Abe Jacobson
30 April 2014

Initial comments on draft thesis of Michael Hutchins

Michael,
Overall, the thesis draft is very good.

Actually, its very *very* good. Actually, it's awesome. I've enjoyed reading most of it (exception is Chapter 2; this must be cleaned up.) The stuff on array bootstrap cal is fabulous. Ditto the ocean/land contrast effect.

Overall, the thesis draft is substantive and also very wide-ranging in its inclusion of several related by separate topics. More than one part of the thesis reflect significant novelty and originality, not to mention persistence. This is one of the most useful and practical theses I've seen in this shop.

My comments below are mainly about clarity and precision. I've included only the criticisms and calls for cleanup or rewording, so I might appear as a scold with nothing good to say. That is not the case; I am just saving on space by omitting the good news, which is the majority.

All of my requests below are for clarity and precision. You need to pay particular attention to clarity and precision as you embark on a career in informatics and technology. Large teaming projects absolutely require, and value, clear and accurate technical communication that "gets it right the first time". Thus, perhaps, my comments below are not purely irrelevant and pedantic.

Chapter 2 was the only consistent problem I had. I mention a number of times when you are treating stroke "power". My mentions of this topic might sound acrid, but please understand that my only goal in making those comments is to get you to clear up your treatment of this very minor part of your thesis so as to segue into the excellent material to follow without having pre-irritated your reader!

In my view it is very confusing and unenlightening how you define stroke power. It is time for you to make a clean-up of chapter 2 and to go purely to stroke *energy*, which you can estimate very clearly. As for whether you can validly compare your results to previous workers' measurements of *peak stroke power*, that is something I would need to be convinced of, and that hasn't happened.

I have retired from copy-editing others' work, but let me just mention a chronic glitch in your grammar. You often use a comma where a semi-colon is appropriate. For example, on p. 44 you write

"As shown in the previous sections the relative detection efficiency values in a given day are derived from the WWLLN observed stroke energy distribution from the previous seven days, this allows..."

The comma before “this allows” should be a semicolon. Ditto for literally scores of other places; you find ‘em. Note, this is not “Ditto for literally scores of other places, you find ‘em.” !

****

Whenever you write “triggering window”, a more accurate phrase is “recording window”. The trigger is located at one sample within the recording window. When I hear the phrase “triggering window”, I assume the writer means that the trigger can occur sometime during this window. But that is clearly not your intent, is it?

Chapter 1, p. 9: Egad, Michael, I thought I had gotten you to correct this confusing terminology. But here it pops up again, like the monster we thought we had killed in the B horror movie. You write

“The stroke energy is calculated from the time- integrated root mean square (RMS) electric field...”

Jeesh, I certainly hope you haven’t done that! Time-integrating the “root mean square electric field”? Let us hope not! I hope that you are taking the square root of the time-integrated squared electric field. This is not quibbling over words, this is a real night-and-day difference. In the language of quantum mechanics, “time-integrated” and “square-root” do not commute.

P. 10: You write

“This is due to variable WWLLN station coverage and the strong a↲ect on VLF radio propagation from orography and ionospheric conditions along the great-circle path of the wave.”

No, “orography” refers to altitude variations, and has only a minor effect on VLF propagation. Rather, don’t you mean “surface electrical conductivity”?

P. 10: You write

“VLF waveforms (1 Hz - 256 kHz)”. No, that’s VLF/LF.

P. 11: You write

“LIS is useful as it is **a lightning detection system with no overlapping detections methods with ground based networks,** and the sensor performance has not changed over time.”

Huh? What does that mean?

Chapter 2, P. 16: You write

“With this technique the global median stroke power seen by the network is 1.0 x 10^6 watts”

Gee, it sure would have been more clear to cast all of your statements like this in terms of radiated energy (Joules). The "power" quoted here is completely confusing to the readership, as the averaging is over a window duration that needs to be stated every time you specify a power level. Up to this point you have not stated the window duration, so it is meaningless to quote a "power" level when in fact you mean averaging a narrow pulse power over a much longer averaging window. Later, when results get serious, you go back to joules and I am happy. Make sure to change this to get rid of power- it is totally dependent on your integration window duration (not even stated to this point) and therefore physically meaningless.

P. 12: You write

“Our global median stroke power is found to be four orders of magnitude lower than reported earlier for the measurements including the nearby ground and sky wave.”

This is not an interesting or useful discussion as you have phrased it. Part of the discrepancy is that you are defining power in a confusing (less polite words come to mind!) way (scaled by the inverse of a mysterious integration window), which you then compare with other folks’ “power” that is based on a different definition.

P. 18: You write

“As we will show here, the network measures a median global VLF stroke power in the far-field waveguide mode of 1.0 x 10^6 W with an average uncertainty of 17%. Previous measurements have shown the power radiated by strokes is often near 10^{10} W (Krider and Guo, 1983). This difference is due to methodology in the measurements. Past measurements used a broad band peak power measurement taken in range of the ground wave (100 km), while WWLLN measures the power in the 6-18 kHz band from the RMS electric field at much longer distances. When these factors are accounted for the median power from WWLLN located strokes is comparable to the previously reported value of 10^{10} W peak power.”

Again, you are hoisted by your own petard (your confusing definition of power). You measure VLF *energy*, not power. The power as you unfortunately define it depends on the waveform details and the width of the waveform compared to the width of the integration window. The whole kerfuffle over previous workers’ peak power and your confusing power is poorly framed and very confusing. If you can use your data to retrieve peak power, then explain how, then compare it with previous workers' peak power. On the other hand, if you cannot retrieve peak power, then this little section of the thesis should be deleted. It is a tiny (but unappetizing) appetizer before a large and excellent meal.

P. 20: You write

"Since energy of the stroke is the time integrated electric field, and the power measurements are from the RMS electric field value, the energy of the stroke can be found from the size of the triggering window: $E_{stroke} = P_{stroke}$ ⇤ t_{window}, with the current triggering window set at 1.33 ms."

First, it's really confusing to use the same symbol (E) for both the electric field and the stroke energy. Second, your first clause is wrong. Third, you've at last revealed the integration width on which your definition of "power" depends. And when a future WWLLN person doubles that width, all the powers will go down by a factor of two. Is that sensible? I think not.

P. 24: You write

"measurements of return stroke absolute peak current"

No. There is no such measurement unless you have a shunt on the bottom of the current channel and actually measure it. Rather, VLF/LF location arrays (e.g. NLDN) make a rather poor WAG (...guess) based on a theory of the channel that may or may not apply. The theory has been validated ONLY on triggered lightning, which is a lo-current, copper-plasma affair that probably differs a whole lot from natural lightning. With a VLF/LF location array, all you measure is the peak voltage at the sensor; getting to a "peak current" is a huge leap of faith.

P. 26 : Twice you mention "RMS power". Leaping Lizards! Those words mean: the square root of the average of the square power. I don't think you mean that.

P. 27: You write:

"The biggest effect on the received stroke power is caused by the distances involved in the measurement and most particularly the di↵erences between the ground and sky wave near the stroke and the waveguide propagated signal (and hence received powers). Most of the VLF power measurements in the literature have measured waveforms at around 100km from the stroke. These distances are near enough that the ground and sky wave has not yet been attenuated by the structure of the Earth-ionosphere waveguide."

(a) I totally don't understand that. It corresponds to nothing I can glean from my own work in this area. And your paragraph ignores the 600-lb gorilla- your nonphysical definition of power is being compared to other workers' peak power.

P. 27: You write

“sample at a Nyquist frequency of 48 kHz”. No, you sample at 96 kSamp/s. Nyquist refers to the allowable bandwidth, not the sampling rate (which is twice the Nyquist).

P. 28: “RMS power” rears its head again. Grr*&%^

Figure 2.7: You have used 100 logarithmic bins, or 0.05 change in the logarithm per bin. So as you go rightward on the abscissa, each bin refers to a bigger chunk of stroke-power real estate. This is misleading. You should divide the histogram by the binsize (in real units, not logarithm) so as to present a probability density vs stroke power. Then I can look at a peak and say, “The probability peak occurs there!” In its present form that cannot be done. In addition, this product should deal with stroke energy, not stroke power.

######Celebration: I can put aside my pain from your treatment of the stroke power and just enjoy the rest of this thesis, which is excellent!#####

Chapter 3, P. 33: “Orography” again.

P. 34: You write

“This causes the range of electron- neutral collision frequencies to overlap with the range of sferic wave frequencies, increasing the attenuation rate of the sferics.”

You wording seems to imply that you believe there is a resonance between the collision rate and the wave frequency. What do you mean?

Figure 3.1: I am so, so happy that you’ve switched to *stroke energy*.

Hallelujah, may the Lord be Praised! Amen and Amen.

Table 3.1: You should not carry unrealistic numbers of digits, suggesting more precision than you can justify. E.g., do you really know the minimum detectable energy to four or five (Tel Aviv) significant figures? I suspect it’s at most 2 figures precise.

Figure 3.5: This is very compelling and effective.

Figure 3.8 et seq: Several of your plots use alphabetic month designators on a time axis.

Do the tick marks correspond to the start of the listed month, to the middle, or what?

Figure 3.9: Where is the “green line”?

Figure 3.11: Zone-local or solar-local time?

Chapter 4, p. 56: You write

“LIS is used as it is a lightning detection system with no overlapping detections methods with the ground network”

I don’t know what that means.

Figure 4.1. You compare LIS and ENTLN stroke densities. It looks like you have not rescaled LIS according to the inverse of its viewing time. ENTLN stares at every place all day, without interruption. LIS, by contrast, has a very low time-on-station within view of a storm. You need to rescale LIS densities upward, to what they would have been had TRMM hovered stationary in view of the storm all the time.

P. 57: You write

“Days with less than 30 LIS flashes are not used in this evaluation.”

Do you mean thirty LISflashes within a 0.5X0.5 deg area per day?

Pp. 57-8: You write

“The viewpoint granules give the start and end times of LIS observation for 0.5 ⇥ 0.5 bins. Full coverage for a given viewpoint granule is determined by the start and end times of the viewpoint granules that are the adjacent corner granules. When LIS has at least partial coverage of the four corner viewpoint granules the center viewpoint granule will be in full view of the sensor.”

Very confusing.

Chapter 6, p. 76: You write

“The electric field measurements are normalized to the radiated very low frequency (VLF) stroke energy to allow direct comparisons of the many stroke-station paths seen by WWLLN.”

Aren’t they normalized by the square root of the stroke energy?

Figure 6.2: Ditto here in the caption, aren’t they normalized by the square root of the stroke energy?

Figure 6.3 caption: “Compiled station...” Do you mean “Combined station”?

Chapter 7, p. 87: You refer twice to the Maxwell current. In a thesis, initial use of an esoteric technical term should be explained. I will come out of the closet, and admit a perverted detail of my life: I don’t know what is the physical mechanism referred to by “Maxwell current”.

Chapter 8, Figure 8.1: Just a thought. You had earlier stated that WWLLN detects essentially all thunderstorms, as assertion that I may have made with not much care a few years ago. Is it right? I’m no longer sure. Certainly, if we consider ENTLN as groundtruth over the Lower 48 land area, then WWLLN had a storm detection efficiency as bad as 10-20 %, no??? And what does that imply for your global storm count (600 at any instant) and your storm-# multiplier on Mach et al current per storm? (I believe you assumed WWLLN sees all storms.)

Figure 8.6: In figures such as this, you need to clarify what the bin size is, and if it is constant. Also, is the bin constant in time or in log(time). This makes a *huge difference* in the reader’s understanding of the spikes at the left for WWLLN.

Speaking of that, why are the spurious features spiky? The ordinate value means nothing until the reader is told what the binning scheme is. I suspect that the spike is an artifact of the logarithmic bins being resolvable by eye. You really need to go to a probability density (vs time, not log) time. You are then free to display that probability density on a log abscissa, but the “per ....” in the density has to be per unit time. Also, graph it bar-chart style to avoid creating misleading spikes.