Analysis of Tropical Storm Damage using Buffered Probability of Exceedance

Justin R. Davis[1] and Stan Uryasev[2]

[1]justin.r.davis@essie.ufl.edu / 352-294-7810
Engineering School of Sustainable Infrastructure & Environment
University of Florida, Gainesville, FL 32611

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

First Draft: August 17, 2015; This Draft: February 8, 2016

Abstract

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. bPOE is demonstrated using tropical storm damage along the Atlantic and Gulf of Mexico coasts of the U. S. from 1900-2013 (average yearly damage $14.776 billion). It can be shown, under general assumptions, that bPOE is more than double the probability of exceedance (POE). For instance, with a $50 billion threshold of damages per year, POE is 10.6% while the bPOE is 26.1%. We also considered expected excess (EE) over some threshold (deductible), which is a minimum premium that insurer should charge per year. We demonstrate that EE equals the difference between the damage associated with POE and bPOE multiplied by bPOE. For instance, for a 25% tail probability, the value-at-risk (quantile) determined using POE is $11.181 billion, while the conditional-value-at-risk (average damage in excess of the quantile) determined by the bPOE is $51.753 billion. This $40.572 difference leads to EE of 0.25*$40.572=$10.143 billion over the threshold $11.181 billion. Furthermore, 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 EE 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 in the 114 years 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 $\geq$ 74 mph) are referred to as "hurricanes". Recent examples of notable hurricanes include Hurricane Katrina in 2005 (1,833 deaths / $108 billion) [2] and Hurricane Sandy in 2012 (147 deaths and $50 billion) [3].

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 ($14.776 billion per year) [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 [4] 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?

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, the standard deviation of yearly damage is $31.774 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) [5] which gives important information on tails of these distributions. The bPOE concept has been introduced as a generalization of the so called Buffered Probability of Failure (BPF) [6] 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[7]. 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) only describes the likelihood that some specified threshold will be**

exceeded. However, "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.

This can be restated in another way as: the threshold used in the calculation of the POE provides no other information. While the threshold used in the calculation of the bPOE is strictly specified to be the averaged of the data points in the tail. Thus, while the calculation of both the POE and bPOE use thresholds to yield probabilities, the bPOE also yields important information about the shape of the distribution: the magnitude of the tail. Additionally, if a common threshold value is used for the calculation of both POE and bPOE, it can be shown, under general assumptions, that bPOE is more than double the probability of exceedance (POE). An example of how the POE and bPOE relate to the area under a probability density function is shown in Figure 1.

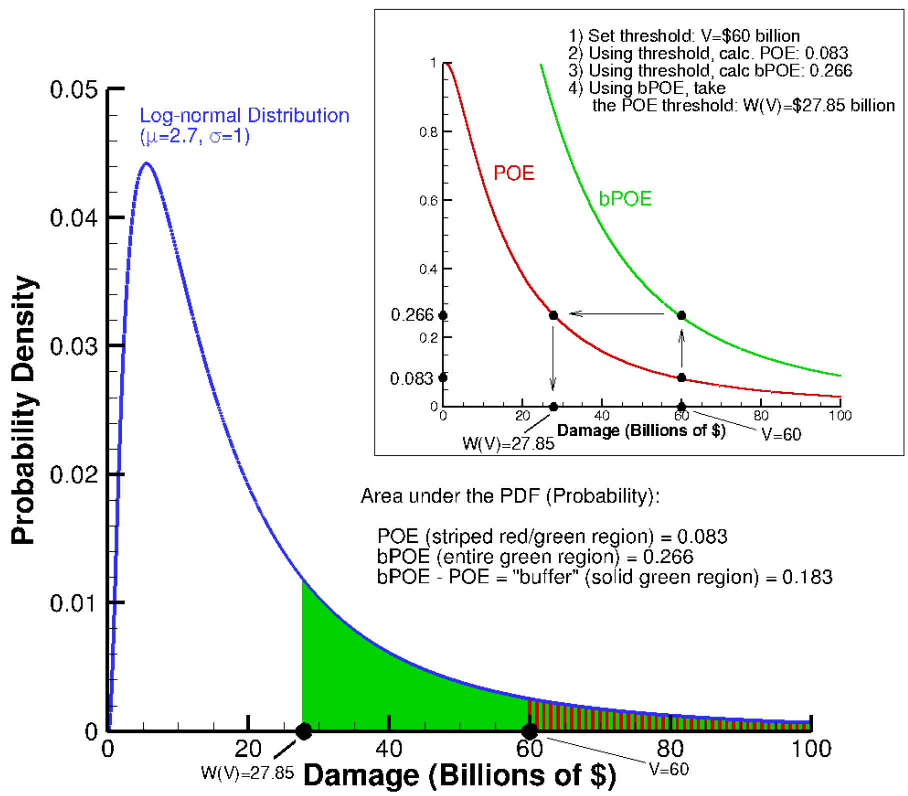

Figure 1 Hypothetically assuming tropical storm economic damage follows a log-normal distribution with a mean 2.7 and standard deviation of 1, the areas under the probability density function which represent the POE and bPOE for a threshold of $60 billion are shown. A geometric representation of how the areas are calculated is shown in the inset.

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 [8], [9]. Thus, rather than using the storm 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 (242 storms over 114 years), combining the presented methodology with an analytic distribution function could be used gain more confidence in the analysis or 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

[5] defines bPOE as one minus the superdistribution function (see definition of superdistribution in [10]). [11] calls the superdistribution function by the Buffered Service Level (BSL) in terms of a server having some capacity and probabilistic demand. While [5] presents theoretical/mathematical properties of bPOE at a highly mathematical level, [11] uses relatively simple terms. Following [11], 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) [6]. 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 [10].

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 (*X*) relative to some given threshold (*W*). The Probability of Exceedance (POE) for that threshold W is defined by

$$p_W(X) = P[X > W] \ . \tag{1}$$

The threshold *W* in $p_W(X)$ is commonly referred to in economics and finance as the "Value at Risk" (VaR). For example, there is a 25% chance that economic damage *X* per from all landfalling storms in a year will exceed the threshold *W*=$11.181 billion. Two characteristics then describe this damage distribution: *W* and $p_W(X) = 0.25$. Note that these two characteristics do not provide any information about the magnitude of damages of *X* exceeding *W*. Sometimes, the information on the possible values of *X* 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 <u>maximum</u> amount of damage a community may experience? Suppose that the damages *X* are normally distributed; in this case (theoretically), there is no limit on the amount of damages experienced for all possible values of *X.* 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 <u>average</u> amount of damage expected? We say that the economic damages $V = E[X|X > W]$ measures *on average* the damages exceeding *W*, where $E[X|X > W]$ is the conditional expectation that the damage *X* exceeds *W.* By definition, economic damage *V has a bPOE equal to* $\bar{p}_V(X) = p_W(X)$. The threshold *V* in $\bar{p}_V(X)$ is commonly referred to in economics and finance as the "Conditional Value at Risk" (CVaR). The formal definition of bPOE [5] can be written as:

Definition 1.

Assume *X* is a continuously distributed random variable and *V* is a threshold in the range, $E[X] \leq V \leq \sup X$. Let us define a threshold W(V), as a function of V, satisfying the following expression:

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

Then by definition, bPOE equals

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

---

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:

$$[X - W(V)]^+ = \max[0, X - W(V)] \ . \tag{2}$$

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

$$V = E[X|X > W(V)] = E[W(V) + X - W(V)|X > W(V)] \tag{3}$$

$$= E[W(V)|X > W(V)] + E[X - W(V)|X > W(V)] \tag{4}$$

$$= W(V) + E[X - W(V)|X > W(V)] \, . \tag{5}$$

By definition of conditional probability

$$E[X - W(V)|X > W(V)] = \frac{E[1_{X>W(V)}(X - W(V))]}{P[X > W(V)]} \, , \tag{6}$$

where

$$1_{X>W(V)} = \begin{cases} 1, & for\ X > W(V) \\ 0 & otherwise \end{cases} \, . \tag{7}$$

Since $1_{X>W(V)}(X - W(V)) = E[X - W(V)]^+$ and P(X>W(V))= $\bar{p}_V(X)$, Equations (5) and (6) imply

$$V = W(V) + \frac{E[X - W(V)]^+}{\bar{p}_V(X)} \ . \tag{8}$$

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

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

$$E[X - W(V)]^+ = \bar{p}_V(X)[V - W(V)] \ . \tag{9}$$

Equivalently we have

$$\bar{p}_V(X) = \frac{E[X - W(V)]^+}{V - W(V)} . \tag{10}$$

The expected excess $E[X - W(V)]^+$ 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 per year. For example, assume that an insurance company is creating a policy to cover tropical storm damage losses per year in excess of some deductible, *W(V)*. Then, amount of money in premiums that should be collected would be the expected excess, $E[X - W(V)]^+$. While some years may incur payable losses below this value and some above, over a large number of insured years, 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 *X*, we have

$$\left(1 - \bar{p}_V(X)\right) \cdot E[X | X \leq W(V)] + \bar{p}_V(X) \cdot E[X | X > W(V)] = E[X]. \tag{11}$$

As a simple analytic example of a continuous distribution, we illustrate the bPOE for a common heavy tailed distribution. However, before individual examples can be presented, it must be noted that so far, the random variable X has been written in upper case. In the following example (and others that appear later), particular realizations of the random variable are written in lower case, for example, $x_1, x_2, \cdots, x_N$.

The probability distribution function (PDF) for an exponential distribution of a random value X is $\lambda e^{-\lambda x}$ with an expected value of $E[X] = 1/\lambda$, where $\lambda > 0$. As such, the cumulative distribution function (CDF) is $1 - e^{-\lambda x}$. Accordingly, the POE is

$$p_x(X) = e^{-\lambda x}. \tag{12}$$

Section 4 of [12]showed that the bPOE for an exponential distribution is equal to

$$\bar{p}_x(X) = e \cdot e^{-\lambda x} = e \cdot p_x(X), \quad \text{for} \quad x > 1/\lambda \ . \tag{13}$$

Thus, the ratio between bPOE and POE is fixed

$$\frac{\bar{p}_x(X)}{p_x(X)} = e, \quad \text{for} \quad x > 1/\lambda \, , \tag{14}$$

and the probability "buffer" is

$$\bar{p}_x(X) - p_x(X) = e \cdot p_x(X) - p_x(X) = (e-1)p_x(X), \quad \text{for} \quad x > 1/\lambda \quad . \tag{15}$$

An illustration of the values of POE and bPOE for an exponential distribution for $\lambda = 1$ is in Figure 2.

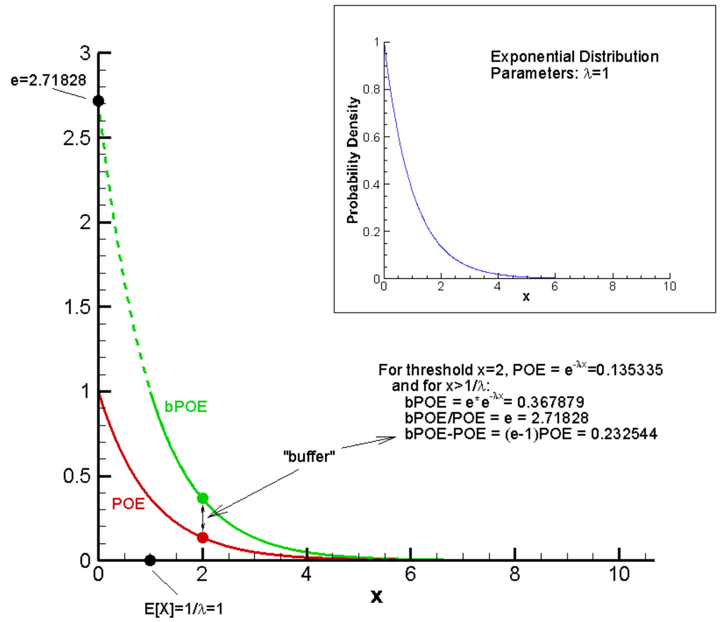

Figure 2 An illustration of the probability of exceedance (POE) (solid red line) and buffered probability of exceedance (bPOE) (solid green line) for the exponential distribution highlighted in the inset.

While the bPOE provides information about the mean value of the tail, it does not provide information on the shape of the tail. As such, it is possible for two different distributions to have the same bPOE at the same threshold as illustrated using a log-normal distribution shown in Figure 3. However, it can be proved that if the bPOE is known for all thresholds, the original distribution (and hence its shape) can be recovered. The proof of which is beyond the scope of this paper.

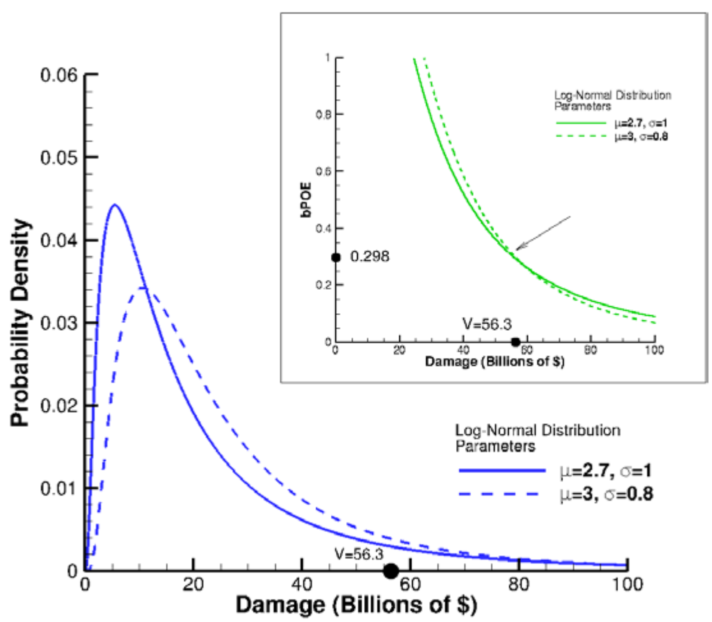

Figure 3 Hypothetically assuming tropical storm economic damage follows one of two possible log-normal distributions, it can be seen that for a threshold of $56.3 billion, the bPOE is equal to 0.298 for both distributions.

Calculation and Optimization of Buffered Probability of Exceedance

While (10) 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 [5], [13]):

$$\bar{p}_V(X) = \begin{cases} 0 & V \geq \sup X \\ \min\limits_{W<V} \dfrac{E[X-W]^+}{V-W} & E[X] < V < \sup X \\ 1 & otherwise \end{cases} . \tag{16}$$

An optimal *W(V)* in Equation (16) satisfies the equation $V = E[X|X > W(V)]$, the proof of which for general distributions is shown in [5]. This minimum formula for the bPOE is very similar to the minimum formula for CVaR, which was established in [14]. The minimum in this formula is achieved on VaR with the 1-bPOE confidence level.

The advantage of (16) can be demonstrated with the *convex* optimization problem formulation obtained in [5], [13]. Let us consider that *X* is a linear function depending on control vector $w \in R^n$ with a random vector of coefficients, θ, i.e., $X = w^T\theta$. We are interested in minimizing the bPOE with *V*=0, which corresponds to Buffered Probability of Failure in [6]. This minimization problem can be formulated as the following convex minimization problem, which can be reduced to *linear programming*:

$$\min_{w \in R^n} E[w^T\theta + 1]^+ \ . \tag{17}$$

bPOE Calculation Algorithm

As stated in (16), 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 $\bar{p}_V$ *t*hen 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 *X* be a discrete random variable with atoms $x_1, x_2, \cdots, x_N$ and probabilities $p_1, p_2, \cdots, p_N$, where $x_i \leq x_{i+1}, i = 1,2,\cdots, N-1$. Denoting confidence levels α_j as

$$\alpha_j = \begin{cases} 0 & \text{for } j = 0 \\ \sum_{i=1}^{j} p_i & \text{for } j = 1,2,\cdots,N-1 \end{cases}. \tag{18}$$

CVaR can be written as [14]

$$\bar{x}_j = \frac{1}{1-\alpha_j} \sum_{i=j+1}^{N} p_i x_i \quad \text{for } j = 0,1,\cdots,N-1. \tag{19}$$

By definition, bPOE equals

$$\bar{p}_{\bar{x}_j} = 1 - \alpha_j \quad \text{for } j = 0,1,\cdots,N-1. \tag{20}$$

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

$$\bar{p}\big(\mu\bar{x}_j + (1-\mu)\bar{x}_{j+1}; X\big) = \left(\frac{\mu}{\bar{p}(\bar{x}_j;X)} + \frac{1-\mu}{\bar{p}(\bar{x}_{j+1};X)}\right)^{-1}. \tag{21}$$

Earlier efforts [15] 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 (18)-(21) can be used to compute CVaR and BPOE, two simple examples are presented: Example #1 consists of a sample dataset in which all data values are unique and equally probable. In Example #2, 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

$$\alpha_0 = 0,$$
$$\alpha_1 = 0.25,$$
$$\alpha_2 = 0.25 + 0.25 = 0.5,$$
$$\alpha_3 = 0.25 + 0.25 + 0.25 = 0.75.$$

CVaR equals

$$\bar{x}_0 = \frac{1}{1-0}(0.25 * 1 + 0.25 * 2 + 0.25 * 5 + 0.25 * 7) = 3.75,$$
$$\bar{x}_1 = \frac{1}{1-0.25}(0.25 * 2 + 0.25 * 5 + 0.25 * 7) = 4.6\bar{6},$$
$$\bar{x}_2 = \frac{1}{1-0.5}(0.25 * 5 + 0.25 * 7) = 6,$$
$$\bar{x}_3 = \frac{1}{1-0.75}(0.25 * 7) = 7.$$

With a corresponding bPOE of

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

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.

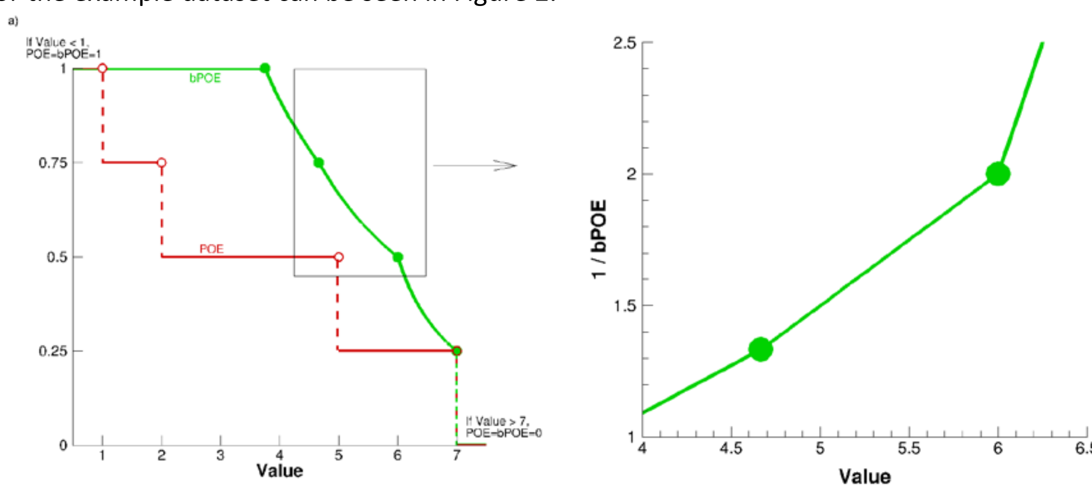

Figure 4 a) A comparison between the POE (red) and bPOE (green) for the example dataset. b) A "zoom in" of the values between 4 and 6.5 demonstrating that 1/bPOE is linear between atoms.

Example #2 – 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

$$\alpha_0 = 0,$$
$$\alpha_1 = 0.2,$$
$$\alpha_2 = 0.2 + 0.2 = 0.4,$$
$$\alpha_3 = 0.2 + 0.2 + 0.4 = 0.8.$$

CVaR equals

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

With a corresponding bPOE of

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

Suppose we are interested in determining the value of $\bar{p}_V$ where $\bar{x} = 6$, then by Equation (21),

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

and

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

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:

> **What is the expected total economic damage be for all of the tropical storms making landfall along the U. S. coastline during a given storm season?**

To answer this question, two 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. During the 114 year historical record, a total of 242 storms from 1900 (Galveston)-2012 (Sandy) (none in 2013) 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]. As a result of this storm, 1926 also experienced the greatest amount of damages per year.
2. In order to compare storms which made landfall of such an extended period of time, it is assumed that the methods described in [16] 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.

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 5. The bPOE is calculated using the algorithm presented in the Appendix. For POE, values are calculated at the damage levels (atoms) associated for each year. Note how the line in Figure 5 which represents POE is not smooth.

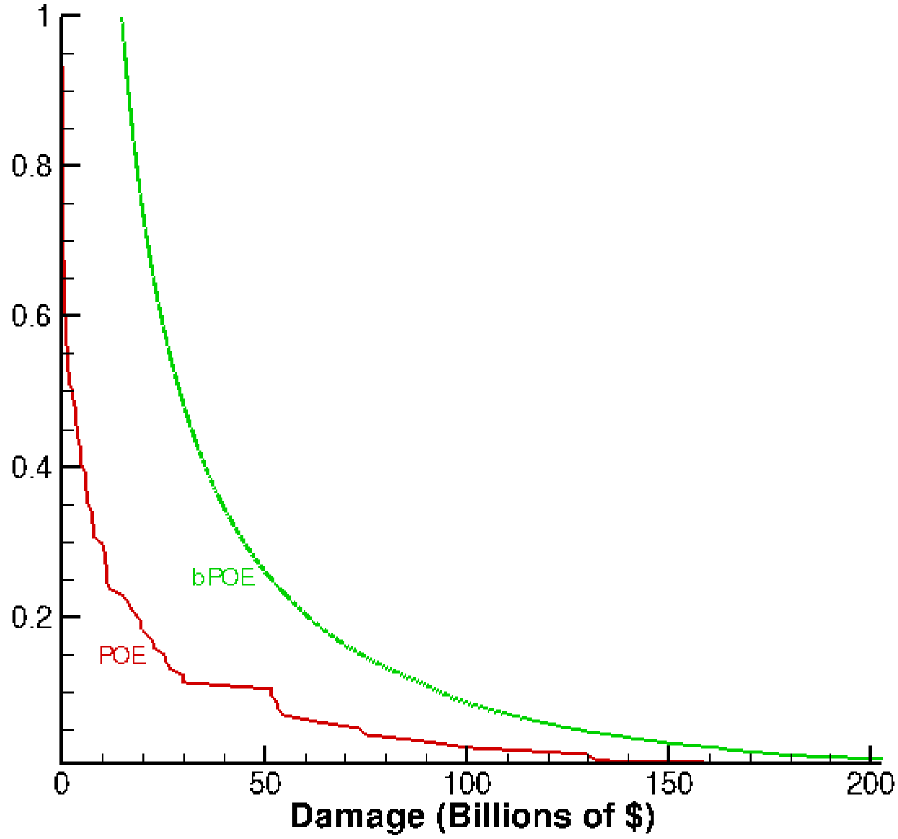

Figure 5 A comparison between the probability of exceedance (POE) (solid red line) and buffered probability of exceedance (bPOE) (solid green line) for damage in the upper tail of the distribution per year caused by all storms that made landfall during 1900-2013. For the POE, the x-axis represents the Value-at-Risk (VaR) which is denoted by W(V) herein, while for the bPOE, the x-axis represents the Conditional-Value-at-Risk which is denoted by a V herein.

Now, let us make the following notations to differentiate the expected value in the upper and lower tails:

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

As an example, we made calculations with bPOE equals 50% and the $W(V)$ threshold corresponding to this bPOE.

> **For a given storm season, there is a 50% probability that the total economic damages caused by all landfalling storms will exceed \$2.450 billion dollars. If the damages do exceed \$2.45, then the average amount of damage will be $\overline{CVaR}$ =\$29.113 billion dollars. If the damages do not exceed this amount, then the average amount of damage will be $\underline{CVaR}$ =\$0.439 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 (9), 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 6 demonstrates how this value works out to be $13.332 billion dollars. Returning to our insurance example mentioned previously, for a $2.45 billion deductible, the insurance company needs to collect at least $13.332 billion dollars in premiums to ensure that it breaks even over a large number of years.

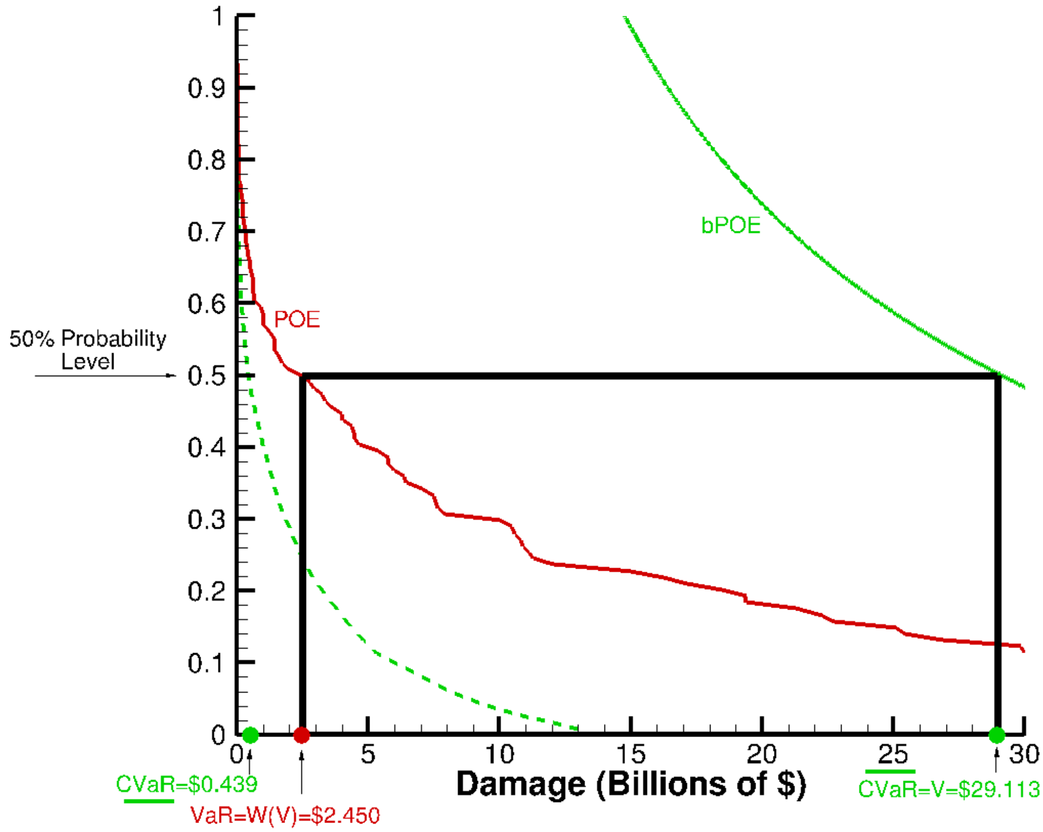

Figure 6 For the 50% probability level, a "zoom in" of the probability of exceedance (POE) (red line) and buffered probability of exceedance (bPOE) (solid green line) for damage in the upper tail along with a line representing the expected value in the lower tail (dashed green line). Expected excess, $E[X - W(V)]^{+}$, is then the area outlined by the box and is equal to (V-W) $\bar{p}_V$ = ($29.113 billion - $2.450 billion) * 0.5 = $13.332 billion. For the POE, the x-axis represents the Value-at-Risk(VaR) in the upper tail which is denoted by a W(V) herein, while for the bPOE, the x-axis represents the Conditional-Value-at-Risk in the upper tail ($\overline{CVaR}$) which is denoted by a V herein. For the line representing the expected value in the lower tail, the x-axis represents the Conditional-Value-at-Risk in the upper tail ($\underline{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 $11.181 billion in damages. If exceeded, $\overline{CVaR}$ is 463% ($40.572 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%	\$132.025	\$210.202	\$12.688	\$78.177	\$0.782
5%	\$73.773	\$127.279	\$8.831	\$53.506	\$2.675
10%	\$51.784	\$93.486	\$6.020	\$41.702	\$4.170
25%	**\$11.181**	**\$51.753**	**\$2.449**	**\$40.572**	**\$10.143**
50%	\$2.450	\$29.113	\$0.439	\$26.663	\$13.332
75%	\$0.183	\$19.690	\$0.033	\$19.508	\$14.631
99%	\$0.000	\$14.925	\$0.000	\$14.925	\$14.776

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 [5] 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 (11):

Lower Tail Contribution + Upper Tail Contribution = Total Expected Damage

0.739*(\$2.336 billion) + 0.261*(\$50 billion) = \$1.726 + \$13.05 billion = \$14.776 billion.

Upper Tail Contribution: Examining the \$50 billion threshold (marked by bold face in Table 3), the POE equals 10.6%. 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 26.1%; therefore, the expected damage in the 26.1% upper tail equals \$50 * 0.261= \$13.05 billion.

Lower Tail Contribution: The lower tail has probability 100% - 26.1% = 73.9%. The expected damage in the lower tail equals the difference of total expected damage and expected damage in the upper tail: \$14.776 - \$13.05 = \$1.726 billion. So, the conditional expected damage in the lower tail equals its expected damage divided by its probability: \$1.726 / 0.739 = \$2.336 billion.

Thus it can be seen that the expected damage for the original damage distribution containing 114 observations can be “packed” into a discrete distribution with two atoms: lower tail expected damage \$1.726 billion is the product of \$2.336 billion (lower tail conditional expected damage) and probability of this tail 0.739 and the upper tail expected damage \$13.05 billion is the product of \$50 billion (upper tail conditional expected damage) and probability 0.261. Such representation shows importance of damages in the upper tail of the distribution.

The lower tail contains 114 * 0.739 ≈ 84 observations and the upper tail, 114 - 84 = 30 observations. By ordering damages in the dataset, this split between the upper and lower tails at the 30th observation occurs at a damage of \$11.1 billion. Again, referring to our insurance example, suppose an insurance fund is setup assuming a deductible of \$11.1 billion (VaR). Then, according to Equation (9), the expected excess in the upper tail (the minimum amount to charge in premiums) equals the product of bPOE = 0.261 and difference of the conditional expected damage in the upper tail, $50 billion (CVaR), and the deductible $11.1 billion:

Expected excess = ($50 - $11.1) * 0.261 = $10.15 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	2.7%	8.6%
$75	4.5%	14.7%
$50	**10.6%**	**26.1%**
$25	14.9%	58.8%
$10	29.8%	100.0%
$5	39.9%	100.0%
$1	58.3%	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 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. It should be noted that this number of observations is fairly limited (114 years' worth of data), thus any conclusions drawn from the values of bPOE must take into account this fact. However, any other statistical characteristics, such as POE also are based on the same limited number of observations and they prone to the same weaknesses. However, emergency planners, insurance companies, etc. are using this same set of observations for their own analysis, so in that context, bPOE still improves upon other statistics measures such as POE that are being used by the practicing community.

Analysis of the damage data shows that for the damage associated with all landfalling storms in a given year, for a $50 billion threshold, the POE is 10.6% while the bPOE if 26.1%. 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 $11.181 billion, while the conditional-value-at-risk (average of damage received) determined by the bPOE is $51.753. For a fixed probability, this $40.572 billion difference between the damage associated with POE and bPOE leads to an expected excess of 0.25*$40.572=$10.143 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. [17]). 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] R. D. Knabb, J. R. Rhome, and D. P. Brown, "Hurricane Katrina: 23-30 August 2005," National Hurricane Center, Tropical Cyclone Report, Sep. 2011.

[3] 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.

[4] "Total Gross Domestic Product by State for Florida," *Federal Reserve Bank of St. Louis*. [Online]. Available: http://research.stlouisfed.org/fred2/series/FLNGSP.

[5] A. Mafusalov and S. Uryasev, "Buffered Probability of Exceedance: Mathematical Properties and Optimization Algorithms," Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, Research Report 2014-1, Jan. 2016.

[6] 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.

[7] R. T. Rockafellar, "Safeguarding Strategies in Risky Optimization," presented at the International Workshop on Engineering Risk Control and Optimization, Gainesville, Florida, Feb-2009.

[8] W. D. Nordhaus, "The Economics of Hurricanes and Implications of Global Warming," *Climate Change Economics*, vol. 01, no. 01, pp. 1–20, May 2010.

[9] C. Blackwell, "Investigating Power Laws in Hurricane Damages," *SSRN Electronic Journal*, 2012.

[10] R. T. Rockafellar and J. O. Royset, "Random variables, monotone relations, and convex analysis," *Mathematical Programming*, Aug. 2014.

[11] 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.

[12] A. Mafusalov, A. Shapiro, and S. Uryasev, "Estimation and Asymptotics for Bufferered Probability of Exceedance," Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, Research Report 2015-5, Oct. 2015.

[13] M. Norton and S. Uryasev, "Maximization of AUC and Buffered AUC in Classification," Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, Research Report 2014-2, Jan. 2015.

[14] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, no. 3, pp. 21–41, 2000.

[15] *Portfolio Safeguard*. American Optimal Decisions, Inc.

[16] 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.

[17] 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," *Water Resources Research*, vol. 43, no. 7, Jul. 2007.

Appendix: bPOE Algorithm

INPUT:
N data points (x):

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

their associated probability of occurrence (p):

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

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):

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

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

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

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

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

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

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

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

$$\left[\left(V_1, \bar{p}_{v_1}\right), \left(V_2, \bar{p}_{v_2}\right), \cdots, \left(V_N, \bar{p}_{v_N}\right)\right]$$

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:

$$\left[\left(V_1, \hat{p}_{v_1}\right), \left(V_2, \hat{p}_{v_2}\right), \cdots, \left(V_N, \hat{p}_N\right)\right]$$

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

$$V = \begin{cases} Indeterminant & \hat{p}_{v_N} > \hat{p}_v \geq 1 \\ V_i + (V_{i+1} - V_i)\dfrac{(\hat{p}_v - \hat{p}_{v_i})}{(\hat{p}_{v_{i+1}} - \hat{p}_{v_i})} & \hat{p}_{v_i} \geq \hat{p}_v \geq \hat{p}_{v_{i+1}} \quad where\ i = 1 \cdots (N-1) \\ V_1 & \infty > \hat{p}_v > \hat{p}_{v_1} \end{cases}$$

-or-

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

$$\hat{p}_v = \frac{1}{\bar{p}_v} = \begin{cases} 1 & V_N > V > -\infty \\ \hat{p}_{v_i} + \left(\hat{p}_{v_{i+1}} - \hat{p}_{v_i}\right)\dfrac{(V - V_i)}{(V_{i+1} - V_i)} & V_i \geq V \geq V_{i+1} \quad where\ i = 1 \cdots (N-1) \\ 0 & \infty > V > V_1 \end{cases}$$