involve
a journal of mathematics

A statistical study of extreme nor'easter snowstorms

Christopher Karvetski, Robert B. Lund and Francis Parisi

mathematical sciences publishers

2009 Vol. 2, No. 3

INVOLVE 2:3(2009)

A statistical study of extreme nor'easter snowstorms

Christopher Karvetski, Robert B. Lund and Francis Parisi

(Communicated by Sat Gupta)

This short paper studies the statistical characteristics of extreme snowstorms striking the eastern seaboard of the United States — the so-called nor'easters. Poisson regression techniques and extreme value methods are used to estimate return periods of storms of various snow volumes. Return periods of several memorable events are estimated, including the superstorm of 1993, the North American blizzard of 1996, and the blizzard of 1888. While nor'easters are found to occur more frequently in late winter than early winter, no evidence of increasing/decreasing storm frequencies in time or dependencies on the North Atlantic oscillation is found.

1. Introduction

A nor'easter is a large-scale winter storm that impacts the east coast of the United States. A nor'easter can drop copious amounts of snow and may also cause flood and wind damage. Nor'easters occur from late fall through early spring. The superstorm of 1993 (March 12–15), for example, was the largest snowstorm affecting the United States in the last century. This storm deposited over 60 inches of snow in some places, is blamed for 300 fatalities, and caused an estimated six to ten billion dollars of damage.

While nor'easters are sometimes referred to as *winter hurricanes*, literature studying their frequency properties is sparse when compared to that for *summer hurricanes*. The goal here is to quantify the nor'easter snowstorm hazard. The timing of this work coincides with attempts by several insurance risk modeling firms to quantify the hazard.

The total snow volume of each storm will be used as the measure of storm severity. Return periods for various snow volume accumulations will then be estimated. A snow volume return period is how long one waits, on average, until a nor'easter with a preset snow volume or greater strikes. For example, the superstorm of 1993

MSC2000: 60G55, 62G07, 62G32, 62M99.

Keywords: extreme values, North Atlantic oscillation, peaks over threshold, Poisson processes, snowstorms.

is also nicknamed the *Storm of the Century*, giving connotations of a 100-year storm. Later, we will see that this storm was more than a 100-year event.

2. The data

The data for this analysis were taken from [Kocin and Uccellinni 2004a; 2004b]. These references contain much information about the individual storms. Nor'easter documentation is scant when compared to that for summer hurricanes (for the latter, see [Blake et al. 2005; Parisi and Lund 2008]). Our data consists of 65 storms occurring during the years 1953–2003 inclusive; the record for this time period is complete. The individual storms are chronologically listed in Table 1; no storms occurred in 1953, 1954, or 1955. As the pre-1953 record is incomplete, we cannot include memorable pre-1953 storms (such as the New England blizzard of 1888), without biasing the overall results.

For each storm in the table, Kocin and Uccellinni 2004a report the areas that accumulated more than four inches, ten inches, twenty inches, and thirty inches, respectively. For each storm, we compute a crude volume estimate via the following rubric. For the superstorm of 1993, an area of 386.0×10^3 squared miles experienced snow accumulations of at least four inches, an area of 283.5×10^3 squared miles saw accumulations of at least ten inches, an area of 142.4×10^3 squared miles had accumulations of at least twenty inches, and an area of 12.9×10^3 squared miles received accumulations of over thirty inches. A volume estimate for the superstorm of 1993 is hence

$$386 \times 4 + 283.5 \times (10 - 4) + 142.4 \times (20 - 10) + 12.9 \times (30 - 20) = 4798,$$

where the units on the volume are 10^3 inches times squared miles. The volumes can be converted to cubic meters upon multiplication by 6.5024×10^7, but we will not do this as the analysis below is invariant of any linear scale change on the volume units. Because of this, volume units will be henceforth suppressed for simplicity.

Whereas our estimated volume underestimates actual values (areas receiving less than 4 inches of snow, for example, are not included in the volume estimates), the estimated volumes are reasonable measures of storm intensity; moreover, all volumes are underestimated in the same way, which makes storm-by-storm comparisons meaningful. The smallest volume was 291.2 and the largest volume was 4798.0. On average, there are about 1.27 storms per season. The number of storms in a single season ranges from zero to five.

We emphasize that this data only contains large-scale nor'easters and not more localized events such as Great Lake effect snowfalls. Alberta Clipper-type storms, whose snowfall volumes tend to be much less than nor'easters, are also not represented in this data.

Date (midstorm)			Sq. miles covered with				NAO	Rough
yr	mo	day	4 in	10 in	20 in	30 in	average	volume
1956	3	16	195.5	92.3	0	0	−1.07	1335.8
1956	3	18	64.9	28.6	2.6	0	−1.07	457.2
1957	12	4	87.2	9.4	0	0	0.073	405.2
1958	2	15	282.6	129.2	20.2	3.4	0.073	2141.6
1958	3	19	146.7	62.1	13.8	3.5	0.073	1132.4
1959	3	12	215.3	121.1	7.7	0	−0.13	1664.8
1960	2	14	353.9	142.1	23.3	0	0.093	2501.2
1960	3	3	590.4	140.8	7.6	0	0.093	3282.4
1960	12	12	302.9	78.5	0.6	0	1.993	1688.6
1961	1	19	144.9	62.3	5.7	0	1.993	1010.4
1961	2	3	369.3	114	19.4	1.4	1.993	2369.2
1961	12	24	105.5	14.8	0	0	0.497	510.8
1962	2	14	101.4	33.8	0.4	0	0.497	612.4
1962	3	6	148.6	70	19.3	0	0.497	1207.4
1963	12	22	374.2	51.3	0	0	−0.763	1804.6
1964	1	12	356.5	129.6	10.3	0	−0.763	2306.6
1964	2	19	169.7	53.4	3.5	0	−0.763	1034.2
1965	1	16	214.5	15.3	0	0	−1.42	949.8
1966	1	22	296.4	145.1	6.6	0	−0.05	2122.2
1966	1	30	371.4	111.7	12.3	1.5	−0.05	2293.8
1966	12	24	292.2	89.8	9.9	0	1.14	1806.6
1967	2	6	246	50.9	0	0	1.14	1289.4
1967	3	21	62.3	7	0	0	1.14	291.2
1969	2	9	107.5	66.4	11.6	0	−2.177	944.4
1969	2	25	101.7	48.4	40.8	24.2	−2.177	1347.2
1969	12	26	250.6	138.7	37.6	0	−0.107	2210.6
1970	12	31	151	46.4	4.4	0	−0.267	926.4
1971	3	4	195.7	101.6	23.3	0	−0.267	1625.4
1971	11	26	163.4	73.4	6.6	0	0.013	1160
1972	2	19	206.3	140.9	13.5	0	0.013	1805.6
1978	1	17	364.4	122.1	0	0	−0.593	2190.2
1978	1	20	295.2	167.7	8.3	0	−0.593	2270
1978	2	6	220.2	132.3	30.7	0.9	−0.593	1990.6
1979	2	18	304	88.2	4.3	0	−1.973	1788.2

Table 1. The nor'easter data (continued on next page).

Date (midstorm)			Sq. miles covered with				NAO	Rough
yr	mo	day	4 in	10 in	20 in	30 in	average	volume
1982	1	14	382.2	133.9	0	0	−0.223	2332.2
1982	4	6	258.3	79.3	2.1	0	−0.223	1530
1983	2	11	157.1	112.6	33.7	0.9	2.07	1650
1984	3	8	120.9	54.6	0	0	1.697	811.2
1984	3	28	124.6	53.3	2.1	0	1.697	839.2
1987	1	1	164.6	76.6	0	0	0.353	1118
1987	1	22	286.9	153.7	2	0	0.353	2089.8
1987	1	25	74.3	38	0	0	0.353	525.2
1987	2	22	61.3	28.3	0.3	0	0.353	418
1988	1	7	488.5	129.7	0	0	−0.13	2732.2
1990	12	26	166	12.7	0	0	0.73	740.2
1992	12	11	118.7	61.6	21.5	0	1.41	1059.4
1993	3	13	386	283.5	142.4	12.9	1.41	4798
1994	1	4	222.3	76.4	10.5	0	1.173	1452.6
1994	2	9	280	57.7	4.4	0	1.173	1510.2
1994	3	3	165.4	109.1	0	0	1.173	1316.2
1995	2	3	200.1	98	0	0	2.897	1388.4
1995	12	20	260.3	85.4	0	0	−2.24	1553.6
1996	1	7	313.8	200.1	90.2	15.1	−2.24	3508.8
1996	2	3	157.3	44.1	0.9	0	−2.24	902.8
1996	2	16	136.7	12.2	0	0	−2.24	620
1997	3	31	76.4	32	13.1	3.1	−0.463	659.6
1996	1	7	313.8	200.1	90.2	15.1	−2.24	3508.8
1996	2	3	157.3	44.1	0.9	0	−2.24	902.8
1996	2	16	136.7	12.2	0	0	−2.24	620
1997	3	31	76.4	32	13.1	3.1	−0.463	659.6
1999	3	14	180.3	58.8	1.4	0	1.55	1088
2000	1	25	205.6	74.2	0.3	0	2.283	1270.6
2000	12	30	103.8	56.5	3.7	0	−0.44	791.2
2001	3	5	161.1	105.1	30.4	1.8	−0.44	1597
2002	12	4	269.7	6.1	0	0	0.17	1115.4
2002	12	25	345.3	91.3	13.8	4.4	0.17	2111
2003	1	3	211.1	77.4	11	0	0.17	1418.8
2003	2	6	88.4	6.1	0	0	0.17	390.2
2003	2	16	303.5	142	51.9	2.7	0.17	2612

Table 1 (continued)**.** The nor'easter data.

3. Arrival properties of nor'easters

To estimate return periods, we need to model the storm arrival times. Following [McDonnell and Holbrook 2004] and [Parisi and Lund 2008], the storm count in season t is modeled as a Poisson random variable with mean λ_t, where

$$\lambda_t = \exp\left(\beta_0 + \alpha t + \beta_1 \mathrm{NAO}_t\right).$$

Here, the parameter α allows for a linear trend in the storm counts (we will assess whether or not this parameter is zero below) and the $\beta_1\mathrm{NAO}_t$ component allows for possible influences of the North Atlantic oscillation (NAO). The average NAO index over December, January, and February during season t is used for NAO_t. Kocin and Uccellinni [2004a] suggest that the NAO may influence nor'easter storm counts (see [Van den Dool et al. 2006] for generalities about the NAO and North American climate).

Poisson regression techniques were used to fit the above model. The estimated parameters are $\hat{\alpha} = 0.008$ and $\hat{\beta}_1 = -0.0149$. Intervals of 95% confidence for α and β are $[-0.009, 0.025]$ and $[-0.339, 0.041]$. As both of these intervals contain zero, these two parameters are statistically indistinguishable from zero with 95% confidence. All possible subsets of the regression model structure were also fitted and produced insignificant parameters at the 95% level. Thus, we do not find evidence of trends or NAO influences in the storm counts. A Kolmogorov–Smirnov test fails to reject a Poisson distribution for the annual storm counts at the 95% level. In short, the annual nor'easter storm counts pass as statistical white noise with a Poisson marginal distribution with a mean of approximately 1.27 storms per season.

Although the number of storms from season to season appears time-homogeneous, the storms do not arrive uniformly within a season. To investigate this aspect, the kernel intensity estimate in Figure 1 was constructed. This graphic presents a probabilistic description of when storms occur within a season. As the earliest storm in our data record occurred on November 26 (and the latest on April 6), we have chosen to measure a storm's arrival date as the number of days after October 1 that the storm's midpoint took place on (the midpoint, or the average of the storms' beginning and ending dates, is used since some storms last multiple days). Figure 1 displays estimates of the intensity function $\hat{\lambda}(t)$ at time t defined by

$$\hat{\lambda}(t) = \frac{1}{N_{\mathrm{yr}}} \sum_{i=1}^{N_{\mathrm{st}}} h^{-1} K\left(\frac{t-d_i}{h}\right), \quad 0 \le t < 365.$$

Here, $N_{\mathrm{yr}} = 51$ is the number of seasons of data, $N_{\mathrm{st}} = 65$ is the total number of storms, K is the Gaussian kernel function

$$K(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}}, \quad -\infty < x < \infty,$$

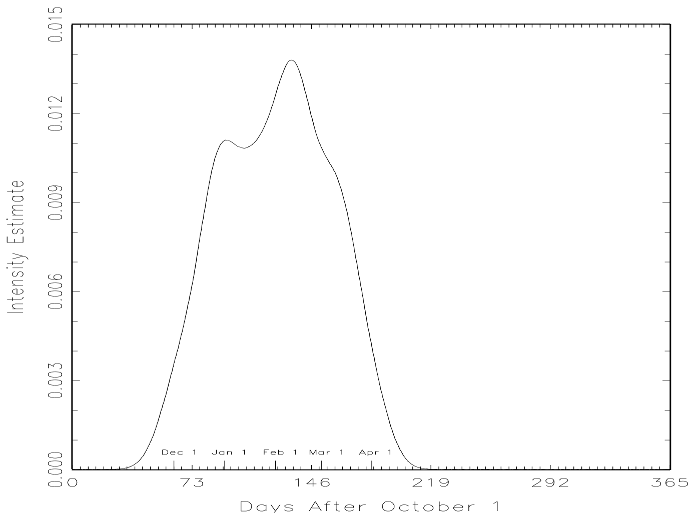

Figure 1. Estimated seasonal intensity function.

$h = 10$ is a bandwidth parameter that controls the amount of smoothing done (see [Sheather and Jones 1991] for discussion about selecting an appropriate h), and d_i is the number of days after October 1 that the midpoint of the ith storm occurs on for $1 \leq i \leq N_{\text{st}}$. The interpretation of $\hat{\lambda}(t)$ is that the probability of a nor'easter occurring in the time interval $(t, t+h]$ is approximately $\hat{\lambda}(t)h$ for small h.

The intensity function in Figure 1 peaks at about 133.8 days after October 1, or around February 11. Hence, nor'easters are slightly more likely to occur after midwinter (which is about January 20) than before midwinter. Though we cannot offer a meteorological explanation for this pattern, we note that summer hurricane arrivals also peak in the later half of their season.

4. Return periods

Our next task lies in estimating the return periods of nor'easters. The return periods derived below apply to the northeastern United States as a whole and not to a specific geographic location. Elaborating, a return period of a twenty inch snowfall for New York City should be estimated from snowfall data taken in New York City proper (whose record is much longer than our 51 years) and not the nor'easter data set in Table 1.

The return period of a volume x storm is simply how long one waits, on the average, until a storm occurs that deposits a snow volume of x or more. We measure all return periods from October 1. For example, one waits an average of 839.5 days after October 1 of any calendar year for a 2.3 year nor'easter to occur.

To estimate return periods, we need to model the snow volumes of the storms. For this, we appeal to the peaks over threshold paradigm (see [Embrechts et al. 1997] for general discussion). Elaborating, there is very general mathematical justification for fitting the Pareto cumulative distribution

$$P(V_i - u \leq x | V_i > u) = 1 - \left(1 + \xi \frac{x}{\sigma}\right)_+^{-1/\xi}, \quad x \geq 0, \tag{4.1}$$

to the nor'easter snow volumes $\{V_i\}_{i=1}^{N_{\mathrm{st}}}$. The three parameters in the Pareto model are the shape parameter ξ, the scale parameter σ, and the threshold $u > 0$. In Equation (4.1), $x_+ = \max(x, 0)$. The threshold $u = 290$ is selected. This threshold allows all nor'easters in Table 1 to be considered and passes the rudimentary diagnostic checks suggested in [Davison and Smith 1990]. The maximum likelihood parameter estimates of the other Pareto parameters and 95% confidence intervals are $\hat{\xi} = -0.2899$ ($[-0.436, -0.144]$) and $\hat{\sigma} = 1544.931$ ($[1174.10, 2032.90]$). A Kolmogorov–Smirnov goodness of fit test fails to reject the fitted Pareto distribution with a p-value of 0.2175. Due to the support set of the distribution in (4.1), the negative estimate of ξ implies that snow volumes of nor'easters cannot exceed $u - \sigma/\xi$, which is approximately 5619.2 in this case. Finally, we regressed the snow volumes on the average December, January, and February NAO index to ascertain if the NAO influences snow volumes (Section 3 shows that NAO does not influence storm counts). No statistically significant relationship was found.

With the above model, return periods can be estimated via simulation. One season of the process is simulated as follows. First, a nonhomogeneous Poisson process with the intensity function in Figure 1 is simulated. This intensity function integrates to approximately 1.27 over an annual cycle, which is the mean number of storms per season. For each generated storm in this cycle, we then simulate a snow volume from the fitted Pareto model.

For a volume of x, the waiting time of the simulation is the elapsed time, measured from October 1, until the first storm whose snow volume exceeds x is encountered. If no snow volume of x is encountered in this season, then one adds a year to the waiting time and simulates another season. This process is repeated until a snow volume of x or more is encountered.

The above scheme generates one fair draw of a "level x" waiting time. Every time a snow volume of x or more is encountered, the simulation run is over and the next run starts from scratch (October 1). An estimate of the return period is based on empirically averaging many independent waiting times.

Simulating the necessary processes is reasonably easy; see [Ross 2002] for general detail. One aspect, however, does merit some elaboration: how to generate a Poisson process from the intensity function in Figure 1. This is done by Poisson thinning. Specifically, to generate one season of storm arrival times, we generate a

Storm name	Volume	Estimated return period (years)
Blizzard of 1888	1837.0	2.4
February blizzard of 1978	1990.6	2.8
Presidents' Day storm of 2003	2612.0	5.5
North American blizzard of 1996	3508.8	19.0
Superstorm of 1993	4798.0	500.1

Table 2. Estimated return periods for some historical storms.

time-homogeneous Poisson process with arrival rate λ^* satisfying $\hat{\lambda}(t) \leq \lambda^*$ for all $t \in [0, 365]$. If a storm occurs at time t in the time-homogeneous process, we then independently flip a coin with heads probability $\lambda(t)/\lambda^*$. If the coin is heads, the storm is kept; if the coin is tails, we disregard the storm. The "thinned process" of retained storms is indeed a sample from a nonhomogeneous Poisson process with arrival rate $\hat{\lambda}(t)$ at time t (see [Ross 1996, page 80] for a proof).

Figure 2 plots estimated return periods for various volume levels as estimated by the model. For example, a snow volume of 4250 has an estimated return period of 82.4 years. This return period was estimated by averaging one hundred thousand independent waiting time draws; hence, there is little simulation error.

While there is little simulation error in this return period estimate, significant model uncertainties may well be present. One could use asymptotic normality of the Pareto parameter estimators to quantify the Pareto uncertainties in the return period estimates (the uncertainties in the Poisson arrival cycle are somewhat harder

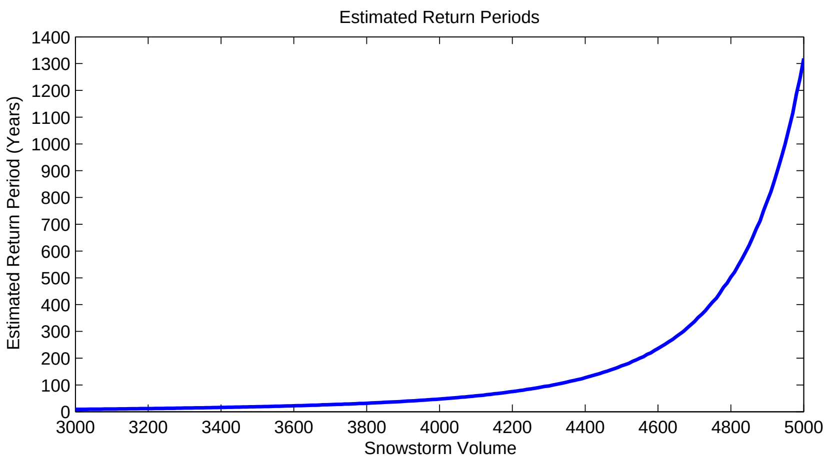

Figure 2. Estimated return periods by snow volume.

to quantify); however, such schemes do not appear to work well in practice (see [Tajvidi 2003] for discussion and possible remedies). While we will not delve into model uncertainties further, Bayesian methods may be promising.

The graphic in Figure 2 shows that a nor'easter with a volume of 3000 occurs about once every nine years on average; a hundred-year volume is about 4300. The return periods increase rapidly for volume levels above 4000. In fact, a volume of 4800 (which is only 2 more than the superstorm of 1993) has a return period of about 500 years. Indeed, it appears that the superstorm of 1993 deserves its "Storm of the Century" nickname.

Table 2 shows estimated return periods of selected historical storms. The blizzard of 1888 has a return period of about 2.4 years, a relatively common event given its historical lore. This estimate is, however, reasonable: while dropping very heavy snow, the storm did not affect a large area. The recent Presidents' Day blizzard of 2003 has a return period of about 5.5 years. The only two storms in our data set with volumes above 3500 are the North American blizzard of 1996 (a volume of 3508.8) and the superstorm of 1993 (a volume of 4798). The return period for the North American blizzard of 1996 is estimated at 19.0 years and the superstorm of 1993's return period is estimated at a whopping 500.1 years. The superstorm of 1993 is clearly an extreme event; indeed, its volume lies close to the statistical boundaries of what is deemed possible. Whereas this return period estimate likely contains considerable error due to model uncertainty, it was indeed an impressive event. In fact, accounts of pre-1953 blizzards do not suggest an event of this magnitude over the last 300 years (the Great Storm of February 1889 and the Great Snow of 1717 seem the closest in magnitude; see [Burt 2004] for descriptions of these storms).

References

[Blake et al. 2005] E. S. Blake, E. N. Rappaport, J. D. Jarrell, and C. W. Landsea, "The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2004 (and other frequently requested hurricane facts)", NOAA technical memorandum, National Weather Service, Miami, FL, 2005.

[Burt 2004] C. C. Burt, *Extreme weather: guide and record book*, W. W. Norton & Company, New York City, 2004.

[Davison and Smith 1990] A. C. Davison and R. L. Smith, "Models for exceedances over high thresholds (with discussion)", *Journal of the Royal Statistical Society, Series B* **52** (1990), 393–442.

[Van den Dool et al. 2006] H. M. Van den Dool, P. Peng, A. Johansson, M. Chelliah, A. Shabbar, and S. Saha, "Seasonal-to-decadal predictability and prediction of North American climate — the Atlantic influence", *J. Climate* **19** (2006), 6005–6024.

[Embrechts et al. 1997] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modeling extremal events*, Applications of Mathematics **33**, Springer, Berlin, 1997.

[Kocin and Uccellinni 2004a] P. J. Kocin and L. W. Uccellinni, *Northeast snowstorms, I: overview*, American Meteorological Society, Boston, 2004.

[Kocin and Uccellinni 2004b] P. J. Kocin and L. W. Uccellinni, *Northeast snowstorms, II: the cases*, American Meteorological Society, Boston, 2004.

[McDonnell and Holbrook 2004] K. A. McDonnell and N. Holbrook, "A Poisson regression model of tropical cyclogenesis for the Australian-Southwest Pacific Ocean region", *Weather Forecasting* **19** (2004), 440–455.

[Parisi and Lund 2008] F. Parisi and R. B. Lund, "Return periods of continental U.S. hurricanes", *Journal of Climate* **21** (2008), 403–410.

[Ross 1996] S. M. Ross, *Stochastic processes*, 2nd ed., John Wiley and Sons, New York City, 1996. Zbl 0888.60002

[Ross 2002] S. M. Ross, *Simulation*, 3rd ed., Academic Press, Inc., Orlando, 2002.

[Sheather and Jones 1991] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation", *J. Roy. Statist. Soc. Ser. B* **53**:3 (1991), 683–690. MR 1125725 Zbl 0800.62219

[Tajvidi 2003] N. Tajvidi, "Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions", *Extremes* **6** (2003), 111–123. Zbl 1063.62037

Received: 2009-02-02 Accepted: 2009-04-22

ckarvetski@gmail.com *Department of Systems and Information Engineering, The University of Virginia, PO Box 400747, Charlottesville, VA 22904, United States*

lund@clemson.edu *Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975, United States*

francis_parisi@standardandpoors.com *Standard & Poor's, Structured Finance Research, 55 Water Street, New York City, NY 10041, United States*