Nat. Hazards Earth Syst. Sci. Discuss.,
doi:10.5194/nhess-2016-198-AC1, 2016

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "Spatio-temporal smoothing of lightning climatologies" *by* Thorsten Simon et al.

Thorsten Simon et al.

thorsten.simon@uibk.ac.at

Received and published: 8 August 2016

Thank you very much for your informative and detailed comments. It's obvious that we will have to work on two aspects of the paper during the revision process: firstly, We have to motivate the need for this method better and secondly we have to present the method in a more accessible way.

As for the motivation:

The main motivation to process the raw data by a statistical model is to improve the signal-to-noise ratio. The raw lightning data contains a lot of noise due to the high variability of processes generating lightning. In general this is not only true for lightning, but also for most other atmospheric variables, e.g. precipitation, wind speed and direction, etc. The GAM applied in our study aims at filtering effects/signals associated

with altitude, day of the year and space and to separate these from the noise. We will extend the introduction in the next version of the manuscript with respect to this point in order to enhance the motivation of the method.

Furthermore, we have to present the method in a more accessible way for readers unfamiliar with advanced regression methods. We are aware that the mathematics behind the methods is complex and maybe even daunting for readers with no or little statistical background. However, we are willing to work on that issue, i.e., presenting the method in a more accessible way, in order to encourage more scientists from the lightning community to work with GAMs.

–

General point 1: I have read this paper dealing with the lightning climatology in Austria. While the paper is well written, clear and at a high level of English, I am not sure why a model is needed to describe the lightning climatologies, when the raw data is already available to the authors. The authors state that for risk assessment or when climatology is used as a benchmark weather forecast this model will be valuable. But why do we need a model when we have the actual real lightning climatology. If we need to know what the probability is of lightning hitting location A, we can calculate this from the raw data.

Answer: This point is already partly addressed by the motivation given above. Regarding the estimation of the probability of lightning at location A we have to admit that this is also possible by averaging the observed lightning days over all years in the data base and maybe also to include some neighboring locations in order to smooth the estimate. However, things are getting more complex when this has to be applied to several locations simultaneously. This is especially true for regions with complex terrain, where a smoothing is desired not only over space and time but also over the altitude. Estimating climatological values by averaging easily leads to an arbitrary selection of smoothness. Thus our aim is to present a neat method that helps researchers to produce valuable climatologies from their raw lightning data.

General point 2: In addition, the model is developed using the lightning data itself, and then tries to predict the same lightning data. So the model input and output are not independent of each other. A correct model should use parameters A, B and C to predict D. not A, B and D to predict D. Furthermore, the model should be developed for a specific period, i.e. 1992 to 2000 (for example) and then tested on the year 2001 to see if the model can reproduce the lightning of 2001. In fact, it would be interesting and valuable to compare the model output (2001) with the real data (2001). How well correlated are the lightning estimates by the model for 2001 (based on a model constructed with input data from 1992-2000) with the real lightning from 2001. That is a legitimate test of the model.

Answer: This is absolutely right. We applied this procedure by cross-validation. 6 years of data are available. The parameters of the model are estimated based on 5 years of the data and validated on the remaining year. This is done 6 times with every single year serving as validation period once. All scores presented in the study are based on this procedure which is state-of-the-art in statistics. We will describe this procedure in more detail in the next version of the manuscript.

General point 3: Finally, if the model is a physical model, then it should be applicable to other regions of Austria. How well does this model predict lightning in other regions of Austria (or Europe)? If it is only good for Carinthia, then why bother? Just use the real observed data for risk maps.

Answer: It is not a physical model, but a statistical one. The presented method can still be applied to other regions.

–

Specific comments:

Page 2 Line 2: of Austria vary

Answer: Will be corrected.

Line 24: data are

Answer: Will be corrected.

line 30: period are

Answer: Will be corrected.

Page 3 line 1: What is the detection efficiency of the intra-cloud flashes relative to the CG flashes?

Answer: Because it is impossible to determine the detection efficiency of intra-cloud flashes without a locally installed VHF network (e.g. LMA), there was no attempt made in Schulz (2016) to characterize the detection efficiency of intra-cloud flashes.

line 6: what about the detection efficiency in %?

Answer: Schulz (2016) show that the flash detection efficiency is greater than 96% (100%) if one of the return strokes in a flash had a peak current greater than 2kA (10kA).

line 13: data are

Answer: Will be corrected.

line 24: Is it a correct assumption to assume a random process?

Answer: Yes, it is a general assumption made for statistical modelling.

Page 4: This reminds me of the KISS principle.....(Keep it Simple, Stupid)

Answer: Yes, we tried to keep it simple. However, I agree that the reader might feel overwhelmed. Nevertheless, eq. (3) is important as it introduces the smoothing parameter lambda which is tuned by cross-validation.

Page 5 line 27: The raw data also shows that the main lightning season is from June until end of August. So what is so great about this model? Why do we need it? To tell us something we already know from the raw data? I don't understand the logic behind the model. What can it tell us that we don't already know.

Answer: Cf. general point 1.

Page 8 line 21: But if we HAVE the climatology, why do we need this tool?

Answer: Cf. general point 1.

line 24: I do not understand why this is a useful too when the raw data give a better estimate of the climatology.

Answer: Cf. general point 1.

line 26: smooth estimates can also be obtained by averaging the raw data temporally and spatially.

Answer: That is true. This procedure would refer to a k-nearest-neighbor estimation. However, one would have to find the optimal width for the smoothing windows in time and space. The analogy between finding the smoothing parameter lambda in a GAM framework and finding the width for a smoothing window is illustrated by Hastie et al. (2009, chapter 6.2).

line 27: Why not simply use the real raw data? I do not understand why a model is needed.

Answer: Cf. general point 1.

Page 12 Figure 2 caption: cells with

Answer: Will be corrected.

Page 13 Figure 3: What are units of y-axis in upper plots?

Answer: No units. The values are on the scale of the additive predictor, i.e. the right hand side of eq. (1).

Page 14 Figure 5: How does this differ from the raw data climatology. Maybe show one next to the other.

Answer: Thanks. That's a good idea. I will consider this and add empirical estimates from the raw data for comparison.

Page 15 Figure 7: What are the numbers in the key of the figure on right. 2.1? 7?.

Answer: The legend shows expected number of flashes accumulated over the lightning season.

–

References:

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning (2nd edition), Springer-Verlag, 763 pages, 2009.

Schulz, W., Diendorfer, G., Pedeboy, S., and Poelman, D. R.: The European lightning location system EUCLID – Part 1: Performance analysis and validation, Nat. Hazards Earth Syst. Sci., 16, 595-605, doi:10.5194/nhess-16-595-2016, 2016.

–

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-198, 2016.