a detached observer a false sense of security. To help quantify the true size of the damages in the tail, other statistical measures such standard deviation are often cited. For example, for the U. S. as a whole, historically, this value is \$19.879 billion, whereas for storms that make landfall in Florida, this value is \$25.476 billion [1]. While this statistic shows a better picture of what is happening, it is difficult to use it in any practical resiliency/planning/management context.

To better capture the true impacts of these heavy tail distributions, researchers have very recently developed a new method of probabilistic analysis called Buffered Probability of Exceedance (bPOE) [7] which gives important information on tails of these distributions. The bPOE concept has been introduced as a generalization of so called Buffered Probability of Failure (BPF) [8] used in reliability analysis. BPF was first introduced by R. T. Rockafellar during a presentation at the International Workshop on Engineering Risk Control and Optimization[9]. A simple high-level comparison between POE and bPOE can be stated as:

For a given set of data and its associated probability distribution, the “Probability of Exceedance” (POE) describes the likelihood that a specified threshold will be exceeded whereas “Buffered Probability of Exceedance” (bPOE) describes the chance that the average of the data points in the upper tail of the distribution will equate to a specified threshold.

bPOE thus combines both the concepts of probability and magnitude. In this paper we analyze tropical storm economic damage data and show how bPOE gives an important information accurately portraying both the probability and extent of the damages historically observed during landfalling tropical storms.

It is noted, that while the bPOE analysis presented herein is being applied to a discrete set of data, it can also be applied to analytical distribution functions. For example, some researchers suggest that tropical storm economic damages follow a Pareto (power-law) distribution [10], [11]. Thus, rather than using the data itself, a distribution function could be generated and then analyzed using bPOE to further look into the properties of the distribution tail. Alternatively, as the pool of landfalling storms is relatively small, combining the presented methodology with an analytic distribution function could be used to breakdown historical damage trends by regional, temporal, or storm intensity differences where only a few storms are available in the historical record.

Buffered Probability of Exceedance (bPOE)

Background

[7] defines bPOE as one minus the superdistribution function (see definition of superdistribution in [12]). [13] calls the superdistribution function by the Buffered Service Level (BSL) in terms of a server having some capacity and probabilistic demand. While [7] presents theoretical/mathematical properties of bPOE at a highly mathematical level, [13] uses relatively simple terms. Following [13], this section seeks to define bPOE and its associated properties but using terminology more understandable to environmental geoscience researchers.

To begin, bPOE owes its origin to the so called Buffered Probability of Failure (BPF) [8]. As compared to standard Probability of Failure, BPF was shown to have exceptional mathematical properties (e.g. it is a continuous function with respect to the threshold). In optimization problems, a constraint on BPF can be reduced to a convex constraint on the superquantile (Conditional Value-at-Risk). BPF is actually an inverse function of superquantile at point zero. The inverse function of superquantile itself can be shown to be a new distribution function [12].

Buffered Service Level and Buffered Probability of Exceedance

Suppose there is a tropical storm about to make landfall along a coast. Using historical economic damage data as a guide, we are interested in knowing the amount of damage (L) relative to some given threshold (W). The Probability of Exceedance (POE) for that threshold W is defined by

$$p_W(L) = P[L > W] . \quad (1)$$

The threshold W in $p_W(L)$ is commonly referred to in economics and finance as the “Value at Risk” (VaR). For example, there is a 25% chance that economic damage L from a landfalling storm will exceed the threshold $W=\$4.875$ billion. Two characteristics then describe this damage distribution: W and $p_W(L) = 0.25$. Note that these two characteristics do not provide any information about the magnitude

of damages of L exceeding W . Sometimes, the information on the possible values of L exceeding the threshold may be quite important. For instance, in nuclear engineering, the quantity of radiation released (when a component fails to meet its design specification) may be very important!

We can ask the question: If the damage exceeds W , what is the maximum amount of damage a community may experience? Suppose that the damages L are normally distributed; in this case (theoretically), there is no limit on the amount of damages experienced for all possible values of L . You may say that the normal distribution is a pure mathematical concept which does not really exist in practice. Unfortunately, quite frequently in practical applications, damages have so-called “heavy tails”, i.e., damages can be very high with very small probabilities (as is the typical case in environmental applications). We can pose a less demanding question: if a damage exceeds W , what is the average amount of damage expected? We say that the economic damages $V = E[L|L > W]$ measures *on average* the damages exceeding W , where $E[L|L > W]$ is the conditional expectation that the damage L exceeds W . By definition, economic damage V has a bPOE equal to $\bar{p}_V(L) = p_W(L)$. The threshold V in $\bar{p}_V(L)$ is commonly referred to in economics and finance as the “Conditional Value at Risk” (CVaR). The formal definition of bPOE [7] can be written as:

Definition 1.

Assume L is a continuously distributed random variable and V is a threshold in the range, $E[L] \leq V \leq \sup L$. Suppose that for a different threshold W which is a function of V , the following equation is satisfied:

$$V = E[L|L > W(V)] .$$

Then by definition, bPOE equals

$$\bar{p}_V(L) = P[L > W(V)] .$$

bPOE answers the question: what is the fraction of the largest damages which, on average, equals V ? The random excess damage over the threshold W , by definition, equals:

$$[L - W(V)]^+ = \max [0, L - W(V)] . \quad (2)$$

The conditional expectation $V = E[L|L > W(V)]$ is related to the expected excess $E[L - W(V)]^+$ as follows:

$$V = E[L|L > W(V)] = W(V) + \frac{E[L - W(V)]^+}{P[L > W(V)]} = W(V) + \frac{E[L - W(V)]^+}{\bar{p}_V(L)} . \quad (3)$$

We will call V the *Buffered Threshold*; the buffered threshold V always exceeds $W(V)$.

Therefore, with bPOE, $\bar{p}_V(L)$, we can express the expected excess, as:

$$E[L - W(V)]^+ = \bar{p}_V(L)[V - W(V)] . \quad (4)$$

Equivalently we have

$$\bar{p}_V(L) = \frac{E[L - W(V)]^+}{V - W(V)} . \quad (5)$$

The expected excess $E[L - W(V)]^+$ is quite popular in various engineering applications; in financial applications it is called Partial Moment One, in stochastic programming it is called Integrated Chance. In

statistics, Quantile Regression is based on the Koenker and Bassett error function, which is actually a weighted sum of two expected excesses of a residual. Another application, directly related to the topic of tropical storm damage, would be the use of expected excess to determine the minimum amount to charge for insurance premiums. For example, assume that an insurance company is creating a policy to cover tropical storm damage losses in excess of some deductible, $W(V)$. Then, amount of money in premiums that should be collected would be the expected excess, $E[L - W(V)]^+$. While some events may incur payable losses below this value and some above, over a large number of insured events, in terms of losses versus collections, the company will break even if they charge this amount. If the company charges more than this amount in premiums, they will earn money to fund their operations or profit. If the amount charged is less, then the company will go out of business as payouts will exceed premiums collected. Therefore from these examples it can be seen how various engineering and mathematical areas can benefit from the concept of buffered probability of exceedance, which interprets the expected excess as a constant damage with some probability.

Furthermore, for a continuously distributed L , we have

$$(1 - \bar{p}_V(L)) \cdot E[L|L \leq W(V)] + \bar{p}_V(L) \cdot E[L|L > W(V)] = E[L]. \quad (6)$$

Calculation and Optimization of Buffered Probability of Exceedance

While (5) can be used for the calculation of the bPOE, it can also be used as an alternative definition of the bPOE. A significant advantage of this formula comes from the following minimization representation (which is stated in other format in [7], [14]):

$$\bar{p}_V(L) = \begin{cases} 0 & V \geq \sup L \\ \min_{W < V} \frac{E[L - W]^+}{V - W} & E[L] < V < \sup L \\ 1 & \text{otherwise} \end{cases} . \quad (7)$$

An optimal $W(V)$ in (7) satisfies the equation $V = E[L|L > W(V)]$, the proof of which for general distributions is shown in [7]. This minimum formula for the bPOE is very similar to the minimum formula for CVaR, which was established in [15]. The minimum in this formula is achieved on VaR with the 1-bPOE confidence level.

The advantage of (7) can be demonstrated with the *convex optimization problem formulation* obtained in [14] and [7]. Let us consider that L is a linear function depending on control vector $w \in R^n$ with a random vector of coefficients, θ , i.e., $L = w^T \theta$. We are interested in minimizing the bPOE with $V=0$, which corresponds to *Buffered Probability of Failure* in [8]. This minimization problem can be formulated as the following convex minimization problem, which can be reduced to *linear programming*:

$$\min_{w \in R^n} E[w^T \theta + 1]^+ . \quad (8)$$

bPOE Calculation Algorithm

As stated in (7), the bPOE can be determined through the solution of the one dimensional optimization problem. While this approach is efficiently stated, it has two distinct disadvantages. First, the typical

engineer or applied scientist who could find benefit in the bPOE, is not likely to be familiar with such formulations such that they could write their own computer program to solve it nor would they likely have experience with optimization software. Second, the optimization problem provides a single solution to a single set of sample data and bPOE/thresholds. However, for many applications, one is not interested in a single bPOE but rather many over a range of values simultaneously (25%, 50%, 75%, 99%, 99.9%, etc.). In such a case, the entire set of calculations necessary to solve the optimization problem would need to be re-computed for each bPOE which is needlessly computationally inefficient.

To address these issues, the problem is reformulated in terms of conditional-value-at-risk (CVaR) (V) and bPOE (\bar{p}_V) calculated at every sample data point (atom). While also not a common approach outside of financial engineering, these statistics can be simply calculated by first sorting the data and then computing an average. With V and \bar{p}_V then computed at each atom, it just becomes a matter of interpolating between the atoms to get the exact \bar{p}_V from V (or vice versa). The values of V and \bar{p}_V at the atoms for a set of sample data can be pre-computed so that many inter-atom values of V and \bar{p}_V can be determined very efficiently.

Let L be a discrete random variable with atoms x_1, x_2, \dots, x_N and probabilities p_1, p_2, \dots, p_N , where $x_i \leq x_{i+1}, i = 1, 2, \dots, N - 1$. Denoting confidence levels α_j as

$$\alpha_j = \begin{cases} 0 & \text{for } j = 0 \\ \sum_{i=1}^j p_i & \text{for } j = 1, 2, \dots, N - 1 \end{cases}. \quad (9)$$

CVaR can be written as [15]

$$\bar{x}_j = \frac{1}{1 - \alpha_j} \sum_{i=j+1}^N p_i x_i \quad \text{for } j = 0, 1, \dots, N - 1. \quad (10)$$

By definition, bPOE equals

$$\bar{p}_{\bar{x}_j} = 1 - \alpha_j \quad \text{for } j = 0, 1, \dots, N - 1. \quad (11)$$

[7] showed that 1/bPOE is a piecewise linear function of CVaR with the following interpolation formula providing inter-atom values

$$\bar{p}(\mu \bar{x}_j + (1 - \mu) \bar{x}_{j+1}; L) = \left(\frac{\mu}{\bar{p}(\bar{x}_j; L)} + \frac{1-\mu}{\bar{p}(\bar{x}_{j+1}; L)} \right)^{-1}. \quad (12)$$

Earlier efforts [16] developed a similar formula to provide CVaR as a function of bPOE.

Thus, the formerly difficult to understand and computationally inefficient optimization problem to calculate bPOE merely becomes a few simple steps involving sorting and a series of sums/divisions (which can be pre-computed) followed by a simple linear interpolation to any combination of bPOE and CVaR desired. This follows from the fact that CVaR and 1/bPOE are piece-wise linear functions for discrete distributions. Furthermore, to make the algorithm more suited to parallel computing algorithms, the data are first sorted from highest to lowest values. Then, with this small modification

and a slight algebraic modification in the algorithm, the calculation of CVaR becomes a simple “prefix sum” (see Appendix).

To illustrate how Equations (9)-(12) can be used to compute CVaR and BPOE, three simple examples are presented: Example #1 consists of a sample dataset in which all data values are unique and equally probable. Example #2 is similar to Example #1 in that all data values are equally probable, however, in this example, one data value is repeated. In Example #3, all data values are again unique, however, one data point is twice as likely to occur as the other data points.

Example #1 – All x values are unique and equally probable

If we have a sample dataset, one case would be that all values are unique and we have a distribution of data with equal probability. Let us give an example with 4 outcomes. Suppose that

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 5, \quad x_4 = 7$$

and

$$p_1 = 0.25, \quad p_2 = 0.25, \quad p_3 = 0.25, \quad p_4 = 0.25.$$

Then for confidence levels α_j

$$\begin{aligned}\alpha_0 &= 0, \\ \alpha_1 &= 0.25, \\ \alpha_2 &= 0.25 + 0.25 = 0.5, \\ \alpha_3 &= 0.25 + 0.25 + 0.25 = 0.75.\end{aligned}$$

CVaR equals

$$\begin{aligned}\bar{x}_0 &= \frac{1}{1-0} (0.25 * 1 + 0.25 * 2 + 0.25 * 5 + 0.25 * 7) = 3.75, \\ \bar{x}_1 &= \frac{1}{1-0.25} (0.25 * 2 + 0.25 * 5 + 0.25 * 7) = 4.66, \\ \bar{x}_2 &= \frac{1}{1-0.5} (0.25 * 5 + 0.25 * 7) = 6, \\ \bar{x}_3 &= \frac{1}{1-0.75} (0.25 * 7) = 7.\end{aligned}$$

With a corresponding bPOE of

$$\begin{aligned}\bar{p}_0 &= 1 - 0 = 1, \\ \bar{p}_1 &= 1 - 0.25 = 0.75, \\ \bar{p}_2 &= 1 - 0.5 = 0.5, \\ \bar{p}_3 &= 1 - 0.75 = 0.25.\end{aligned}$$

Thus, we can see, \bar{x}_2 equals the average of the largest two outcomes $x_3 = 5$ and $x_4 = 7$. We see also that $\bar{p}_{\bar{x}_2} = 1 - \alpha_2 = 1 - 0.5 = 0.5$. So the bPOE for the threshold $\bar{x}_2 = 6$ equals the probability of the tail (0.5) with the expected value in this tail equal to 6. A complete comparison between POE and bPOE for the example dataset can be seen in Figure 2.

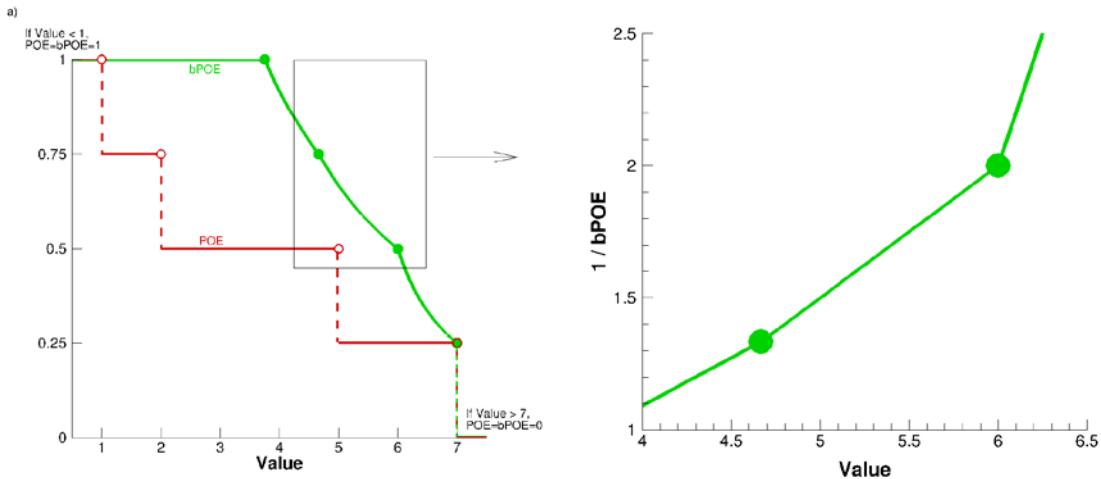


Figure 1 a) A comparison between the POE (red) and bPOE (green) for the example dataset. b) A “zoom in” of the values between 4 and 6.5 demonstrating that $1/bPOE$ is linear between atoms.

Example #2 – One x value is repeated but all are equally probable

If we have a sample dataset, one case would be that some values are duplicated but we still have a distribution of data with equal probability. Let us give an example with 5 outcomes. Suppose that

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 5, \quad x_4 = 5 \quad x_5 = 7$$

and

$$p_1 = 0.2, \quad p_2 = 0.2, \quad p_3 = 0.2, \quad p_4 = 0.2 \quad p_5 = 0.2.$$

Then for confidence levels α_j

$$\begin{aligned}\alpha_0 &= 0, \\ \alpha_1 &= 0.2, \\ \alpha_2 &= 0.2 + 0.2 = 0.4, \\ \alpha_3 &= 0.2 + 0.2 + 0.2 = 0.6, \\ \alpha_4 &= 0.2 + 0.2 + 0.2 + 0.2 = 0.8.\end{aligned}$$

CVaR equals

$$\begin{aligned}\bar{x}_0 &= \frac{1}{1-0} (0.2 * 1 + 0.2 * 2 + 0.2 * 5 + 0.2 * 5 + 0.2 * 7) = 4, \\ \bar{x}_1 &= \frac{1}{1-0.2} (0.2 * 2 + 0.2 * 5 + 0.2 * 5 + 0.2 * 7) = 4.75, \\ \bar{x}_2 &= \frac{1}{1-0.4} (0.2 * 5 + 0.2 * 5 + 0.2 * 7) = 5.6\bar{6}, \\ \bar{x}_3 &= \frac{1}{1-0.6} (0.2 * 5 + 0.2 * 7) = 6, \\ \bar{x}_4 &= \frac{1}{1-0.8} (0.2 * 7) = 7.\end{aligned}$$

With a corresponding bPOE of

$$\begin{aligned}\bar{p}_0 &= 1 - 0 = 1, \\ \bar{p}_1 &= 1 - 0.2 = 0.8, \\ \bar{p}_2 &= 1 - 0.4 = 0.6, \\ \bar{p}_3 &= 1 - 0.6 = 0.4, \\ \bar{p}_4 &= 1 - 0.8 = 0.2.\end{aligned}$$

Example #3 – All x values are unique and but not equally probably

If we have a sample dataset, one case would be that the values are unique but we the distribution of data has an unequal probability. Let us give an example with 4 outcomes (Figure 1). Suppose that

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 5, \quad x_4 = 7$$

and

$$p_1 = 0.2, \quad p_2 = 0.2, \quad p_3 = 0.4, \quad p_4 = 0.2.$$

Then for confidence levels α_j

$$\begin{aligned}\alpha_0 &= 0, \\ \alpha_1 &= 0.2, \\ \alpha_2 &= 0.2 + 0.2 = 0.4, \\ \alpha_3 &= 0.2 + 0.2 + 0.4 = 0.8.\end{aligned}$$

CVaR equals

$$\begin{aligned}\bar{x}_0 &= \frac{1}{1-0} (0.2 * 1 + 0.2 * 2 + 0.4 * 5 + 0.2 * 7) = 4, \\ \bar{x}_1 &= \frac{1}{1-0.2} (0.2 * 2 + 0.4 * 5 + 0.2 * 7) = \frac{14}{3} = 4.75, \\ \bar{x}_2 &= \frac{1}{1-0.4} (0.4 * 5 + 0.2 * 7) = 5.6\bar{6}, \\ \bar{x}_3 &= \frac{1}{1-0.8} (0.2 * 7) = 7.\end{aligned}$$

With a corresponding bPOE of

$$\begin{aligned}\bar{p}_0 &= 1 - 0 = 1, \\ \bar{p}_1 &= 1 - 0.2 = 0.8, \\ \bar{p}_2 &= 1 - 0.4 = 0.6, \\ \bar{p}_3 &= 1 - 0.8 = 0.2.\end{aligned}$$

Suppose we are interested in determining the value of \bar{p}_V where $\bar{x} = 6$, then by Equation (11),

$$\mu * 5.6\bar{6} + (1 - \mu) * 7 = 6 \quad \text{or} \quad \mu = 0.75$$

and

$$\bar{p}(6; L) = \left(\frac{0.75}{0.6} + \frac{1 - 0.75}{0.2} \right)^{-1} = 0.4$$

which is exactly \bar{x}_3 as shown in Example 2.

Tropical Storm Damage Data

Using the concepts of POE and bPOE presented herein, an analysis is presented to answer the practical question:

If a tropical storm is approaching the U. S. coastline and landfall is imminent, what would the expected total economic damage be?

To answer this question, three assumptions are made:

1. It is assumed that there are no long-term trends in hurricane characteristics which would impact future events (e.g. there are no potential changes in storm tracks / intensity caused by climate change). While this would be an interesting research topic, it is outside of the scope of this effort. The historical record shows that a total of 242 storms from 1900 (Galveston)-2012 (Sandy) with landfalls occurring most often in Florida, Texas, Louisiana, and North Carolina (Table 1). The most damaging storm being the Great Miami Storm of 1926 which made landfall in Florida as a Category 4 storm with 145 mph winds and inflicting \$203B in damage [1].
2. In order to compare storms which made landfall of such an extended period of time, it is assumed that the methods described in [17] produce sufficiently accurate damage estimates. Those authors normalize economic damages obtained from several sources to reflect current inflation, wealth, and population from what existed at the time of the actual storm activity. More specifically, damage data for storms making landfall all the way to 2013 was directly obtained from the ICAT Damage Estimator[1] and analyzed in constant 2014 dollars.
3. It is assumed that the probability of occurrence of each of the 242 storms was equal.

Table 1 A summary of the number of tropical storms grouped by state, month, and strength at the time of landfall. Storm strength is defined using the Saffir-Simpson Hurricane Wind Scale.

State	Count	Month	Count	Strength	Count
Alabama	4	May	2	Tropical Depression	4
Florida	83	Jun	20	Tropical Storm	63
Georgia	3	Jul	20	Category 1	59
Louisiana	35	Aug	61	Category 2	43
Massachusetts	3	Sep	100	Category 3	52
Maine	1	Oct	33	Category 4	18
Mississippi	6	Nov	6	Category 5	3
North Carolina	31				
New York	9				
Rhode Island	1				
South Carolina	13				
Texas	53				

Application of bPOE to Tropical Storm Damage Data

Using the set of tropical storm economic damage data described previously, a comparison between the POE and bPOE is shown in Figure 2. The bPOE is calculated using the algorithm presented in the Appendix. For POE, values are calculated at the damage levels (atoms) associated for each of the 242 storms. Note how the line in Figure 2 which represents POE is not smooth.

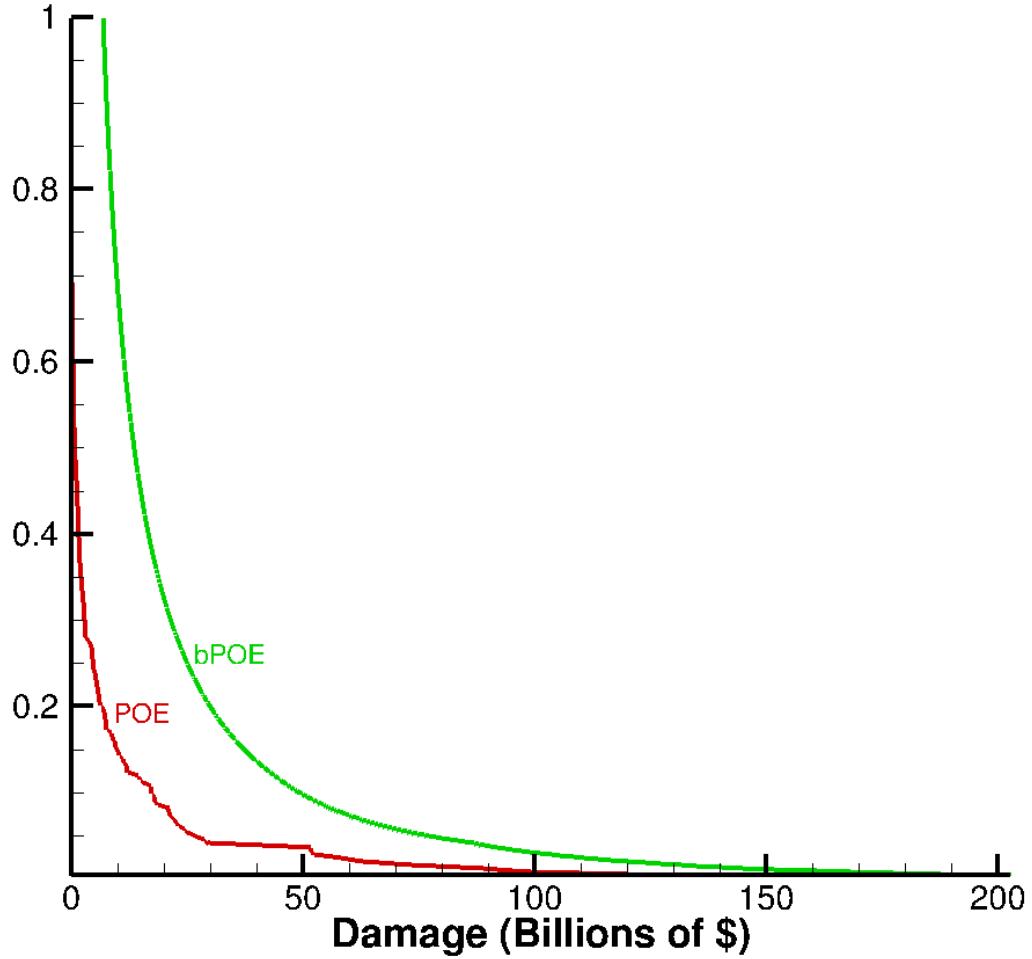


Figure 2 A comparison between the probability of exceedance (POE) (solid red line) and buffered probability of exceedance (bPOE) (solid green line) for damage in the upper tail of the distribution caused by all storms that made landfall during 1900-2013. For the POE, the x-axis represents the Value-at-Risk (VaR) which is denoted by a W herein, while for the bPOE, the x-axis represents the Conditional-Value-at-Risk which is denoted by a V herein.

Now, let us denote the following symbols to differentiate the differences between the expected value in the upper tail, with the value in the lower tail:

$$\begin{aligned}\overline{CVaR} &= E[L|L > W(V)] = \text{Expected value in the upper tail,} \\ \underline{CVaR} &= E[L|L \leq W(V)] = \text{Expected value in the lower tail.}\end{aligned}$$

Then, further interpretation of 50% probability threshold could be written as:

If a storm is about to make landfall, there is a 50% probability that the total economic damages caused by the storm will exceed \$0.910 billion dollars. If the damages do exceed \$0.910, then the average amount of damage will be $\overline{CVaR} = \$13.693$ billion dollars. If the damages do not exceed this amount, then the average amount of damage will be $\underline{CVaR} = \$0.228$ billion dollars.

The large difference between \overline{CVaR} and \underline{CVaR} shows how storm damage is in essence a “binary event”. That is, an all-or-nothing scenario occurs, significant damage or minimal damage.

As mentioned previously, Equation (4), the difference between \overline{CVaR} and VaR multiplied by the BPOE for a fixed level is referred to as “expected excess”. For the 50% probability level, Figure 3 demonstrates how this value works out to be \$6.392 billion dollars. Returning to our insurance example mentioned previously, for a \$0.910 billion dollar deductible, the insurance company needs to collect at least \$6.392 billion dollars in premiums to ensure that it breaks even over a large number of tropical storms.

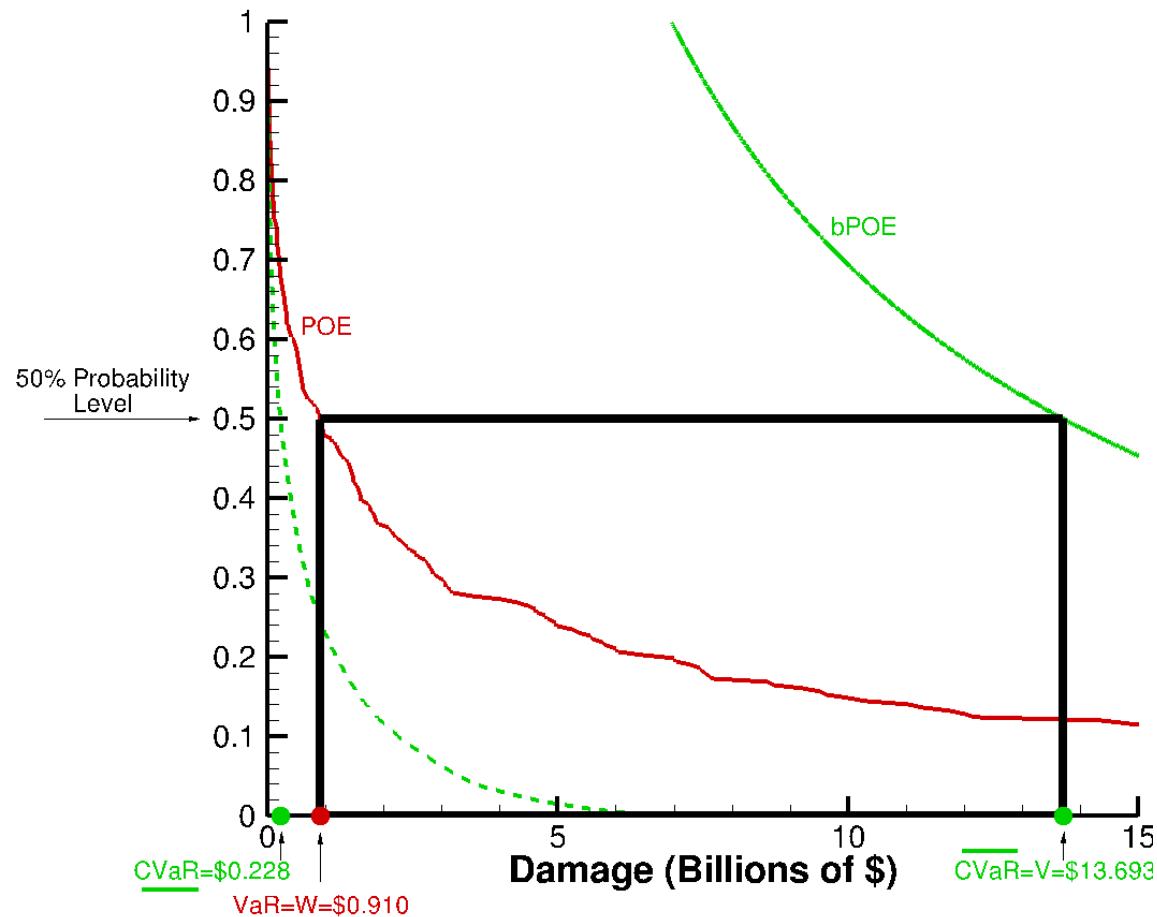


Figure 3 For the 50% probability level, a “zoom in” of the probability of exceedance (POE) (red line) and buffered probability of exceedance (bPOE) (solid green line) for damage in the upper tail along with a line representing the expected value in the lower tail (dashed green line). Expected excess, $E[L - W(V)]^+$, is then the area outlined by the box and is equal to $(V-W) * \bar{p}_V = (\$13.693 \text{ billion} - \$0.910 \text{ billion}) * 0.5 = \6.392 billion . For the POE, the x-axis represents the Value-at-Risk(VaR) in the upper tail which is denoted by a W herein, while for the bPOE, the x-axis represents the Conditional-Value-at-Risk in the upper tail (\overline{CVaR}) which is denoted by a V herein. For the line representing the expected value in the lower tail, the x-axis represents the Conditional-Value-at-Risk in the upper tail ($CVaR$).

Continuing with this example, through further interpretation of the dataset, specific values at risk for various probability thresholds can be tabulated (Table 2). Several interesting things fall out of this analysis. First, while it was expected that $V > W(V)$, the true extent of this conditional expected excess damage is quite significant not just in terms of percentages but also in terms of real dollars. Taking the 25% probably row (marked by bold face in Table 2) for example, there is a 25% chance that a landfalling storm will cause damage exceeding \$4.785 billion in damages. If exceeded, \overline{CVaR} is 526% (\$20.391

billion!) greater than VaR. Second, as was noted when looking at the 50% probability level earlier, storm damages prove to be binary events for ALL probability levels.

Table 2 A comparison between the value at risk (VaR) and conditional value at risk (CVaR) for various probability thresholds. Dollars are reported in billions. For a given confidence level, VaR is interpolated from surrounding atoms.

$1-\alpha$	VaR	$\overline{CVaR}(V)$	$\underline{CVaR}(V)$	$\overline{CVaR}(V) - \underline{CVaR}(V)$	Expected Excess
1%	\$93.808	\$154.272	\$5.433	\$60.464	\$0.605
5%	\$26.465	\$76.703	\$3.289	\$50.238	\$2.512
10%	\$17.410	\$49.191	\$2.268	\$31.781	\$3.178
25%	\$4.785	\$25.176	\$0.889	\$20.391	\$5.098
50%	\$0.910	\$13.693	\$0.228	\$12.783	\$6.392
75%	\$0.130	\$9.265	\$0.047	\$9.135	\$6.851
99%	\$0.000	\$7.031	\$0.000	\$7.030	\$6.960

Repeating this example, but from a slightly different point of view, the analysis can be inverted to compare POE with bPOE for various damage thresholds (Table 3). It is observed that bPOE is significantly higher than POE which is consistent with [7] that showed, with reasonable assumptions, that bPOE is at least two times larger than POE.

Let us now discuss how damage can be decomposed into a probability distribution with just two atoms as shown in Equation (6):

$$\text{Lower Tail Contribution} + \text{Upper Tail Contribution} = \text{Total Expected Damage}$$

$$0.903 * (\$2.311 \text{ billion}) + 0.097 * (\$50 \text{ billion}) = \$2.087 + \$4.874 \text{ billion} = \$6.961 \text{ billion.}$$

Upper Tail Contribution: Examining the \$50 billion threshold (marked by bold face in Table 3), the POE equals 3.7%. Thus damages exceeding \$50 billion are rare and it looks like these excess damages may not impact significantly the overall average damage. The bPOE equals 9.7%; therefore, the expected damage in the 9.7% upper tail equals $\$50 * 0.097 = \4.874 billion.

Lower Tail Contribution: The lower tail has probability $100\% - 9.7\% = 90.3\%$. The expected damage in the lower tail equals the difference of total expected damage and expected damage in the upper tail: $\$6.961 - \$4.874 = \$2.087$ billion. So, the conditional expected damage in the lower tail equals the expected damage in this tail divided by probability of this tail: $\$2.087 / 0.903 = \2.311 billion.

Thus it can be seen that the expected damage for the original damage distribution containing 242 observations can be “packed” into a discrete distribution with two atoms: lower tail expected damage \$2.087 billion is the product of \$2.311 billion (lower tail conditional expected damage) and probability of this tail 0.903 and the upper tail expected damage \$4.874 billion is the product of \$50 billion (upper tail conditional expected damage) and probability 0.097. Such representation shows importance of damages in the upper tail of the distribution.

The lower tail contains $242 * 0.903 \approx 219$ observations and the upper tail, $242 - 219 = 23$ observations. By ordering damages in the dataset, this split between the upper and lower tails at the 23 observation occurs at a damage of \$18.130 billion. Again, referring to our insurance example, suppose an insurance

fund is setup assuming a deductible of \$18.130 billion (VaR). Then, according to Equation (4), the expected excess in the upper tail (the minimum amount to charge in premiums) equals the product of bPOE = 0.097 and difference of the conditional expected damage in the upper tail, \$50 billion (CVaR), and the deductible \$18.130 billion:

$$\text{Expected excess} = (\$50 - \$18.130) * 0.097 = \$3.091 \text{ billion.}$$

Table 3 A comparison between the probability of exceedance (p_w) and the buffered probability of exceedance (\bar{p}_v) for various damage thresholds. Dollars are reported in billions.

Damage Threshold	p_w	\bar{p}_v
\$100	0.8%	3.0%
\$75	1.6%	5.2%
\$50	3.7%	9.7%
\$25	5.4%	25.2%
\$10	14.8%	69.4%
\$5	23.9%	100.0%
\$1	47.8%	100.0%

While the original set of data is somewhat limited, it is still possible to subdivide the storms into groups based on landfalling state, month and strength and end up with sample sizes reasonably large enough to warrant further analysis. An example of the expected excess damage of these different subdivisions is shown in Table 4. In addition to Florida being in the crosshairs of a landfalling storms most often, the excess damage is also significantly larger than other high risk states. In particular, for a storm that makes landfall and causes damages exceeding the 50% probability level, Florida can expect 3.3x more in damages than North Carolina. Historically, September is the most active month for storm landfall and also proves to be the month with the highest excess damage although the other active months are not that significantly different. In terms of storm strength, the more intense storms prove to have the larger excess damages, which make sense as the damages overall would be expected to be larger for larger storms.

Table 4 A summary of the expected excess damage at the 50% confidence level for landfalling tropical storms by state, month, and strength. Values are only shown for groups of landfalling storms which have a sample size count ≥ 30 (see Table 1). Storm strengths are defined by the Saffir-Simpson Hurricane Wind Scale. Dollars are reported in billions.

State	Expected Excess	Month	Expected Excess	Strength	Expected Excess
Florida	\$8.111	Aug	\$7.486	Tropical Storm	\$0.991
Louisiana	\$4.658	Sep	\$8.057	Category 1	\$2.249
North Carolina	\$2.430	Oct	\$6.227	Category 2	\$2.932
Texas	\$6.733			Category 3	\$6.967

Summary

Relying on POE alone or in combination with other statistical measures such as mean or standard deviation does not give a clear picture of data which have heavy tailed distributions. In this paper, the new concept of bPOE has been explained and advanced. Through averaging of data in the tail, bPOE combines the elements of both probability as well as magnitude into a single statistic. The concept is

then reduced to a simple algorithmic form which requires only sorting of the data combined with linear interpolation to get the bPOE from the CVaR or vice versa. While the bPOE analysis presented herein is being applied to a discrete set of data, it can also be applied to analytical distribution functions. Finally, using a set of normalized economic damage data for tropical storms making landfall in the Atlantic and Gulf coasts of the U. S. from 1900-2013, an analysis using bPOE has been presented.

For instance, analysis of the data shows that for a landfalling storm, for a \$50 billion threshold, the POE is 3.7% while the bPOE is 9.7%. Consistent with analysis of data in other fields, the risk observed by bPOE is more than double that of POE. For a 25% probability, the value-at-risk determined using POE is \$4.785 billion, while the conditional-value-at-risk (average of damage received) determined by the bPOE is \$25.176. For a fixed probability, this \$20.391 billion difference between the damage associated with POE and bPOE leads to an expected excess of $0.25 * \$20.391 = \5.098 billion and represents the damage above and beyond the damage indicated by the POE. Subdividing the economic damage data by landfall state, at the 50% probability level, Florida, the state most often hit by tropical storms has the highest value for expected excess at \$8.111 billion.

The concepts of bPOE and CVaR also appear to be highly informative to a wide variety of other environmental geoscience applications (e.g. [18]). A preliminary analysis of salinity in Apalachicola Bay, FL shows that salinity values are significantly closer to a key threshold at which oyster larvae cannot survive possibly better explaining unexplained recently unexplained oyster die offs.. Another application to a dataset of concentrations of particular matter in the air in Bogota, Colombia shows that when values enter the unhealthy range, they are very unhealthy, much more so that the POE analysis would indicate which has potentially significant implications to how air quality regulations are determined.

References

- [1] "ICAT Damage Estimator," 26-Aug-2014. [Online]. Available: <http://www.icatdamageestimator.com/>.
- [2] "hurricane," *Merriam-Webster.com*. 26-Aug-2014.
- [3] M., Taube, Karl Miller, *The gods and symbols of ancient Mexico and the Maya: an illustrated dictionary of Mesoamerican religion*. London: Thames and Hudson, 1997.
- [4] R. D. Knabb, J. R. Rhome, and D. P. Brown, "Hurricane Katrina: 23-30 August 2005," National Hurricane Center, Tropical Cyclone Report, Sep. 2011.
- [5] E. S. Blake, T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven II, "Hurricane Sandy (AL182012): 22-29 October 2012," National Hurricane Center, Tropical Cyclone Report, Feb. 2013.
- [6] "Total Gross Domestic Product by State for Florida," *Federal Reserve Bank of St. Louis*. [Online]. Available: <http://research.stlouisfed.org/fred2/series/FLNGSP>.
- [7] A. Mafusalov and S. Uryasev, "Buffered Probability of Exceedance: Mathematical Properties and Optimization Algorithms," Department of Industrial and Systems Engineering, University of Florida, Research Report 2014-1, Sep. 2014.
- [8] R. T. Rockafellar and J. O. Royset, "On buffered failure probability in design and optimization of structures," *Reliability Engineering & System Safety*, vol. 95, no. 5, pp. 499–510, May 2010.
- [9] R. T. Rockafellar, "Safeguarding Strategies in Risky Optimization," presented at the International Workshop on Engineering Risk Control and Optimization, Gainesville, Florida, Feb-2009.
- [10] W. D. Nordhaus, "The Economics of Hurricanes and Implications of Global Warming," *Climate Change Economics*, vol. 01, no. 01, pp. 1–20, May 2010.
- [11] C. Blackwell, "Investigating Power Laws in Hurricane Damages," *SSRN Electronic Journal*, 2012.
- [12] R. T. Rockafellar and J. O. Royset, "Random variables, monotone relations, and convex analysis," *Mathematical Programming*, Aug. 2014.
- [13] S. Uryasev, "Buffered Probability of Exceedance and Buffered Service Level: Definition and Properties," Department of Industrial and Systems Engineering, University of Florida, Research Report 2014-3, Oct. 2014.
- [14] M. Norton and S. Uryasev, "Maximization of AUC and Buffered AUC in Classification," Department of Industrial and Systems Engineering, University of Florida, Research Report 2014-2, Oct. 2014.
- [15] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, no. 3, pp. 21–41, 2000.
- [16] *Portfolio Safeguard*. American Optimal Decisions, Inc.
- [17] R. A. Pielke, J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, "Normalized Hurricane Damage in the United States: 1900–2005," *Natural Hazards Review*, vol. 9, no. 1, pp. 29–42, Feb. 2008.
- [18] G. M. Yamout, K. Hatfield, and H. E. Romeijn, "Comparison of new conditional value-at-risk-based management models for optimal allocation of uncertain water supplies: CVaR MODELS FOR UNCERTAIN WATER SUPPLIES OPTIMAL ALLOCATION," *Water Resources Research*, vol. 43, no. 7, p. n/a–n/a, Jul. 2007.

Appendix: bPOE Algorithm

INPUT:

N data points (x):

$$[x_1, x_2, \dots, x_N],$$

their associated probability of occurrence (p):

$$[p_1, p_2, \dots, p_N],$$

and the desired:

a) \bar{p}_v (Buffered Probability of Exceedance-bPOE) -or- b) V (Conditional Value at Risk-CVaR)

OUTPUT:

The corresponding:

a) V -or- b) \bar{p}_v

PROCEDURE:

Step 1) Combine x and p together as a 2-tuple (ordered pair):

$$[(x_1, p_1), (x_2, p_2), \dots, (x_N, p_N)]$$

Step 2) Sort the elements of (x, p) by x from large to small ($x_1 > x_2, x_2 > x_3$, etc.):

$$[(x_1, p_1), (x_2, p_2), \dots, (x_N, p_N)]_{sorted}$$

Step 3) Calculate the prefix sum of the individual items of each element:

$$\left[(x_1, p_1), (x_1 + x_2, p_1 + p_2), (x_1 + x_2 + x_3, p_1 + p_2 + p_3), \dots, \left(\sum_{i=1}^N x_i, \sum_{i=1}^N p_i \right) \right]$$

Step 4) Divide the first item in each element by its ranked order ($1^{st}, 2^{nd}, 3^{rd}$, etc.):

$$\left[\left(\frac{x_1}{1}, p_1 \right), \left(\frac{x_1 + x_2}{2}, p_1 + p_2 \right), \left(\frac{x_1 + x_2 + x_3}{3}, p_1 + p_2 + p_3 \right), \dots, \left(\frac{1}{N} \sum_{i=1}^N x_i, \sum_{i=1}^N p_i \right) \right]$$

The resulting vector consists of V and \bar{p}_v for the input data points (d):

$$[(V_1, \bar{p}_{v1}), (V_2, \bar{p}_{v2}), \dots, (V_N, \bar{p}_{vN})]$$

Step 5) To obtain arbitrary values of V or \bar{p}_v , let $\hat{p}_v = \frac{1}{\bar{p}_v}$, then the resulting vector represents a piecewise linear function:

$$[(V_1, \hat{p}_{v1}), (V_2, \hat{p}_{v2}), \dots, (V_N, \hat{p}_{vN})]$$

a) Linearly interpolate V from the vector and the input \hat{p}_v :

$$V = \begin{cases} Indeterminant & \hat{p}_{vN} > \hat{p}_v \geq 1 \\ V_i + (V_{i+1} - V_i) \frac{(\hat{p}_v - \hat{p}_{vi})}{(\hat{p}_{vi+1} - \hat{p}_{vi})} & \hat{p}_{vi} \geq \hat{p}_v \geq \hat{p}_{vi+1} \quad \text{where } i = 1 \dots (N-1) \\ V_1 & \infty > \hat{p}_v > \hat{p}_{v1} \end{cases}$$

-or-

b) Linearly interpolate \hat{p}_v from the vector and the input V :

$$\hat{p}_v = \frac{1}{\bar{p}_v} = \begin{cases} 1 & V_N > V > -\infty \\ \hat{p}_{vi} + (\hat{p}_{vi+1} - \hat{p}_{vi}) \frac{(V - V_i)}{(V_{i+1} - V_i)} & V_i \geq V \geq V_{i+1} \quad \text{where } i = 1 \dots (N-1) \\ 0 & \infty > V > V_1 \end{cases}$$