Title: West Texas Lightning Mapping Array (WTLMA) Data 17 January 2013

Author(s)

PI: Dr. Eric Bruning
Assistant Professor, Atmospheric Science Group
Department of Geosciences
Texas Tech University, Box 41053, Lubbock, TX 79409-1053 USA
Ofc: 806 834-3120 / Fax: (806)-742-1738
eric.bruning@ttu.edu

1.0 Data Set Overview:

This dataset is comprised of total lightning measurements at VHF from the West Texas Lightning Mapping Array. Included are:

1. VHF source locations produced from raw station data using the standard New Mexico Tech algorithm
2. Lightning flash output from a flash clustering algorithm that combines VHF source information into flashes using space and time criteria
3. Gridded flash products using lmatools.

Time period covered by the data: May 14, 2012-June 30, 2012
Center of LMA: 33.606968 -101.822625
Data source: Texas Tech University
LMA website: http://pogo.tosm.ttu.edu/wtlma/

2.0 Instrument Description:

The West Texas Lightning Mapping Array became routinely operational on 25 April 2012, using the latest LMA hardware available from New Mexico Tech, the latest in series of improvements to the network (Rison et al. 1999, Krehbiel et al. 2000, Thomas et al. 2001). During DC3, 10 stations were active. Detailed error analysis of a similar network operated during the STEPS-2000 campaign is provided in Thomas et al. (2004). Koshak et al. (2004) provides additional error analysis. The relatively uniform distribution of stations in the West Texas network should have less source error variability with azimuth than depicted with the somewhat latitudinally elongated North Alabama network. Altitude retrievals are the most error-prone, and are negligibly minimal within at least 50 km of the network center, with a decline in altitude precision to the range of the New Mexico border from Lubbock. Beyond this range (125 km), altitude errors are large enough that the data should be considered effectively 2D, with additional errors in range becoming increasingly prominent out to the line-of-sight horizon at approximately 250 km.

The West Texas network overlaps in coverage with the Central and Southwest Oklahoma networks operated by the National Severe Storms Laboratory. Future processing should enable co-processing of these networks to permit additional regions of 3D coverage, especially between southwest Oklahoma and West Texas.

3.0 Data Collection and Processing:

Data were collected through a combination of automated and manual collection to ensure data from as many stations as possible were incorporated. Each LMA sensor recorded the peak pulse of all all pulses to break threshold in an 80 μs interval. VHF source locations from at least 5 sensors are used to obtain the VHF source location using a time of arrival technique. A two-pass processing technique was used, with the first pass accepting all retrieved sources with chi-squared less than 5.0, and a second pass that permitted additional points with large chi-squared values.

VHF source locations are then clustered into flashes using temporal and spatial criteria according to McCaul et al. (2005, 2009). All sources with chi-squared less than 5.0, altitude less than 20 km, and detected by at least six stations were candidates for clustering. A 3 km spatial and 0.15 second temporal threshold was used for clustering, and no flash was allowed to exceed 3 s in length. Although no minimum source number per flash was prescribed in the flash data output, we recommend defining as valid flashes as containing ≥ 10 sources per flash to mitigate the presence of noise. The more experienced user may wish to use a different source number per flash threshold to define a valid flash, which is why all potential flashes are provided regardless of source number.

The gridded flash products were created using the lmatools Python software package (Bruning 2013). A minimum of ten sources was required for a flash to appear on the grid. Four product types were created: VHF source density, flash initiation density, flash extent density, and average flash area. These products were created in 5-min windows on a regular lat-lon grid with nominal (at the grid center) spacing of 3 km. For each of these products, except for perhaps flash initiation density, a logarithmic color mapping usually works best. A brief description of these product types follows.

Source density
This product is a simple count of the VHF sources on the grid. It is the most basic form of gridded data that can be derived from the LMA data. It only counts those sources that contributed to a valid flash.

Flash initiation density
For each flash, the first VHF source point in time is taken as the flash initiation location. Therefore, each flash is represented by a single grid cell location A sum of all values on this grid gives the total number of flashes in the domain. The electric field rebuilds relatively rapidly in regions with larger flash initiation rates, since flashes initiate in regions of large electric field.

Flash extent density
Flash extent density gives a column-local flash rate; it is a count of how many flashes passed through that grid cell. This product highlights a key advantage of the LMA: it

shows regions where it was energetically favorable (local maximum in electric potential) for flashes to frequently revisit.

Average flash area (footprint)
The average flash area product calculates the average area of all flashes that passed through a grid cell. It is the sum of the areas of all flashes that passed through a grid cell divided by the flash extent density. This product is good at highlighting regions with infrequent but very extensive flashes, such as MCS stratiform regions and supercell anvils. It can also highlight convective regions that have unusually small, frequent flashes.

4.0 Data Format:

4.0.1 VHF Sources

Data file structure is ASCII, with naming convention for the files being
LYLOUT_yymmdd_hhmmss_duration (seconds).dat

<u>Data Format:</u> Header information is 65 lines long and is in each ten-minute file. Header information includes the following:
LMA center location
Station locations
number of active stations
chi-squared used in reprocessing
minimum number of stations used in processing

<u>List of parameters with units, sampling intervals, frequency, range:</u>
Each line of the file after the "*** data ***" line represents a single geolocated VHF source retrieval.

<u>Sampling rate:</u> Effectively continuous, with a maximum rate of 12500 locations per second, derived from the 80 µs trigger windows at each station. See instrument description section for references that discuss timing precision, which is on the order of 40-50 nanoseconds for well-located sources.
<u>Range:</u> < 50 m 3D error within 125 km of the center of the lightning mapping array. Thus data most accurate within 125 km of the LMA center. See instrument description above.

<u>Columns</u>
time (UTC sec of day), latitude (decimal degrees), longitude (decimal degrees), altitude (m), reduced chi^2 (unit less), Power (dBW), mask (unit less)

format(15.9f 10.6f 11.6f 7.1f 5.2f 5.1f 6x)

When the hexadecimal mask value is converted to an integer and represented as binary, each bit position represents a station, given by the station mask order in the header. Stations whose bit positions are "on" (1) participated in the solution.

4.0.2 WTLMA Flashes

ASCII source data that have been clustered into flashes are provided in HDF5 format. The naming convention of these ten-minute (UTC) files is:
LYLOUT_yymmdd_hhmmss_duration.dat.flash.h5,
where duration is in seconds.

Each file includes individual source data in an "event" group, with an individual table within the event group for each 10-minute clustered file. The columns of an example event table for 5 May 2012 for 0000-0010 UTC are as follows:

```
/events (Group) 'Analyzed detected events'
/events/LMA_120520_000000_600 (Table(9,)) 'LMA_120520_000000_600'
  description := {
  "alt": Float32Col(shape=(), dflt=0.0, pos=0),
  "charge": Int8Col(shape=(), dflt=0, pos=1),
  "chi2": Float32Col(shape=(), dflt=0.0, pos=2),
  "flash_id": Int32Col(shape=(), dflt=0, pos=3),
  "lat": Float32Col(shape=(), dflt=0.0, pos=4),
  "lon": Float32Col(shape=(), dflt=0.0, pos=5),
  "mask": StringCol(itemsize=4, shape=(), dflt='', pos=6),
  "power": Float32Col(shape=(), dflt=0.0, pos=7),
  "stations": UInt8Col(shape=(), dflt=0, pos=8),
  "time": Float64Col(shape=(), dflt=0.0, pos=9)}
```

A flash_id column denotes the flash_id within this 10 minute interval to which the source belongs. The charge column is empty; it is meant for use in performing manual charge identification on sources using a GUI tool in post-analysis. Stations contains the count of stations that participated in the solution. Other columns have the same meaning as in the ASCII source data files.

Each HDF5 file stores flash-level data in a "flashes" group, with an individual table of the same name as the event subtable to which it corresponds. The columns of an example flash table for 5 May 2012 for 0000-0010 UTC are as follows:

```
/flashes (Group) 'Sorted LMA flash data'
/flashes/LMA_120520_000000_600 (Table(9,)) 'LMA_120520_000000_600'
  description := {
  "area": Float32Col(shape=(), dflt=0.0, pos=0),
  "ctr_alt": Float32Col(shape=(), dflt=0.0, pos=1),
  "ctr_lat": Float32Col(shape=(), dflt=0.0, pos=2),
  "ctr_lon": Float32Col(shape=(), dflt=0.0, pos=3),
```

```
"duration": Float32Col(shape=(), dflt=0.0, pos=4),
"flash_id": Int32Col(shape=(), dflt=0, pos=5),
"init_alt": Float32Col(shape=(), dflt=0.0, pos=6),
"init_lat": Float32Col(shape=(), dflt=0.0, pos=7),
"init_lon": Float32Col(shape=(), dflt=0.0, pos=8),
"init_pts": StringCol(itemsize=256, shape=(), dflt='', pos=9),
"n_points": Int16Col(shape=(), dflt=0, pos=10),
"start": Float64Col(shape=(), dflt=0.0, pos=11)}
```

Each row in the table should have a unique flash_id that corresponds to the flash_id assigned that flash's n_points VHF sources in the event table. The first source in the flash is assumed to be the flash initiation time and location, recorded in the start, init_lat, init_lon, and init_alt variables. The ctr_lat, ctr_lon, and ctr_alt variables report the mean latitude, longitude, and altitude, respectively, of all sources in the flash. The duration of the flash is reported in the variable of that name. The flash area (in km^2) is calculated as the convex hull of the VHF sources in plan projection.

For details about the flash sorting methodology, see section 3.0. An intermediate ASCII-format flash file resulting from the McCaul et al. (2009) flash-sorting algorithm is available upon request.

4.0.3 Gridded Flash Products

Gridded flash products are delivered in Climate and Forecast Metadata Compliant (Eaton et al., 2011) NETCDF v.3 files. The naming convention is:
WTLMA_yyyymmdd_hhmmss_3600_10src_0.0328deg-dx_product-type.nc,
where WTLMA indicates the network name, followed by the beginning date and time of the file. 3600 is the duration in seconds of the data in the file, 10src is the minimum point per flash and 0.0328deg-dx is the grid size (in longitude = dx; equivalent to 3 km at the network center location). product-type is one of [flash_extent, flash_initiation, footprint, lma_source]. More information about these product types are available in Section 3.0 above, and in the table below.

<u>Data header:</u>
All gridded LMA data has the following dimensions: longitude and latitude are 134 with 12 5-min time steps "ntimes". The variables are dependent to each flash product file and are as follows:

LMA Source Density	LMA Flash Footprint	Flash Initiation	Flash Extent
int lma_source(ntimes, lon,lat)	float flash_footprint(ntimes, lon, lat)	int flash_initiation(ntimes, lon, lat)	int flash_extent(ntimes, lon, lat)
units = "Count per grid pixel per 5.0 min"	units = "km^2 per flash"	units = "Count per grid pixel per 5.0 min"	units = "Count per grid pixel per 5.0 min"
long_name = "LMA source density"	long_name = "LMA local mean flash area"	long_name = "LMA flash initiation density"	long_name = "LMA flash extent density"

grid_mapping = "crs"	grid_mapping = "crs"	grid_mapping = "crs"	grid_mapping = "crs"
missing_value = -9999	missing_value = -9999	missing_value = -9999	missing_value = -9999

The variables common to all gridded flash products are as follows:

int crs	float time(ntimes)	float longitude(lon)	float latitude(lat)
semi_major_axis = 6378137.f	units = "seconds since 2012-05-03 00:00:00"	units = "degrees_east"	units = "degrees_north"
longitude_of_prime_meridian = 0.f	long_name = "time"	long_name = "longitude"	long_name = "latitude"
grid_mapping_name = "latitude_longitude"		standard_name = "longitude"	standard_name = “latitude"
inverse_flattening = 298.2572f			

5.0 Data Remarks:

Missing data periods: None. WTLMA stations that were not active for each ten-minute period are denoted in the header section of the VHF source files. If the “source” column in the Station Data table is 0, then the station was down.

On days with blowing dust in West Texas, the two-pass processing often retrieved hundreds of thousands of sources, including a great many with large chi-sq values. These sources were filtered out during the flash sorting process, since they are random in space and time and therefore did not cluster like lightning flashes. Users of the raw ASCII data should be aware that 31 May, 20 June, 28 June, and 29 June were particularly extreme examples, at times producing a in excess of two million sources in a few 10 minute files.

Software compatibility (i.e., list of existing software to view/manipulate the data)

The lmatools package (Bruning 2013) contains example code for creating plots from the gridded data files. Sample 12-panel plots (every 5 min) for each product have been created for each hour, providing an at-a-glance visualization of the data provided in the gridded NetCDF data. These plots can be downloaded from http://pogo.tosm.ttu.edu/data/DC3/plots_WTLMA/

XLMA software package. Contact Dr. Ron Thomas, New Mexico Tech, to obtain the software. Visualizes VHF ASCII data files
Email: thomas@nmt.edu

CF-compliant NetCDF gridded fields should be easily read in a number of standard packages, including various community tools from Unidata such as the IDV (http://www.unidata.ucar.edu/software/idv/).

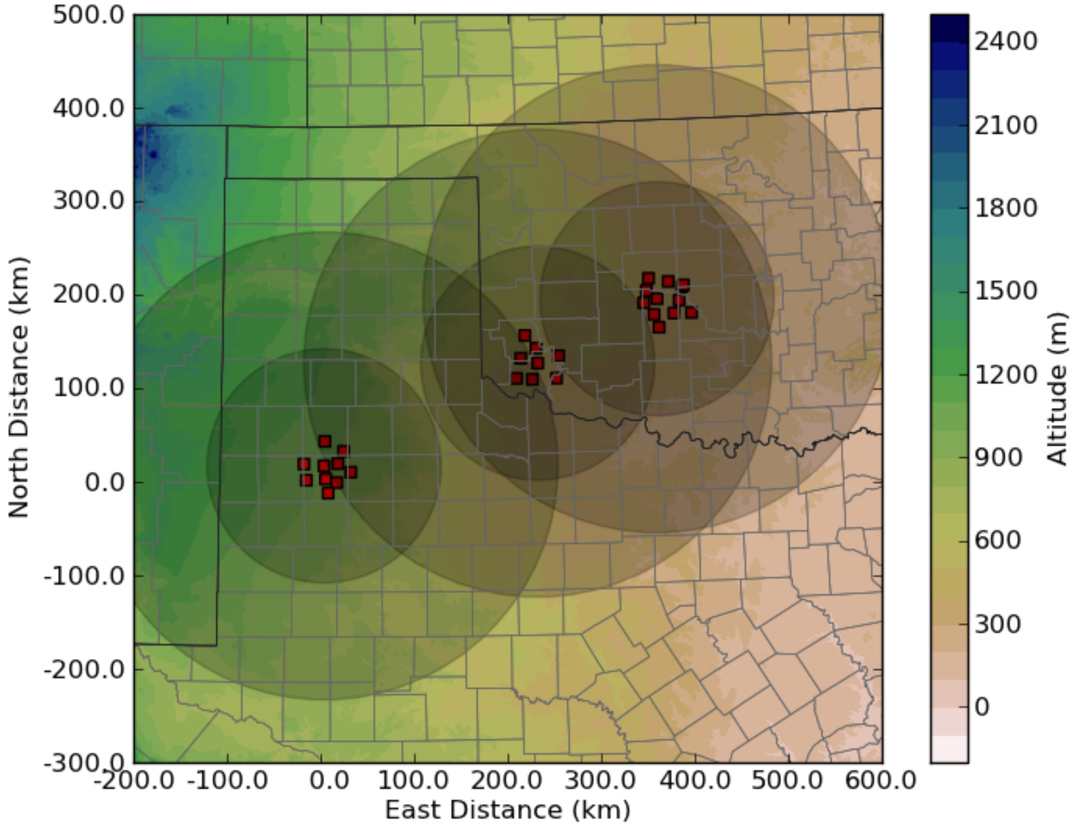

Figure 1 The combined coverage of the West Texas and Oklahoma Lightning Mapping Array installations. Sensor locations are indicated by red squares. Range rings indicate the nominal 3D and 2D coverage areas at 125 km and 250 km range from each network center. The coordinate center is Lubbock, TX.

6.0 References:

Bruning, E. 2013: lmatools source repository, https://bitbucket.org/deeplycloudy/lmatools/commits/3af14db9be08dabecccfcbe8b73f507561b0fe87, retrieved 16 January 2013.

Eaton, B. et al. 2011: *NetCDF Climate and Forecast (CF) Metadata Conventions,* retrieved 17 January 2013 from http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.6/cf-conventions-multi.html

Koshak, W. J., et al., 2004: North Alabama Lightning Mapping Array (LMA): VHF source retrieval algorithm and error analyses. *J. Atmos. Oceanic Tech.*, **21**, 543–558.

Krehbiel, P. R., R. J. Thomas, W. Rison, T. Hamlin, J. Harlin, and M. Davis, 2000: GPS-based mapping system reveals lightning inside storms. *EOS, Trans. Amer. Geophys. Union*, **81 (3)**, 21–25.

McCaul, E. W., J. Bailey, J. Hall, S. J. Goodman, R. Blakeslee, and D. E. Buechler, 2005: A flash clustering algorithm for North Alabama Lightning Mapping Array data. Preprints, Conf. on Meteorological Applications of Lightning Data, San Diego, CA, Amer. Meteor. Soc., 5.2. [Available online at http://ams.confex.com/ams/Annual2005/techprogram/paper_84373.htm.]

——, S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. Wea. Forecasting, 24, 709–729.

Rison, W., R. J. Thomas, P. R. Krehbiel, T. Hamlin, and J. Harlin, 1999: A GPS-based three-dimensional lightning mapping system: Initial observations in central New Mexico. *Geophys. Res. Lett.*, **26**, 3573–3576.

Thomas, R. J., P. R. Krehbiel, W. Rison, T. Hamlin, J. Harlin, and D. Shown, 2001: Observations of VHF source powers radiated by lightning. *Geophys. Res. Lett.*, **28 (1)**, 143–146.

Thomas, R. J., P. R. Krehbiel, W. Rison, S. J. Hunyady, W. P. Winn, T. Hamlin, and J. Harlin, 2004: Accuracy of the lightning mapping array. *J. Geophys. Res.*, **109 (D14207)**.