343

Predicting the Expected Number of U.S. Lightning Fatalities for a Year and for a Date Within that Year

William P. Roeder
Private Meteorologist
Rockledge, FL, U.S.A

1. BACKGROUND:

The number of lightning deaths per year in the U.S. has been declining for many decades (Roeder, 2013; Roeder, 2012; Holle, 2012; Ashley and Gilson, 2009; Holle et al., 2005; Lopez and Holle, 1998). As a result, the running 30-year mean used by National Weather Service for weather fatalities overestimates the current rate. For example, the 30-year running mean (1984-2013 (2014 not yet added)) is 51 deaths (National Weather Service, 2015a), which is considerably more than the 28.9 deaths predicted for the 2015 mean by the new method in this paper. Sometimes the 10-year running mean is used to partially mitigate the problem of the 30-year running mean. While the 10-year running mean (1984-2013) of 33 deaths (National Weather Service, 2015a) is closer than the 30-year running mean, it still overestimates the correct rate. In addition, such a short period average can be skewed by a single extreme event.

2. UPDATES TO 2012 STUDY:

This paper updates the 2012 study of the inter-annual and intra-annual lightning fatalities in the U.S. (Roeder, 2013; Roeder, 2012).

2.a. Updates to Inter-Annual Fatalities

The updates to the inter-annual study are relatively minor. These updates added the three latest years of observed U.S. lightning fatalities (2012-2014) and restricted the exponential best-fit curve to the last 30 years (1985-2014) to make the results more representative of current U.S. annual lightning fatality rates. The annual lightning fatalities in the U.S. are listed in Table-1 and shown graphically in Figure-1 along with a best-fit negative exponential curve.

The best-fit exponential curve allows estimating the expected number of lightning fatalities for any year in the near future that is for that specific year while still being consistent with the pattern of lightning fatalities from the recent past decades. The previous study provided a best-fit curve from 1941 onward, since that was when the onset of exponential decrease of lightning fatalities appeared to begin in the U.S. (Roeder, 2012). The new best-fit negative exponential curve for the expected number of U.S. annual lightning fatalities for the period 1985-2014 is in (1).

$$\text{expected annual fatalities} = 3424.3e^{-0.043(x-1900)} \quad (1)$$

$$r^2 = 0.80$$

where x = year

Table-1.
Number of annual lightning deaths in the U.S. from 1985-2014. Data from 1985-1991 are from Lopez and Holle (1998), 1992-2005 are from National Weather Service (2015a), and 2006-2014 are from National Weather Service (2015b).

Year	Deaths	Year	Deaths	Year	Deaths
1985	85	1995	85	2005	38
1986	78	1996	52	2006	48
1987	99	1997	42	2007	45
1988	82	1998	44	2008	28
1989	75	1999	46	2009	34
1990	89	2000	51	2010	29
1991	75	2001	44	2011	26
1992	41	2002	51	2012	28
1993	43	2003	43	2013	23
1994	69	2004	31	2014	26

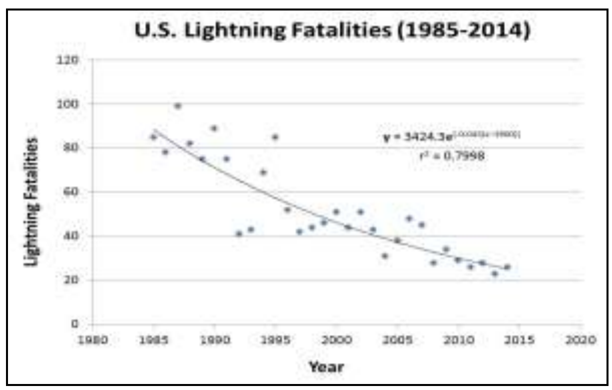

Figure-1. Number of U.S, annual lightning fatalities for 1985-2014 and best-fit negative exponential curve.

Various error bars of the U.S. annual lightning fatalities was also calculated for the data. The average and various percentiles were calculated for each year ± 10 years. Then a best-fit curve was calculated through the years as a smoothing function to filter the noise of natural variation and to provide results that are consistent across the years. The new best-fit negative exponential curve for the expected mean number of U.S. annual lightning fatalities for 1995-2004 is in (2). Note that the period of record is different than for the expected number of days to allow for the range of

Corresponding Author: William P. Roeder,
e-mail: wroeder@cfl.rr.com

± 10 years around the center year used in the calculations. The subtraction of 1900 years in the exponents of equations (1), (2), and (3) is done merely for the convenience of keeping the first coefficients (in front of the exponential function) from being excessively large, e.g. ~10^3 vs. ~10^{24}.

$$\text{mean annual fatalities} = 2283.00e^{-0.038(x-1900)} \quad (2)$$

$$r^2 = 0.99$$

where x = year.

The best-fit negative exponential curve for the median number of U.S. lightning fatalities for 1995-2004 is in (3). The median is preferred over the mean since the percentile distributions are very asymmetric and the

$$\text{median annual fatalities} = 4889.10e^{-0.047(x-1900)} \quad (3)$$

$$r^2 = 0.92$$

where x = year

median has an equal probability of the actual number being higher or lower. The expected median number of lightning fatalities in the U.S. for 2014 was 23.0 deaths, which compared well with the observed number of 26 deaths, and was somewhat less than the expected mean of 30.0 deaths.

Although linear regression fit the median almost as well (r^2 = 0.9050 vs. 0.9156, respectively), the negative exponential regression was chosen since the preceding years clearly had a negative exponential distribution (Roeder, 2012) and so represents reality better. It is only in the later period, when the lightning fatalities are in the tail of the curve that the curve appears to be quasi-linear. Also, the linear regression decreases too rapidly when extrapolated into the future. The best fit negative exponential curve will not have this disadvantage given the decreasing rate of decline in the tail. For example, the negative exponential regression predicts a 2017 median of 20.0 lightning deaths, while the linear regression predicts -3.1 deaths—negative deaths are obviously meaningless.

The final update to the inter-annual analysis added the 2.5th and 97.5th percentiles, in addition to the previous percentiles, so that the entire set is: 2.5th, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 97.5th percentiles. The two new percentiles allow two-tail hypothesis testing at the 95% confidence level. As in the 2012 study, these percentiles were estimated from calculating each percentile for the annual lightning fatalities for each year (1995-2004) ±10 years (Figure-2). The period ends at 2004 to allow for the +10 year part of the estimate since the available data ends in 2014. Next, a best-fit negative exponential curve was calculated for each percentile (Table-2). Normally, more than 21 data points would be desired for a best-fit curve and including earlier years was considered. However, in this case, the choice was made to use the smaller sample of the 1995-2004 period, rather than 1984-2003, to avoid some internal inconsistencies in the results due to some discontinuities in the data in 1993 and earlier.

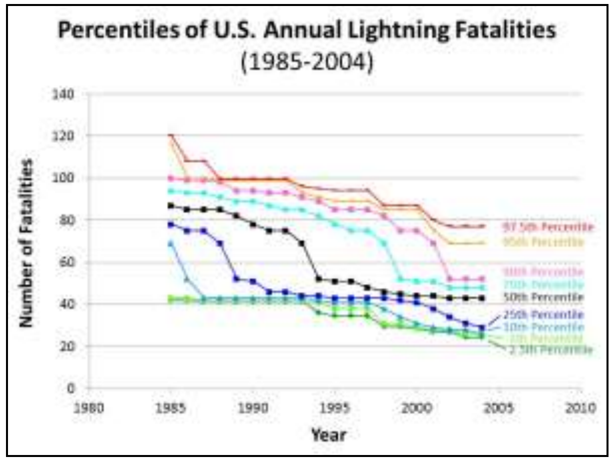

Figure-2 Running 21-year percentiles (center year ± 10 years) of U.S. lightning fatalities for 1984-2004.

Table-2. Equations for various percentiles of the number of U.S. annual lightning fatalities (1995-2004). The 2.5th and 97.5th percentiles form a 95% confidence interval.

Percentile	Equation	r^2
2.5th	$y = 711.30e^{-0.032(x-1900)}$	0.90
5th	$y = 588.14e^{-0.030(x-1900)}$	0.85
10th	$y = 1148.90e^{-0.036(x-1900)}$	0.83
25th	$y = 3013.30e^{-0.045(x-1900)}$	0.87
50th (median)	$y = 4889.10e^{-0.047(x-1900)}$	0.92
mean	$y = 2283.00e^{-0.038(x-1900)}$	0.99
75th	$y = 3605.60e^{-0.041(x-1900)}$	0.88
90th	$y = 1814.50e^{-0.033(x-1900)}$	0.80
95th	$y = 759.41e^{-0.023(x-1900)}$	0.88
97.5th	$y = 594.84e^{-0.020(x-1900)}$	0.92

y = expected number of fatalities; x = year

For example, using the 1984-2003 results from the previous 2012 analysis led to the 2.5th percentile being larger than the 5th percentile in some future years. Using new 1995-2004 also made the results more representative of the current lightning fatality rate. This may be important since there is evidence that the U.S. lightning fatality rate changed beginning in 2008 (see section-3). A graph of the median, and 2.5th and 97.5th percentiles (95% confidence interval) and the best-fit curves projected to 2014 is shown in Figure-3.

Due to the variability in the data and the closeness of some of the percentiles, some of the equations may still produce internally inconsistent results, e.g. the predicted 2.5th percentile can be slightly larger than the 5th percentile for some years. In practical application, if one predicted percentile is inconsistent with another, consider using the more conservative of the two.

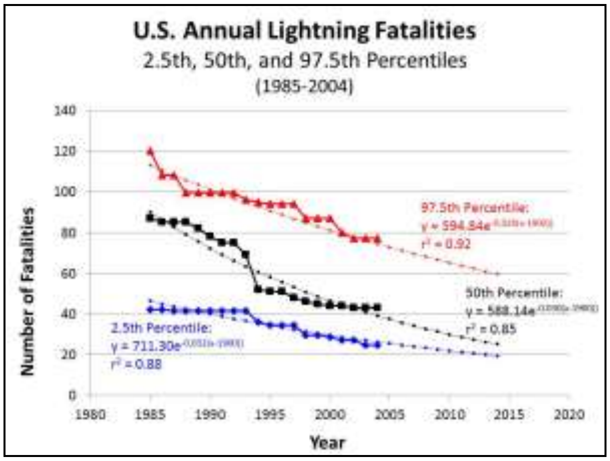

Figure-3. The 2.5th, 50th, and 97.5th percentiles of U.S annual lightning fatalities for 1985-2004 and best-fit negative exponential curves extrapolated to 2014.

As an example, the prediction for 2015 is a median number of U.S. lightning fatalities of 22.0 with a 95% confidence interval of 17.9 to 59.6 fatalities (2.5th and 97.5th percentiles, respectively). These estimates are calculated from the equations in Table-2. Note that for 2014, the last year for which data are available, the model would have predicted a median of 23.0 fatalities with a 95% confidence interval of 18.5 to 60.8 fatalities. This compares very well with the observed number of 26 lightning fatalities.

2.b. Updates to Intra-Annual Fatalities

The updates to the intra-annual study were more substantial than those to the inter-annual study. The 2012 study provided the median dates for selected percentiles of U.S. lighting fatalities based on the dates of those fatalities from 2006-2011. The update in this paper adds the dates of the lightning fatalities from 2012-2014 to the analysis, a 33% increase in number of years. The data are from the NWS lightning safety database (NWS, 2015b) and discussed in Jensenius (2014), Roeder and Jensenius (2013), Roeder and Jensenius (2012a), and Roeder and Jensenius (2012b). In addition, a logistic regression on the percentile dates was added for a more robust overall result. Finally, the 2.5th and 97.5th percentiles were added, to provide a 95% confidence interval, in addition to the previous percentiles.

The distribution of U.S. intra-annual lightning fatalities was determined as follows. The dates for nine percentiles were calculated for each of 9 years 2006-2014: 2.5th, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 97.5th percentiles. The median for each percentile was calculated from the 9 years (Table-2). A best-fit logistic curve was then used to estimate the distribution of these percentiles to ensure consistency and to represent a typical year (Equation 5 and Figure-4). Note that the middle of the U.S. lightning fatality year is 15 July.

A generalized logistic curve is shown in (4).

$$y = 100*(1/(1 + e^{-(\alpha + \beta*(x - \delta))})) \quad (4)$$

where y = percentile (%),
x = day of the year,
e.g. for 1 Jan, x = 1; for 1 Feb, x = 32; and for 31 Dec, x = 365.
α, β, δ are the distribution parameters.

A best-fit logistic curve cannot be calculated analytically and so the best-fit curve was done manually. The three distribution parameters (α, β, δ) were iterated sequentially until the RMSE of the median percentiles from the curve was minimized. Once the first round of optimized parameters was obtained, the sequence of iteration had to be repeated since a change in one parameter changes the optimum value of the other parameters. The sequence of iterative optimization was repeated until the parameters were accurate to four decimal places, i.e. changed by less than 0.00005. The best-fit logistic curve is in (5).

$$y = 100*(1/(1 + e^{-(0.0315 + 0.0420*(x - 195.5400))})) \quad (5)$$

where y = percentile (%),
and x = day of the year.

This best-fit logistic curve had a RMSE of 1.23% and a bias of 0.3%, indicating a good match to the intra-annual percentiles calculated from 2006-2014. This good match can also be seen in Figure-4.

Table-2. Percentiles for the intra-annual of lightning fatalities in the U.S. The percentiles are the median of the percentiles observed in each year for 2006-2014.

Percentile	Median Date
0	1 Jan
2.5th	25 Apr
5th	2 May
10th	22 May
25th	16 Jun
50th	15 Jul
mean (standard deviation)	15 Jul (45.6 days)
75th	9 Aug
90th	5 Sep
95th	18 Sep
97.5th	26 Sep
100th	31 Dec

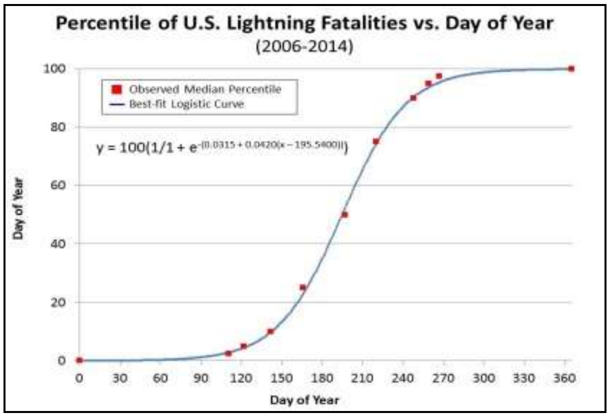

Figure-4. Best-fit logistic curve for the distribution of intra-annual lightning fatalities in the U.S. (2006-2014).

2.c. Applications of Inter-Annual and Intra-Annual Fatality Statistics

The inter-annual and intra-annual analyses can be combined to predict the expected number of accumulated lightning fatalities and their error bars for any day of any year in the U.S. For example, the inter-annual analysis predicts an expected median of 22.0 lightning fatalities in the U.S. for 2015. Through 20 June, the intra-annual analysis indicates that 26.9% of the annual lightning fatalities should have accumulated. Therefore, the expected median number of U.S. lightning fatalities by the end of 20 June 2015 is 5.9 fatalities (26.9% of 22.0).

As another example, the expected 2.5th and 97.5th percentiles of lightning fatalities for 2015 are 17.9 and 59.6 fatalities, respectively, and form a 95% confidence interval. Multiplying these numbers by the intra-annual percentile for 20 June (26.9%) gives a 95% confidence interval of 4.8 to 16.0 lightning fatalities for that date.

An EXCEL tool was built to automate the most useful parts of these analyses (Figure-5). The user needs to only enter the desired date in the yellow shaded cell and the rest of the data are calculated. This tool is useful for monitoring the development of lightning casualties in the U.S. as the year develops, which has some utility in lightning safety education. For example, around mid-2013, it became apparent that 2013 could set a new record low for U.S. lightning fatalities, but that new record might not be statistically significant at the 95% level. This eventually verified with a new record low of 23 observed fatalities. The expected 5th percentile was 21.5 fatalities (using the previous version of the tool, since the new years of data to update the tool had not yet occurred), so the observed value was not statistically different from the expected value at the 95% significance level (1-tail hypothesis test), though it was close—if only two fewer lightning casualties had occurred, the new record would have been statistically significant. A separate EXCEL spreadsheet (not shown) was developed to calculate the probability of a new U.S. record low for the year occurring based on the observed number of lightning fatalities by the date of interest.

Expected U.S. Lightning Fatalties

For Entire Year (POR: 1984-2014)

Date	2.5th Percentile	5th Percentile	Median (50th Percentile)	95th Percentile	97.5th Percentile
20-Jun-15	17.9	18.7	22.0	53.9	59.6
			Mean	Number	
			28.9	25.3	

For This Date (POR: 1984-2004)

Day Of Year	Percentile of Annual Lightning Fatalities (logistic regression)	Expected Number of U.S. Lightning Fatalities by this Date				
171	26.9	2.5th Percentile	5th Percentile	Median	95th Percentile	97.5th Percentile
		4.8	5.0	5.9	14.5	16.0
				Mean	Number	
				7.8	6.8	

Figure-5. Example of EXCEL tool to calculate the expected number of U.S. lightning fatalities by a desired date and desired year and various error bars for those expected numbers.

3. POSSIBLE CHANGE IN U.S. LIGHTNING FATALITIES BEGINNING IN 2008:

There is some evidence that the pattern of U.S. lightning fatalities may have changed starting in 2008. However, 2014 offered some contrary evidence.

3.a. Evidence for the Change

The annual lightning fatalities in the U.S. for 2002-2013 are in Figure-6 (6 years of the changed fatality rate (2008-2013) and 6 years before the change (2002-2007)). A sudden drop in fatalities and decrease in year-to-year variability may have begun in 2008. The drop in fatalities is more apparent if one considers 2004 to be a low outlier for 2002-2007. Linear regressions for the two periods are shown to help shoe the change and suggest the drop was 10.0 fatalities per year.

The mean for 2002-2007 is 42.7 fatalities with a standard deviation of the mean of 2.95. For 2008-2013 the mean is 28.0 with a standard deviation of the mean of 1.48 (Figure-7). A two-tailed Student's t-test with unequal variances for a null hypothesis of equal means (Wilks, 2006) is rejected with a p-value = 0.0067. This strongly supports that the lightning fatality rate in the U.S. did indeed change beginning in 2008. A one-tailed hypothesis test assuming a lower 2008-2013 mean was not done since there is no a priori evidence to suggest a decrease vs. increase in lightning fatalities.

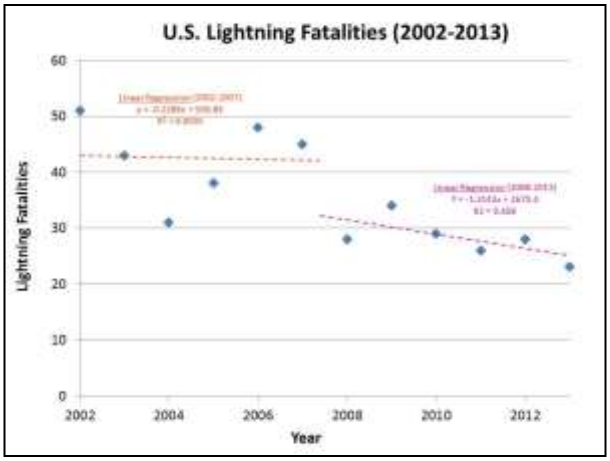

Figure-6. Annual lightning fatalities in the U.S. from 2002-2013. Note the sudden change to lower fatalities beginning in 2008. Linear regressions for 2002-2007 and 2008-2013 are shown to help indicate the change. The decrease was about 10.0 fatalities.

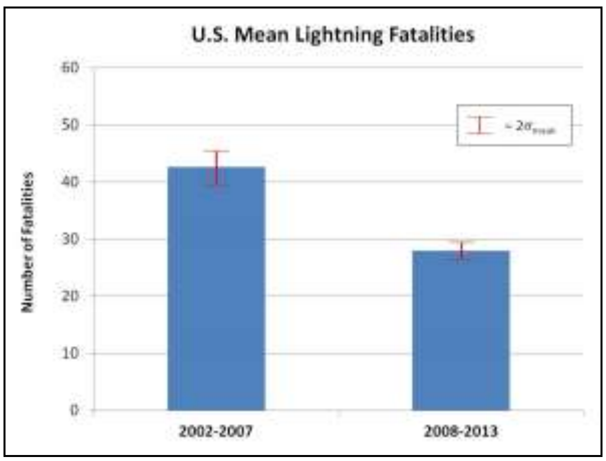

Figure-7. Mean, standard deviation, and standard deviation of the mean for 2002-2007 and 2008-2013. The error bars are <u>two</u> standard deviations <u>of the means</u> and show visually that means are very different statistically, as confirmed by the t-test ($p = 0.0067$).

More evidence suggests that the U.S. lightning fatality rate suddenly lowered in 2008. Each of the six years (2008-2013) was less than expected, as compared to the previous best-fit exponential curve (Roeder, 2012). None of these less-than-expected events was statistically significant, even though two of those years set new record lows. However, taken as a whole, those six consecutive years of fatalities below expectations are statistically significant at nearly the 98.4% level. If the distribution of fatalities was random about the expected values, the probability of six consecutive observations below expectation is $(0.5)^6$, or 1.6%. If this trend continues, a new statistical model of the expected lightning fatalities should be developed, restricting the analysis to the 2008 and later years.

Some members of the lightning safety community have speculated on the cause of this apparent sudden change in U.S. lightning fatalities. The cause was first thought to have been due perhaps to the economic recession that began in 2007-2008—people may have been spending less time doing at-risk outdoor recreation to save money. However, if that were the cause, then the lightning fatalities should return toward expectations as the economy recovers. Since that return to expectations has not been observed (2013 was a new record low in U.S. lightning fatalities), the economic recession may not have been the primary cause of the change in lightning fatalities, or at least not solely the cause. Further speculation considered that the cause may have been a combination of the economic recession and NOAA's lightning safety education campaign that began in 2001 (Jensenius and Franklin, 2012; Jensenius and Franklin, 2008; Hodanish et al., 2008; Jensenius et al., 2008; Jensenius and Franklin, 2006a; Jensenius and Franklin, 2006b). The population may have been primed to change their lightning safety behavior by the education campaign, then once the recession precipitated a change in behavior, that change persisted even after the recession ended.

Another possible cause would be a shift in the amount of lightning in the U.S. However, it is assumed that a significant enough reduction in lightning activity has not occurred over the six year period. For example, there was an intense drought of the U.S. in 2012, especially the Mid-West, with an assumed drop in lightning activity, yet 2012 did not show a corresponding drop in lightning fatalities compared to the previous or subsequent year.

3.b. Evidence Against the Change or That the Change Ended

However, this pattern of below median fatalities may have ended in 2014. The expected median lightning fatalities in the U.S. for 2014 was 25.2 deaths. The observed fatalities were slightly higher, though barely, at 26 deaths. In addition, if the expected number of fatalities (equation 1) were used (26.4 deaths) rather than the expected median, then the observed fatalities would have been less than expected, though barely, continuing the pattern of the previous 6 years. So while 2014 offers evidence against the change, or that the change ended, that evidence is weak.

3.c. Status of the Possible Change

It is now unclear if the new trend of lower fatalities that appeared to begin in 2008 has ended, or was even real, or if the weak break in the pattern in 2014 was a fluke or a return to the previous pattern. More data from future years will be needed to resolve the issue.

4. FUTURE WORK:

The analysis should be updated as more annual data accrue. This is especially true for the intra-annual distribution of lightning fatalities since only 9 years of data have been available so far (2006-2014).

The annual percentile distributions also could be improved. The choice of period record for the percentiles and ± interval over which to measure each annual percentile should be optimized to find the best balance between sample size and making the results more representative of the present.

In addition, the same analysis should be done using per capita lightning fatalities, rather than the observed U.S. lightning fatalities. This should provide a small improvement in the predictions. Since the population has been increasing during this period, the per capita lightning fatalities will indicate a slightly faster drop in the death rate from lightning hazard.

The possible sudden drop in U.S. lightning fatalities in 2008 and possible end of the pattern in 2014 should be studied further. The annual rates should be monitored to see if they remain below expectation. If so, then a new model of expected lightning fatalities, representative of the new rate, should be developed. In addition, the annual lightning flash rate in the U.S. should be investigated to see if changes in flash rate help explain the changes in fatalities. Finally, social scientists should be consulted to try to explain the observed changes in U.S. annual lightning fatalities beginning in 2008.

5. SUMMARY:

An update to the statistical analysis of annual rate of lightning fatalities in the U.S. was presented, including an update to the method estimating the total number of expected lightning fatalities in the U.S. for any year in the near future and the expected number of fatalities on any day during any year. The uncertainty in these estimates was also updated for use as error bars and to allow hypothesis testing. The model predicts a median of 22.0 lightning fatalities in the U.S. for 2015 with a 95% confidence interval of 17.9 to 59.6 fatalities. The apparent sudden decrease of U.S. lightning fatalities beginning in 2008 and possible return to the previous rate in 2014 was discussed.

6. REFERENCES:

Ashley, W. S., and C. W. Gilson, 2009: A reassessment of U.S. lightning mortality, *Bulletin of the American Meteorological Society*, **90**, Oct 09, 1501-1518

Cooper, M. A., 2012: A brief history of lightning safety efforts in the United States, *4th International Lightning Meteorology Conference*, 4-5 Apr 12, 8 pp.

Hodanish, S. J., K. Torgerson, J. S. Jensenius, M. A. Cooper MD, M. Utley, and W. P. Roeder, 2008: Leon the Lightning Safety Lion says: "When Thunder Roars, Go Indoors!"--NOAA's Efforts Regarding Children's Lightning Safety, *3rd Conference on Meteorological Applications of Lightning Data*, Paper 2.15, 4 pp.

Holle, R. L, 2012: Recent studies of lightning safety and demographics, *4th International Lightning Meteorology Conference*, 4-5 Apr 12, 15 pp.

Holle, R. L., R. E. Lopez, and B.C. Navarro, 2005: Deaths, injuries, and damages from lightning in the United States in the 1890s in comparison with the 1990s, *Journal of Applied Meteorology*, **44**, 1563-1573

Jensenius, J., 2014: A detailed analysis of recent lightning deaths in the United States, *5th International Lightning Meteorology Conference*, 20-21 Mar 14, 8 pp.

Jensenius, J., and D. Franklin, 2012: A review of NOAA's lightning safety awareness campaign and its impact across the United States, *4th International Lightning Meteorology Conference*, 4-5 Apr 12, 8 pp.

Jensenius, J., and D. B. Franklin, 2008: Lightning kills – play it safe: NOAA's continuing efforts to educate the public on the dangers of lightning, *2nd International Lightning Meteorology Conference*, 24-25 Apr 08, 7 pp.

Jensenius, J., and D. B. Franklin, and S. Hodanish, 2008: Lightning kills – play it safe: NOAA's efforts to educate the public on the dangers of lightning, *3rd Conference on Meteorological Applications of Lightning Data*, Paper 5.2, 5 pp.

Jensenius, J., and D. B. Franklin, 2006a: Lightning kills – play it safe: NOAA's efforts to educate the public on the dangers of lightning, *19th International Lightning Detection Conference*, 24-25 Apr 2006, 7 pp.

Jensenius, J., and D. B. Franklin, 2006b: Lightning kills – play it safe: NOAA's efforts to educate the public on the dangers of lightning, *1st International Lightning Meteorology Conference*, 26-27 Apr 2006, 7 pp.

Lopez, R. E., and R. L. Holle, 1998: Changes in the number of lightning deaths in the United States during the twentieth century, *Journal of Climate, Vol. 11, No. 8*, Aug 98, 2070-2077

National Weather Service, 2015a: NOAA, National Weather Service Office of Climate, Water, and Weather Services, 1325 East West Highway, Silver Spring, MD 20910, *www.nws.noaa.gov/om/hazstats.shtml*, 2015

National Weather Service, 2015b: NOAA, National Weather Service Lightning Safety, 1325 East West Highway, Silver Spring, MD 20910, *http://www.lightningsafety.noaa.gov/victims.htm*, 2015

Roeder, W. P, 2013: Estimating the expected number of U.S. lightning fatalities during a year, throughout a year, and comparison to other storm phenomena, *6th Conference on Meteorological Applications of Lightning Data,* Paper P1.4, 6-10 Jan 2013, 1 pp.

Roeder, W. P., 2012: A statistical model for the inter-annual and intra-annual fatalities from lightning in the U.S. and comparison to other storm phenomena, *4th International Lightning Meteorology Conference*, 4-5 Apr 12, 6 pp.

Roeder, W. P., and J. Jensenius, 2013: NOAA's new lightning fatality database and Applications in lightning safety education, *6th Conference on Meteorological Applications of Lightning Data,* Paper P1.3, 6-10 Jan 2013, 1 pp.

Roeder, W. P., and J. Jensenius, 2012a: A new high-quality lightning fatality database for lightning safety education, *International Lightning Meteorology Conference*, 4-5 Apr 12, 9 pp.

Roeder, W. P., and J. Jensenius, 2012b: Lightning safety applications of the new NOAA lightning fatality database, *37th Annual Meeting of the National Weather Association*, P2.30, 10 pp.

Wilks, D. S., 2006: Statistical methods in the atmospheric sciences, 2nd edition, *Academic Press, Inc.*, Vol. 91 International Geophysics Series, pp. 627