# The 2017 Statistical Model for U.S. Annual Lightning Fatalities

William P. Roeder
Private Meteorologist
Rockledge, FL, USA

## 1. Introduction

The annual lightning fatalities in the U.S. have been generally declining since the 1940s as shown in Figure-1 (Roeder, 2015; Roeder, 2014; Roeder, 2013; Roeder, 2012; Holle, 2012; Ashley and Gilson, 2009; Holle et al., 2005; Lopez and Holle, 1998). This declining rate creates difficulties in estimating the current fatality rate. Traditionally, the National Weather Service uses a running 30-year mean for annual fatalities from various weather phenomena. While this is appropriate for weather fatality rates that have no overall trend in time, it can lead to misleading results for fatality rates that are changing over time. Since U.S. annual lightning fatality rates have been declining, the 30-year running mean overestimates the current rate. In such cases, the National Weather Service also uses a running 10-year mean. While a lag still exists, it yields a rate more representative of the current rate. However, a running 10-year mean is overly sensitive to extreme events. A better approach is to recognize the lightning annual fatality rates are varying over time and to model that trend. This results an estimate that has no lag, i.e. is representative of the current year, and is also consistent with all the previous years. A best-fit negative exponential curve has been previously used to model this trend and percentile regression used to model the error bars. In addition, logistic regression was used to model how the U.S. annual lightning fatalities accumulate throughout the year.

This paper reports new results that add three more years (2014-2016) of U.S. lightning fatalities to the analysis and is labeled the ‘2017’ version of the model. In addition to the additional years, a sudden drop in the annual lightning fatalities and a decrease in annual variability was identified that began in 2008. The data prior to 2008 were excluded from the updated model since they were no longer representative of the current time. While more than a 9-year period would normally be desired to model annual rates, the data prior to 2008 appear to be no longer representative of the current time.

Two of new years of data (2014-2015) reinforce the previous evidence that another discontinuous drop in U.S. annual lightning fatalities and the annual variability began in 2008. Unfortunately, 2016 did not confirm the apparent new trend having a statistically significant higher than expected number of lightning deaths. It is not known if 2016 was an outlier, or a return to the previous pattern, or a result of changes in National Weather Service lightning safety education in 2016, or if other explanations are needed.

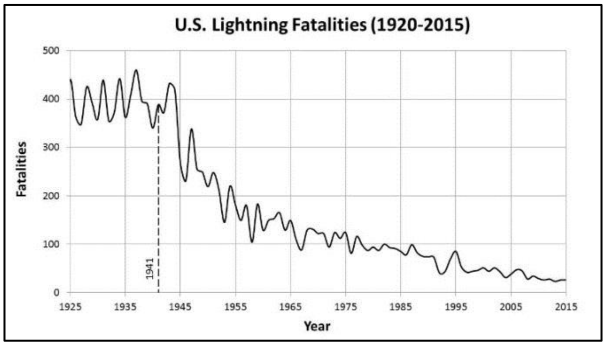


**Figure-1.** U.S. annual lightning fatalities for 1940-2015. Lightning fatalities began a general decline around 1941 that has continued through 2016. Data from 1900-1991 are from Lopez and Holle (1998), 1992-2005 are from National Weather Service (2016a) and the 2006-2016 data are from National Weather Service (2016b).

## 2. The 2017 Model

The history and development of the original model and subsequent updates is summarized in Roeder (2016). This paper focuses on the new 2017 model.

The 2017 model identified a change in the pattern of U.S. annual lightning fatalities that began in 2008 (Figure-2). In addition to a reduction in annual lightning deaths, the year-to-year variability also appeared to decrease. As a result, the new 2017 model restricted the data used to be from 2008-2016. While more than 9 years is desired for the model, the earlier data is no longer representative of the current time.

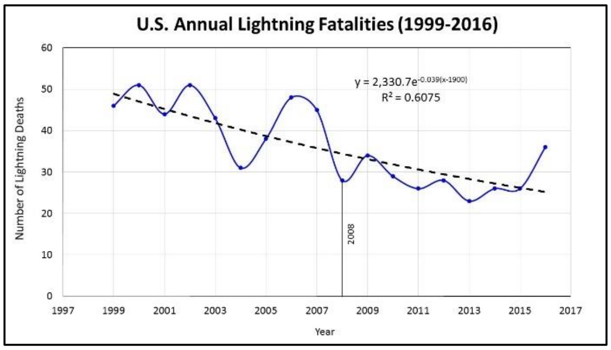


**Figure-2.** The U.S. lightning fatalities U.S. annual lightning fatalities for 1999-2016. The pattern appears to have changed starting about 2008, dropping to a lower rate with less interannual variability.

The 2017 model also adds three more years to the last update (2014-2016) to cover 2008-2016. The model update includes a new best-fit negative exponential curve for the total annual lightning deaths (Table-1). In addition, a new logistic regression for the accumulated percentage of lightning deaths expected throughout the year was also developed (Equation-1) (Table-2) (Figure-3). Since there is no evidence that the pattern of accumulation of lightning fatalities during the year has changed, this curve uses the full available data from 2006-2016 (NWS, 2016b). Both of these curves include 95% confidence intervals. By combining these curves, one can predict the expected number of U.S. lightning fatalities for any date for any future year and the expected 95% error bar for that prediction. A spreadsheet was built that does these calculations automatically—just enter a date and the spreadsheet returns the expected number of U.S. lightning deaths expected by that date, total for that year, and error bars for both (Figure-4).

**Table-1**

The 2017 statistical model of U.S. annual lightning fatalities. The percentiles and mean are based on ±10-year intervals for the years based on 2008-2016. Use of the median for predictions is recommended to minimize errors.

| Parameter | Best-Fit Equation | $r^2$ |
|---|---|---|
| 2.5 Percentile | $y = 477.11e^{-(0.028(x-1900))}$ | 0.898 |
| 5.0 Percentile | $y = 370.98e^{-(0.025(x-1900))}$ | 0.889 |
| 10.0 Percentile | $y = 2{,}935.3^{-(0.045(x-1900))}$ | 0.874 |
| 25.0 Percentile | $y = 23{,}075e^{-(0.064(x-1900))}$ | 0.957 |
| Median | $y = 252.34e^{-(0.071(x-1900))}$ | 0.768 |
| Mean | $y = 2{,}434.3e^{-(0.039(x-1900))}$ | 0.988 |
| Expected Number | $y = 360.79e^{-(0.021(x-1900))}$ | 0.131 |
| 75.0 Percentile | $y = 2{,}597.6e^{-(0.039(x-1900))}$ | 0.690 |
| 90.0 Percentile | $y = 25{,}855e^{-(0.059(x-1900))}$ | 0.846 |
| 95.0 Percentile | $y = 14{,}007e^{-(0.052(x-1900))}$ | 0.812 |
| 97.5 Percentile | $y = 4{,}604.4e^{-(0.040(x-1900))}$ | 0.599 |

y = predicted value
x = year of interest

**Equation-1**

Best-fit logistic regression for the percentage of U.S. lightning fatalities throughout the year (2006-2016). RMSE = 1.23%

$$P = 100[1/\{1 + e^{-(0.0304 + 0.0425(D - 194.8325))}\}]$$

where P = percentage of the U.S. lightning fatality season,
and
D = day of the year, e.g. 31 June = 182

**Table-2**

Median date for various percentiles of accumulated annual U.S. lightning fatalities. Since there is no change in the accumulation of deaths during the year, the full time period for which data are available is used (2006-2016) (National Weaher Service (2016b).

| Percentile | Day Of Year | Date |
|---|---|---|
| 0th | 1 | 1 Jan |
| 2.5th | 111 | 21 Apr |
| 5th | 122 | 2 May |
| 10th | 142 | 22 May |
| 25th | 166 | 15 Jun |
| 50th | 196 | 15 Jul |
| 75th | 220 | 8 Aug |
| 90th | 248 | 5 Sep |
| 95th | 254 | 11 Sep |
| 97.5th | 267 | 24 Sep |
| 100th | 365 | 31 Dec |

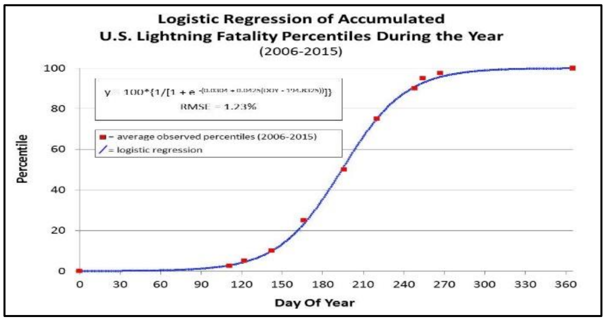
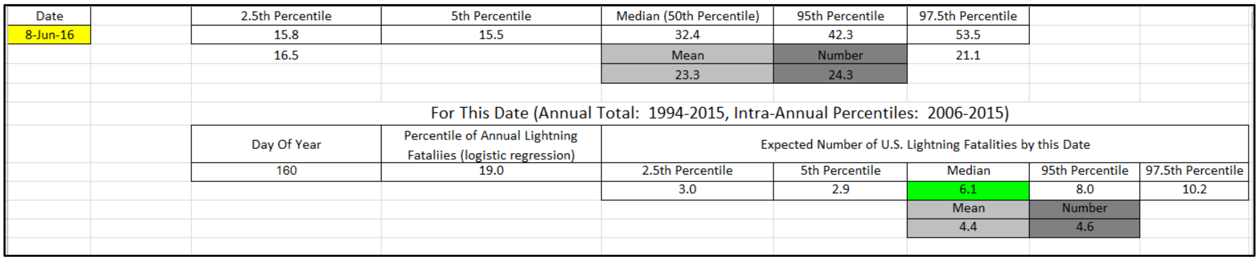


**Figure-3.** The logistic regression for the distribution of U.S. lightning fatalities throughout the year.



| Date | | 2.5th Percentile | 5th Percentile | Median (50th Percentile) | 95th Percentile | 97.5th Percentile |
|---|---|---|---|---|---|---|
| 8-Jun-16 | | 15.8 | 15.5 | 32.4 | 42.3 | 53.5 |
| | | 16.5 | | Mean | Number | 21.1 |
| | | | | 23.3 | 24.3 | |

For This Date (Annual Total: 1994-2015, Intra-Annual Percentiles: 2006-2015)

| Day Of Year | Percentile of Annual Lightning Fataliies (logistic regression) | Expected Number of U.S. Lightning Fatalities by this Date | | | | |
|---|---|---|---|---|---|---|
| 160 | 19.0 | 2.5th Percentile | 5th Percentile | Median | 95th Percentile | 97.5th Percentile |
| | | 3.0 | 2.9 | 6.1 | 8.0 | 10.2 |
| | | | | Mean | Number | |
| | | | | 4.4 | 4.6 | |

**Figure-5**. The spreadsheet that provides the expected number of U.S. lightning fatalities accumulated by a specified date. The user enters the desired date (yellow) and the spreadsheet returns the expected number of U.S. lightning fatalities by that date (green) and its 95% confidence interval. The total number expected for the entire year and its 95% confidence interval is also provided, along with other information.

The final update was a new visualization of the observed versus predicted number of lightning deaths for tracking how the U.S. lightning deaths are accumulating through the year (Figure-5). These new graphics allow a newly recognized application of identifying if the rate of accumulating lightning fatalities is statistically significantly different than expected. If the rate of change in observed lightning fatalities is higher/lower than the upper/lower bound of the 95% confidence interval for the same period, then the observed lightning fatality is statistically significantly different than expectation. Care must be taken to observe the accumulating lightning fatalities over a long enough period to be representative, otherwise even am normal one/no lightning death(s) in 1 day could be identified as statistically high/low. Conversely, one must not observe the accumulating lightning deaths too long are important trends and opportunities to take action may be missed. Anecdotally, an approximately 2-week observing period may be useful. Note that the new visualization for lightning fatalities, discussed below, are a useful way to visually identify possibly statistically significant deviations in the lightning fatality rate.

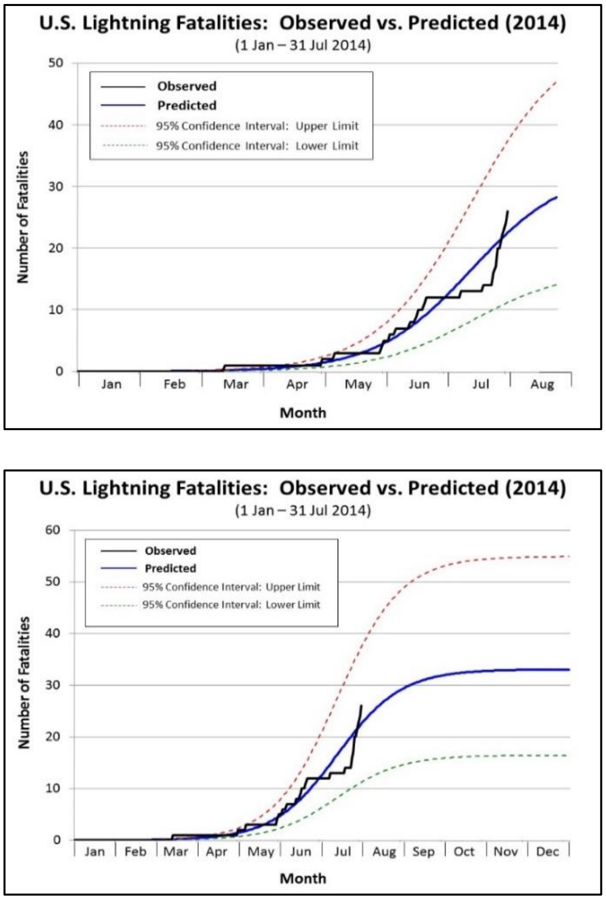


**Figure-5.** The new graphics showing the observed U.S. lightning fatalities as they accumulate throughout the year for easy comparison to the model predictions. Two versions are used, one through the current date and about 4 weeks further for detailed examination (top), and one through the current date and the end of year (bottom).

## 3. What Happened in 2016?

As discussed above, there was strong evidence that a new pattern of fewer and less variable U.S. annual lightning fatalities began in 2008. The data from 2008-2015 supported that conclusion. However, 2016 proved to be an exception. The year of 2016 had 38 lightning deaths which was statistically significantly higher than the previous model prediction of 26.3 deaths with a 95% confidence interval of 18.9 to 35.8 deaths. That was the first time in 9 years that the observed annual lightning deaths was statistically significantly different than the model's prediction.

There are three main possibilities for the 2016 lightning fatalities being higher than expected. Firstly, it may just be a statistically fluke with an extreme value happening sooner than expected from the confidence interval. If so, perhaps the 2017 number of lightning fatalities will regress toward the mean and be closer to the model's prediction. Secondly, it is possible that the change in pattern in U.S. lightning fatalities that apparently began in 2008 has reverted back to the previous pattern as had been speculated if the 2008 change was due to the economic 'great recession' and as the economy recovered. Or perhaps the 2008 change was not real, despite the strong statistical evidence.

A third possibility is that the 2016 event was due to changes in the National Weather Service (NWS) lightning safety education program. In 2016, NWS changed most of their weather safety programs from using national weeks to seasonal/regional weeks. This included the national 'Lightning Safety Week' that had previously been held the last week of June. This may have resulted in less lightning safety information reaching the public. More specifically, the NWS did not increase lightning safety messaging when a large increase in the rate of lightning fatalities occurred in Aug 2016 (Figure-6). Note that the rate of lightning fatalities in Aug 2016 was higher than the upper bound of the 95% confidence interval at that time and so the rate in Aug 2016 was statistically significantly higher than expected. As discussed above, the NWS reacted to a surge in lightning fatalities in 2015 with increased lightning safety messaging that appeared to be very successful. However, NWS did not take similar action with the 2016 surge.

The root cause is unclear for why 2016 had a much higher than expected number of lightning fatalities in the U.S. We must wait for more evidence in 2017 and subsequent years before making a judgment.

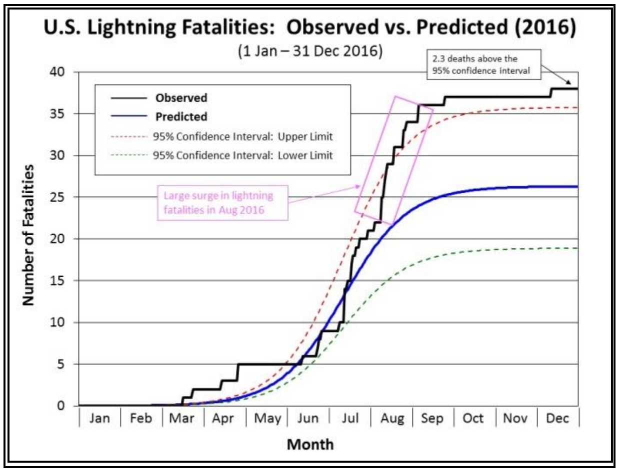


**Figure-6.** The pattern of observed U.S. lightning fatalities for 2016 as compared to the model predictions. Note the large increase in lightning deaths during Aug 2016.

## 4. Applications

There are many applications for the updated equations and graphic for predicting the U.S. annual lightning fatalities and distribution throughout the year. Most importantly, the 95% confidence intervals allow easy identification if the observed lightning fatalities are statistically significantly different from the expected value at the 5% significance level. The other percentiles allow other significance levels. For example, 2011 had a then record low of 26 lightning fatalities in the U.S. This received considerable press coverage. However, the 95% confidence interval for the model at that time was 22.1 to 66.5 deaths. Therefore, even though that year had a record low, it was not statistically significantly different than the expected median of 27.0, though it was very close. Another new record low of 23 lightning deaths was set in 2013. This new record was also not statistically significant since the median and 95% confidence interval was 24.6 and 20.8 to 63.6 deaths.

A second application is identifying if the accumulating lightning fatalities during the year are statistically significantly different than expected. Usually one should wait until enough lightning deaths have occurred before identifying significant trends, typically in June. An example of the utility of tracking accumulating lightning fatalities occurred in 2015. The new graphic (discussed at the end of section-II) showed a sudden upward trend in U.S, lightning fatalities from 21-27 June (Figure-7 (top)). Although not tested formally, the change in trend appears to be statistically significant; its slope is steeper than the slope of the upper bound of the 95% confidence interval for those dates. As a result, the National Weather Service (NWS) increased lightning safety education messaging at that time. The fatality rate quickly dropped thereafter (Figure-7 (bottom)), although a causative effect cannot be scientifically proven. This reduction in lightning fatality appeared to be even more statistically significant than the initial increase. When the new graphic flagged the increase in lightning fatalities, Vaisala, Inc. investigated if the increase for those dates was due to a jump in lightning activity or in frequency of positive polarity high current lightning (Holle, 2015). Vaisala saw no change in U.S. lightning to explain the increase or subsequent decrease in lightning fatalities. This indirectly bolsters the supposition that the NWS messaging caused the decrease in lightning fatality rate. The three main causes of lightning fatality are lightning flash density, population density, and the lightning safety behavior of the population (Roeder et al., 2015). Vaisala indicated that a change in lightning flash density was not involved. Population shifts of this magnitude and timing are unreasonable. Therefore, the likely cause was a change in the lightning safety behavior of the population, presumably due to the NWS's increase in lightning safety messaging.

The third and final application discussed here is increasing interest in lightning safety education by providing timely input. If lightning fatalities are above expectations at the current date, then that can be used in lightning safety education to increase lightning safety awareness and motivate increased good lightning safety practices. Conversely, if the lightning fatalities are below expectations at the current date, that can be used to encourage continued good lightning safety practices.

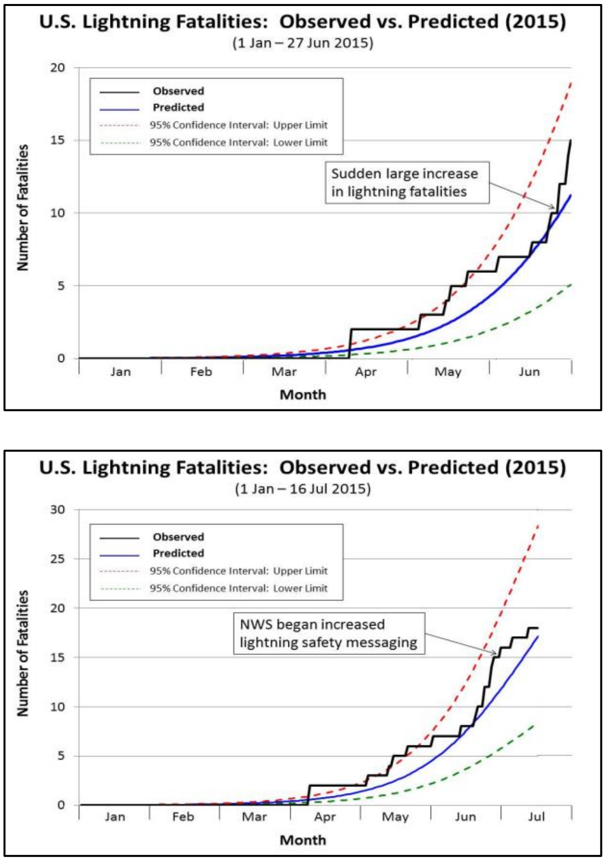


**Figure-7.** The U.S. lightning fatalities began a sudden increase 21-27 June 2015 (top). The NWS increased lightning safety messaging around 27 June 2015 and the lightning fatality decreased soon thereafter (bottom), although a causitive effect cannot be scientifically proven.

## 5. Future Work

The new 2008-2015 model for U.S. annual lightning fatalities only covers 8 years. This model should continue to be updated as more data accumulates. In addition, although a new pattern strongly appears to have began in 2008, the U.S. annual lightning fatalities should continue to be monitored to ensure that the new pattern is persisting.

A statistical test should be developed for significant changes in the lightning fatality rate for any period during the year. As we saw in the second application above, the new graphic for acumulating U.S. lightning fatalities can also be used for detecting changes in the rate of of those fatalities. A statistical test for significant changes in this rate is needed. The optimum observing period should be determined for identifying statistically significant deviations in the rate of accumulating lightning fatalities. A too short of an observing period will lead to misleading decisions and a too long a period will miss opportunities for action to intervene on lightning fatalities that are accumulating much faster than expected. Anecdotally, an approximately 2-week observing period may be useful.

The large confidence intervals are due to the influence of large values from the mid-1980s to early-1990s. Over the next 4-7 years, those influences will decrease and completely dissappear. The model should be updated at the start of that period, and again at the end of that period to significantly improve the error estimates of the model.

A final issue is that U.S. annual lightning fatalities have entered the tail of their negative exponential distribution. As a result, the lightning fatalities are not declining as fast as previously and certainly not as fast as desired. This suggests a change in lightning safety education should be considered. If one conceedes that the lightning fatalities cannot be significantly decreased, then perhaps lightning safety education should be changed to a maintenance style of education—lightning safety awareness has been well established, so perhaps different education techniques would be better at maintaining that awareness, as opposed to the current methods used to build that increased awareness. Or perhaps new education methods are needed to significantly reduce the U.S. annual lightning fatalities further.

## 6. SUMMARY

Two updates to the previous model of U.S. annual lightning fatalities are presented. Both new models continue to use a best-fit negative exponential curve rather than 30-year or 10-year running means that are typically used by the NWS. The first new model restricts the analysis to 1994 onward since the pattern of U.S. lightning patterns changed starting in 1994 so the new model is more representative of the current time. The change was a sudden decrease in the number of fatalities and a decrease in the inter-annual variability. In addition, the two most recent years of observed fatalities were added to the analysis so that the period covers 1994-2015. The equations for predicting the distribution of U.S. annual lightning fatalities for this 1994-2015 model are in Table-I.

The logistic regression for the typical distribution of U.S. lightning fatalities during the year was also updated to include the two most recent years. The period now covers 2006-2015. The best-fit logistic regression for how lightning fatalities accumulate during a year is in (1) and shown in Fig-3.

The expected number of U.S. lightning fatalities for any date can be calculated by multiplying the number of lightning fatalities expected for that year by the percentage expected by that day of the year. In addition, various percentiles of the distribution can also be calculated, providing error bars for the predicted values. The 2.5th and 97.5th percentiles provide a 95% confidence interval, allowing hypothesis tests for whether the observed fatalities differs from the prediction at the 95% significance level. A new graph allows tracking how the U.S. lightning fatalities are accumulating during the year.

Evidence continues to accumulate that another discontinuous change in the pattern of U.S. annual lightning fatalities occurred in 2008 and continues through the present. The probability of this change occurring randomly is 0.39%. A second new model (2008-2015) for these fatalities was created (Table-III). Since the period is relatively small, only 8 years, frequent updates to this new model should be conducted as more data becomes available. The large error bars for both the 1994-2015 and 2008-2015 models will decrease significantly over the next 4-7 years as the influence of large values from the mid-1980s and early-1990s decreases and disappears. In addition, the pattern of annual U.S. lightning fatalities should continue to be monitored to ensure the new pattern is persisting.

Regardless of which model is used, the 1994-2015 or the 2006-2015 model, the U.S. annual lightning fatalities are in the tail of their negative exponential distributions. As a result, the lightning fatalities are no longer declining as fast as previously and certainly not as fast as desired. This suggests a different approach to lightning safety education be considered. If one concedes that the U.S. lightning fatalities are approaching an irreducible minimum, now that lightning safety awareness is well established, perhaps a different style of public education would be better at maintaining that awareness, rather than the previous approach used to increase lightning safety awareness. Or if a significant further reduction in the lightning fatalities is to be achieved, a new approach to lightning safety education may be needed.

## 7. 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.

Holle, R. L (2015), Investigating the increase in lightning fatalities through July 2015, Vaisala Experts Blog, www.vaisala.com/en/weather/lightning/experts-blog/Lists/Posts/AllPosts.aspx, 15 Jul 15, 1 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

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 (2016a), 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, 2016

National Weather Service (2016b), NOAA, National Weather Service Lightning Safety, 1325 East West Highway, Silver Spring, MD 20910, www.lightningsafety.noaa.gov/more.htm, 2016

Roeder, W. P., 2016: Changes in U.S. annual lightning fatalities from 1990-2015, 24th International Lightning Detection Conference/8th Lightning Meteorology Conference, 18-21 Apr 16, 7 pp.

Roeder, W. P., B. Cummins, W. Ashley, R. L. Holle, and K. Cummins (2015), Lightning fatality risk map of the contiguous United States, Natural Hazards, Journal of the International Society for the Prevention and Mitigation of Natural Hazards, Vol. 79, No. 3, doi: 10.1007/s11069-015-1920-6, Aug 15, 1681-1692

Roeder, W. P. (2015), Predicting the expected number of U.S. lightning fatalities for a year and for a date within that year, 7th Conference on Meteorological Applications of Lightning Data, 5-8 Jan 2015, 7 pp.

Roeder, W. P. (2014), An update to predicting the number of U.S. annual lightning fatalities, 5th International Lightning Meteorology Conference, 20-21 Mar 2014, 13 pp.

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 interannual 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.

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