Title: RELAMPAGO - Low Frequency Autonomous Magnetic Field Sensors (LFAMS) Data Documentation

Version 2.0, 3/18/2021

Authors:
PI: Wiebke Deierling
Co-PI: Robert Marshall
André Antunes de Sá
Austin Sousa

Contact Information:
Robert Marshall
3775 Discovery Drive
CCAR, 429 UCB
Boulder, CO 80303
email: robert.marshall@colorado.edu

Wiebke Deierling
3775 Discovery Drive
CCAR, 429 UCB
Boulder, CO 80303
email:wiebke.deierling@colorado.edu

1 Introduction

This document provides a description of the data products associated with the Low Frequency Autonomous Magnetic Field Sensors (LFAMS) deployed from 1 November to 15 December 2018 during the RELAMPAGO campaign in Argentina. The document also provides a brief description of the instrument, its deployment, and the data processing used to generate the data products.

2 Instrument Description

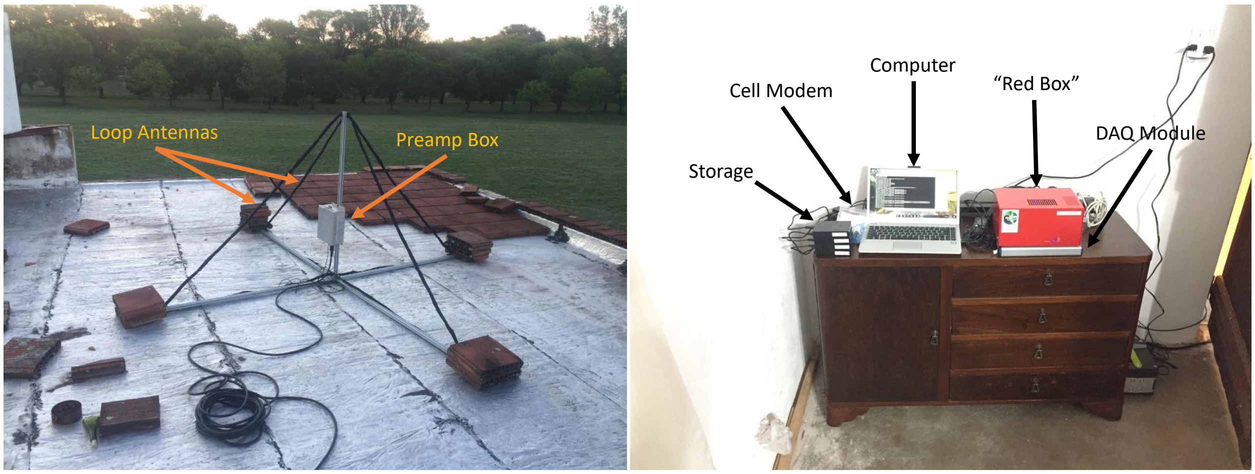

Figure 1: LF Antenna and pre-amp installed at the RELAMPAGO LF3 site (left), and receiver electronics installed inside a house at the RELAMPAGO LF1 site (right).

An array of four Low Frequency Autonomous Magnetic Field Sensors (LFAMS) were deployed during the RELAMPAGO field campaign (Fig. 1). The sensors are based on the 100 kHz sampling rate VLF instrument described by Cohen et al. [2010], modified to extend the bandwidth and increase the sampling rate to 1 MHz as described by Cohen et al. [2018]. The new LF system covers frequencies from ∼300 Hz to 400 kHz, above which the response falls off rapidly due to the anti-aliasing filter.

An LFAMS sensor is comprised of two air-core magnetic loop antennas, across which the magnetic component of a lightning-emitted sferic, $\vec{B}$, induces a voltage, V, as given by Faraday's law:

$$V = -n\frac{\partial}{\partial t}\int \vec{B} \cdot d\vec{A} \quad (1)$$

The two antennas are aligned North-South and East-West (Channel 1 and Channel 2 respectively). The voltage signal is pre-conditioned by a preamp circuit for transmission through a long cable to a receiver box (the "Red Box"). At the receiver, the signal is filtered and amplified before being sent to a National Instruments data acquisition system (DAQ), where the signal is sampled at MS/s. The DAQ sample clock is synchronized to a GPS module's pulse-per-second (PPS) signal. A personal computer interfaces with the DAQ and collects the sampled data in hard disk storage. A cell modem attached to the computer is used for remote maintenance. The system assumes that the physical antenna gain is consistent with the one derived from measured parameters and Faraday's Law, and a comb filter auto-calibration routine gives the system gain from the antenna terminal to the DAQ. Examples of channel frequency response after calibration and channel gain ratio are shown in Figs. 2 and 3.

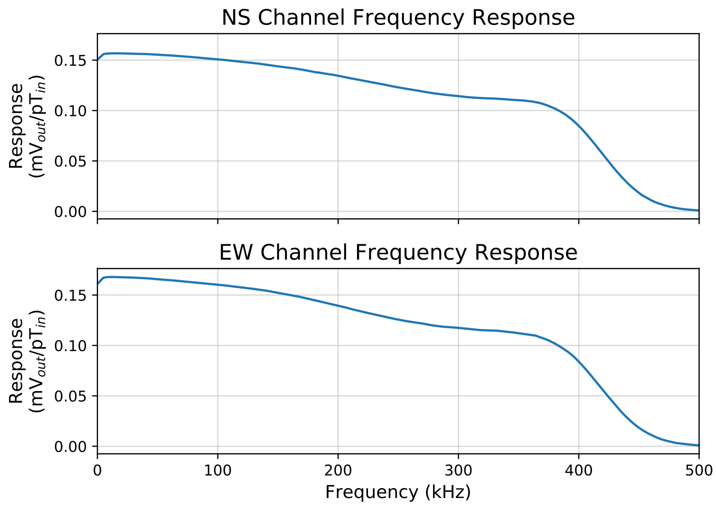

Figure 2: Frequency response for the NS/EW channel of RELAMPAGO LF4 after calibration.

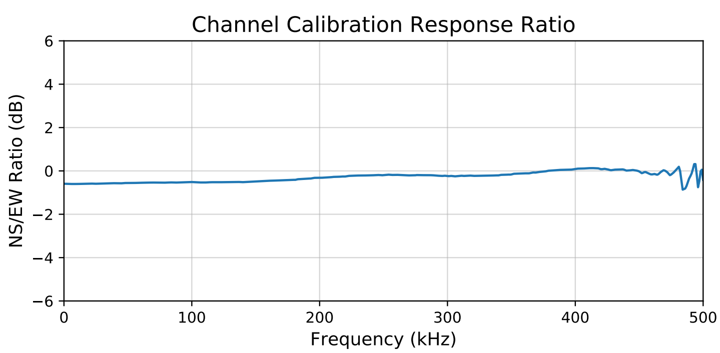

Figure 3: NS-EW channel gain ratio of RELAMPAGO LF4 after calibration.

The LF instruments deployed in RELAMPAGO suffered from a clock offset from GPS time that would be updated to a new offset with every instrument reset. Two of the instruments also had a constant clock drift of 1 $\mu s\,s^{-1}$ due to an aging oscillator that slowed enough to lose one sampling pulse per million. An example of these clock errors are shown in Fig. 4 for November 10. The clock offset can be estimated from a cross-correlation of the Level 1 data with another lightning dataset, e.g., lightning pulse data from the Earth Networks Total Lightning Network (ENTLN). Both types of errors are corrected in the first step of the Level 2 data processing, while Level 0 and Level 1 datasets are not time-corrected.

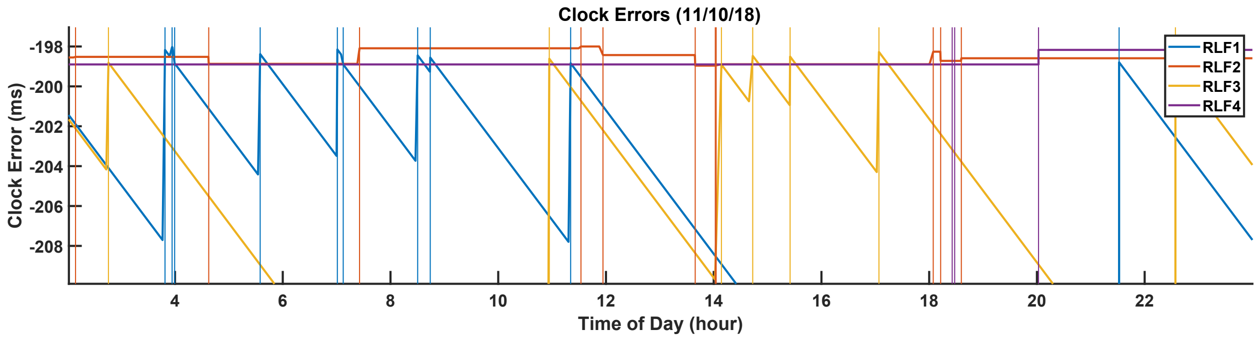

Figure 4: Clock error for the RELAMPAGO stations based on a cross-correlation with ENTLN data. Vertical lines indicate a station power-on from a shutdown or reset.

3 Network

The array of LFAMS was deployed to cover the Mendoza and Cordoba regions during RELAMPAGO (Fig. 5). Table 1 provides the location information on each of these sites.

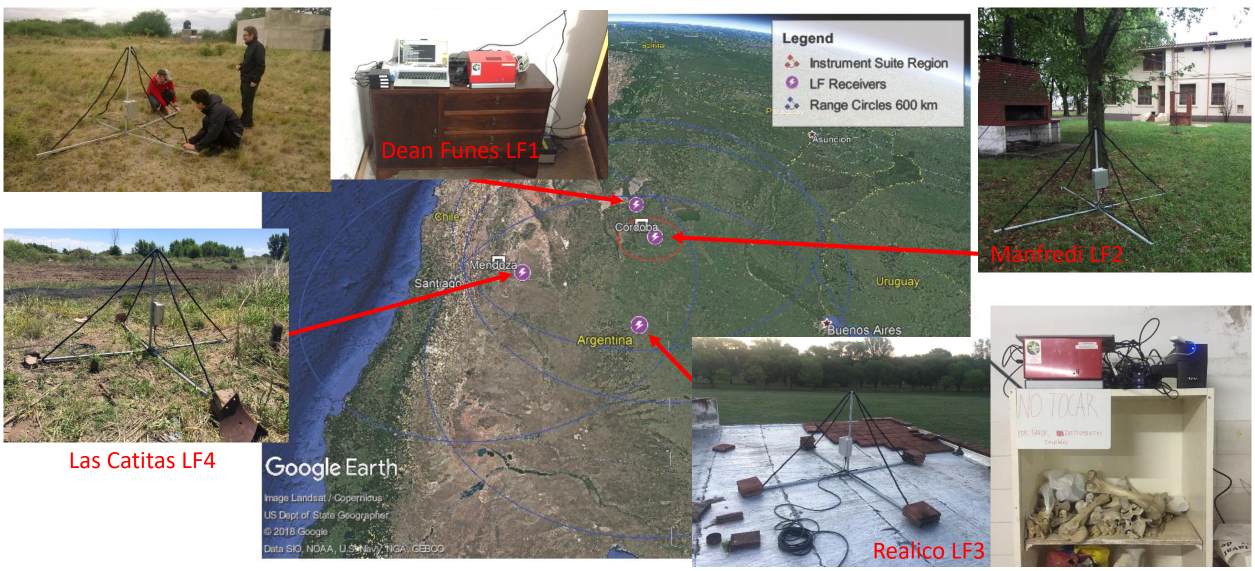

Figure 5: RELAMPAGO site map of the LFAM sensors in Argentina.

Table 1: Table providing the deployed location of the instruments.

Receiver Name	City	Latitude (°)	Longitude (°)
LF1	Dean Funes	-30.426280	-64.369355
LF2	Manfredi	-31.854739	-63.754468
LF3	Realico	-35.163211	-64.280471
LF4	Mendoza	-33.280325	-68.054266

4 Data Coverage

Once deployed, all instruments were programmed to acquire data continuously throughout the campaign, however due to power outages the receivers have periods where no data was captured. These power outages were quite frequent in Argentina during times of thunderstorms. Fig. 6 below presents the data coverage available from LFAMS in RELAMPAGO.

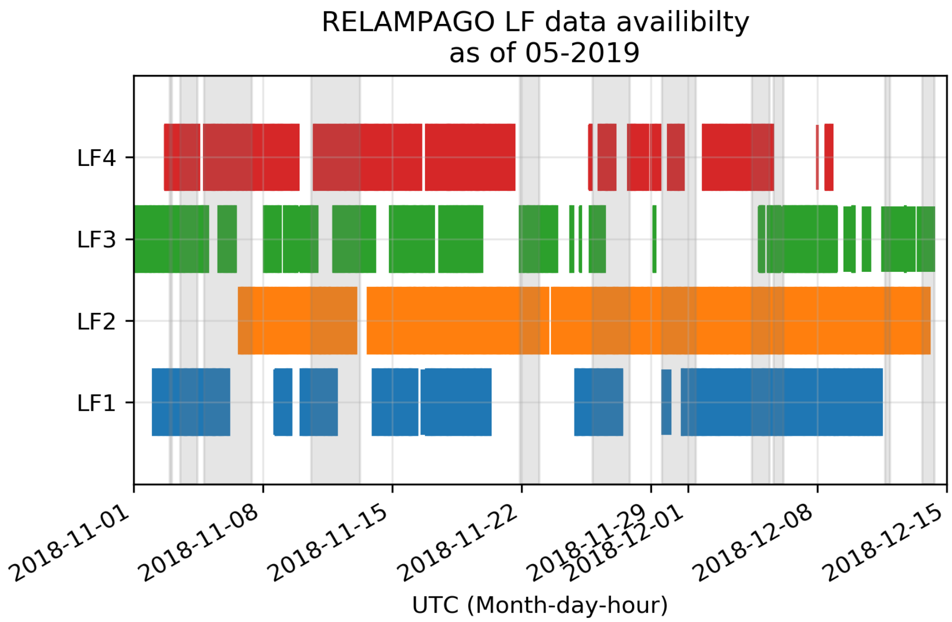

Figure 6: Data coverage during RELAMPAGO for each of the LF sensors. Shaded gray regions represent an Intensive Observing Period (IOP) for the campaign

5 Data Processing

The magnetic field data collected by the LFAMs is post-processed to yield different data products. Level 0 data correspond to raw data from the instrument, i.e., continuous magnetic field measurements, after the calibration has been applied and noise from power line harmonics are filtered out. Level 1 data are a collection of lightning event data (radio atmospherics or sferics) extracted from the Level 0 data. The Level 2 data product presents lightning event location from sferic observations and lightning flash information from clustered events.

5.1 Level 1: Sferic detection is done through a peak search across the raw data (quadrature addition of the two channels) above the noise floor. Once a possible sferic has been identified, a data window of 1.2 ms is extracted with the main peak centered at 200 µs. Fig. 7 illustrates this process by overlaying the extracted sferic windows as shaded yellow bands behind 10 seconds of raw data in purple.

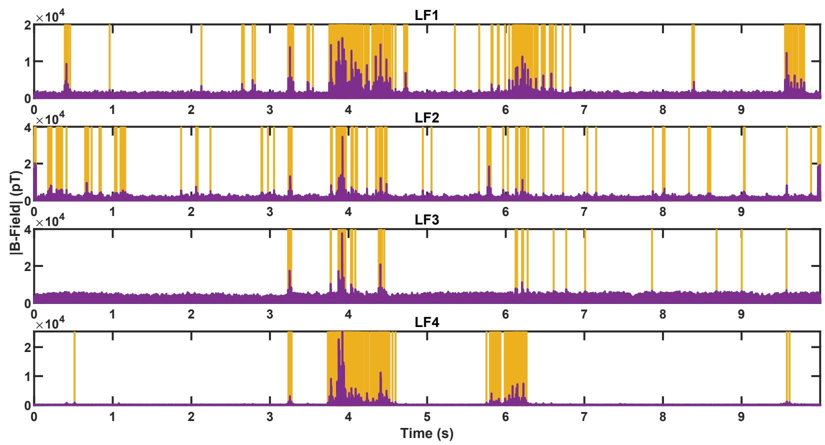

Figure 7: Plot of Level 0 data (purple) and sferic windows to be extracted from the data (yellow bands) for all receivers during 10 seconds on 11/26/2018 04:05 UTC)

5.2 Level 2: The Level 2 data processing is described in Fig. 8, where the final product consists of lightning events (return strokes) and flashes, with their estimated time, latitude, longitude, and peak current. A more detailed analysis of this processing will be presented in a future publication.

Sferic Matching With the time-corrected Level 1 data, the sferics that were observed by each receiver can be matched into lightning events, i.e., a source event corresponding to the radio emission observed as sferics by each LF station. The matching is accomplished by grouping sferics within 0.1 s of each other across stations, and then cross-correlating the whole group to remove a group time of arrival. Sferics that line up in time within 2 ms across at least three stations after the group time of arrival correction are assumed to come from the same source, becoming a lightning event output of this stage. Cross correlation scores are also computed for sferics of an event against a reference sferic, which is chosen to be first sferic observed in that event. The event quality, a measure of data quality reported

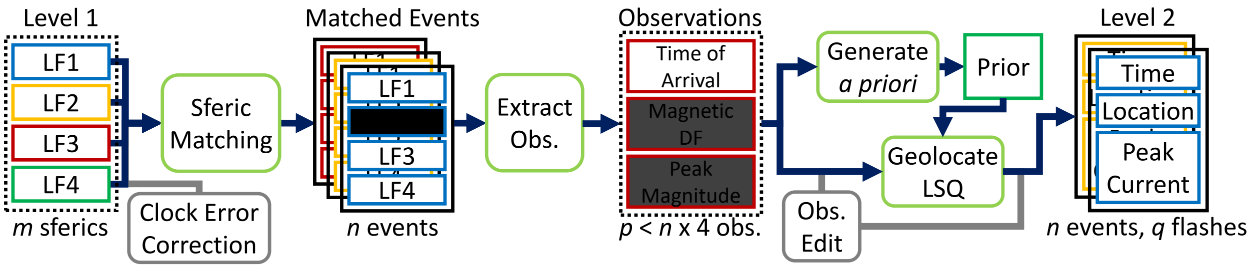

Figure 8: Flowchart describing the data processing for generating the Level 2 data product. The gray ad-hoc processes are only necessary in handling specific issues with the RELAMPAGO dataset.

for each event based on the expected similarity between sferics, is computed from the minimum cross-correlation score of the sferics of an event.

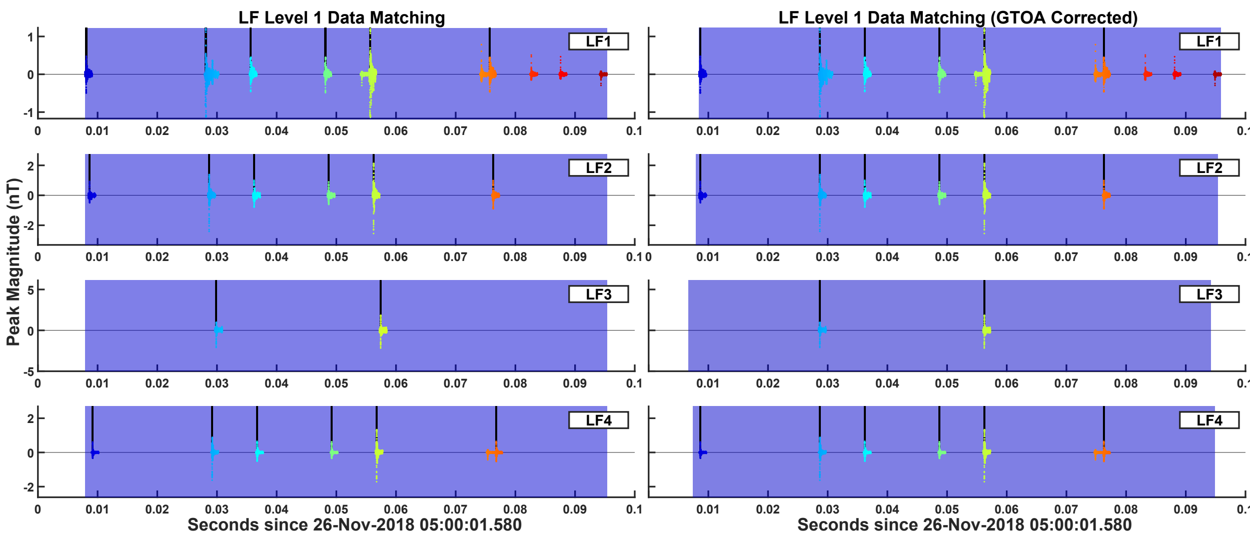

Figure 9: Plot of a group of sferics used in the lightning event matching process. Black vertical bars indicate sferics that were matched. The left plot has the sferics real arrival time at each station, and the right plot shows the sferics with the group time of arrival removed.

Sferic Observations Observations needed for geolocation, e.g., time of arrival (TOA) and magnetic direction finding (MDF), can be extracted for each event at this point. Data editing can be performed here to remove bad quality observations, for example MDF observations that came from saturated sferics. Only TOA observations were used in geolocating the RELAMPAGO dataset, given that saturation of multiple LF receivers in the RELAMPAGO storms make amplitude-based observations unreliable, but peak currents can still be estimated using amplitude information from observations that did not saturate. If all stations saturate, the peak current estimate is given as 'Inf'. Due to nonlinear saturation regions not captured in this model, lightning observations near receiver saturation, in particular lightning near the highest saturation station LF4, may be underestimated.

A priori Based on the TOA observations (sferic peak) for each lightning event, an *a priori* geolocation state is estimated before being given to a least squares filter along with the observations. Observability studies were performed for the given estimation problem, as Cramer-Rao Bound maps of the minimum location uncertainty for lightning in the network's domain, as well as simulations of the *a priori* result error. The *a priori* simulation consisted of lightning events generated all over the network domain outlined in Fig. 10. Their corresponding TOA observations was then given as an input to the *a priori* algorithm. The resulting *a priori* locations and the max location error for that location is shown in Fig. 10 for all station topologies. Based on this *a priori* error a domain mask is generated for each network configuration to include only the regions with error less than 100 km, defining the observable region for the array. The masks are then used in limiting the number of ambiguous solutions geolocated by discarding events that fall outside the mask boundaries.

Event Geolocation Finally, the observations and *a priori* are fed into a linear least-squares statistical (LSQ) filter for an estimate of lightning occurrence time and location along with their uncertainties given by the time of arrival model, assuming an unbiased gaussian distribution of time of arrival uncertainty of 10 µs and negligible model and linearization errors. The time of arrival uncertainty is a best guess based on the station clock correction for the RELAMPAGO LF dataset. For estimating peak current, an attenuation model based on finite-difference time-domain (FDTD) modeling of lightning propagation [Marshall, 2012] is used with the assumption that a known peak radiated field a distance away from the source, e.g., 100 km, is proportional to the source's peak current by a constant parameter [Orville, 1991].

Flash Clustering After geolocation and application of the observability mask, lightning events are clustered into lightning flashes. Inspired by similar LLSs, a simple agglomerative (bottom-up) hierarchical clustering is employed based on a spatiotemporal distance criteria of 10 km and 0.3 seconds. The location uncertainty of an event loosens the distance criteria directly, and, unlike other LLSs, the distance criteria is tightened according to the quality of an event. This quality-based penalty is necessary in the RELAMPAGO dataset since no quality control is performed on the Level 1 data, and a large number of detected events that are not lightning return strokes get improperly geolocated. The penalty introduces a maximum location error of 100 km for zero quality events and quickly decreases to 2 km at 0.2 quality score before tapering to 0 km at 0.6. The largest number of unsuitable events was empirically found to occur below 0.2–0.4 in the RELAMPAGO dataset, and the best penalty distribution was chosen for minimizing false detections while keeping a high flash detection efficiency. The flash inherits time, latitude and longitude of the events by means of a weighted centroid, using event qualities for weights, while the reported flash quality and peak current are from the events with strongest quality and peak current respectively. Flash area is computed from the convex hull of strokes within a flash, and flash duration is the length of time between the first and last strokes in a flash.

6 Data Format

Level 1 and Level 2 data are being made available on the EOL RELAMPAGO data webpage.

6.1 Level 1:

The Level 1 data product is packaged in NetCDF 4 "Classic Mode" containers, with each container corresponding to an hour of detected sferics for each LFAMS station. Each hourly NetCDF file follows the name convention:

RLFs-YYYYMMDDHH.nc

where *s* stands for the LF station number (1, 2, 3, or 4); and *YYYY*, *MM*, *DD* and *HH* corresponds to the year, month, day and hour, respectively, of the data covered in the file. Each file contains five variables as described in Table 2: The *product_datenumber* variable holds an integer number of days since January 0 0000 UTC, in the proleptic ISO calendar, indicating the date of the file dataset; *sferic_ew* and *sferic_ns* are the sferic data held inside a 2D array, where sferic windows of 1.2 ms (1200 points at 1 MHz) are stacked in each row of the array; *sferic_peak_time* is a 1D array representing the time in UTC seconds for the main lightning sferic peak since the beginning of the day for each sferic. As each sferic is centered around the main peak, this time always corresponds to the 200th point in a sferic window. Finally, *sferic_loc* holds the latitude and longitude information for the recording station. Each variable holds attributes relevant to that variable, such as units or a long name description. There are also global attributes available for information regarding the whole data container, such as filename, hardware version, etc.

Note: The Level 1 data released originally was negatively impacted by incorrect scaling of one of the channels and low detection efficiency for that channel. These issues have been addressed and a new version of the Level 1 data is available on the EOL RELAMPAGO data archive.

Table 2: Table of variables and number of attributes available in each NetCDF file for an hour of data coverage. Here *n* represents the number of sferics detected by the station during that hour.

Variables	Dimensions	Number of Attributes
global	N/A	9
product_datenumber	scalar	2
sferic_ew	2D (*n* x 1200)	3
sferic_ns	2D (*n* x 1200)	3
sferic_peak_time	1D (*n* x 1)	2
sferic_loc	1D (1 x 2)	2

6.2 Level 2:

The level 2 data product is distributed in three ASCII text files per day: geolocated events, geolocated flashes, and supplementary information on the LSQ filter. These files use the following naming convention:

RLF_Level2-YYYYMMDDHH-events.txt
RLF_Level2-YYYYMMDDHH-flashes.txt
RLF_Level2-YYYYMMDDHH-info.txt

where *YYYY*, *MM*, and *DD* corresponds to the year, month, and day, respectively, of the data covered in each file.

Each file holds a 2D array of the events/flashes/entry for that day in the row dimension, and corresponding information for that row in column variables, as described in Tables 3, 4, 5. Uncertainties are 1-σ unless specified. The RMS value of the three or four LSQ residuals (observed minus expected), one per time of arrival observation, is reported for each event. Corrections for time of arrival (additive term) and peak magnitude observations (scaling factor) are found in the supplementary information file for each station along with correction instructions, and station availability can be derived from entries with NaN (Not a Number) in the station's clock error estimate. Stations RLF1 and RLF3 need an extra correction term to account for their clock drift, which is described in the info file. The reported UTC time for events and flashes does not need to be corrected. Currently, gain corrections are constant for the whole campaign. Flash IDs correspond to the row number of the flashes file for the same day, and RLF Level 1 IDs correspond to the n-th entry in the corresponding Level 1 data for that station and hour of the day. Flash quality is provided in the events file indicating the quality of the flash associated with the event, and a quality of 0 is reported for events without a flash association. A minimum flash quality of 0.4 is recommended for most users, since there is a large number of mismatched events/flashes in flashes below the 0.4 measure. Events/Flashes with lower quality are still reported for when they can be manually inspected against the Level 1 data or another data source. Cross-correlations scores between stations show a zero for the reference station chosen and NaN for stations that did not observe the event.

References

M. Cohen, U. Inan, and E. Paschal. Sensitive broadband ELF/VLF radio reception with the AWESOME instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 48 (1):3–17, jan 2010. doi: 10.1109/tgrs.2009.2028334.

M. B. Cohen, R. K. Said, E. W. Paschal, J. C. McCormick, N. C. Gross, L. Thompson, M. Higginson-Rollins, U. S. Inan, and J. Chang. Broadband longwave radio remote sensing instrumentation. *Review of Scientific Instruments*, 89(9):094501, 2018.

R. A. Marshall. An improved model of the lightning electromagnetic field interaction with the d-region ionosphere. *Journal of Geophysical Research: Space Physics*, 117(A3): n/a–n/a, mar 2012. doi: 10.1029/2011ja017408.

R. E. Orville. Calibration of a magnetic direction finding network using measured triggered lightning return stroke peak currents. *Journal of Geophysical Research*, 96(D9):17135, 1991. doi: 10.1029/91jd00611.

Table 3: Table describing the column variables for each event row in the Level 2 events file.

Column #	Variable	Units/Comments
1	UTC Time	Seconds since beginning of day
2	Latitude	Degrees
3	Longitude	Degrees
4	Peak Current	kA
5	Time LSQ Uncertainty	Seconds
6	Latitude LSQ Uncertainty	Degrees
7	Longitude LSQ Uncertainty	Degrees
8	Peak Current Uncertainty	kA
9	3σ-Error Ellipsoid Semi-Major Axis	Degrees
10	3σ-Error Ellipsoid Semi-Minor Axis	Degrees
11	3σ-Error Ellipsoid Azimuth Offset	Degrees
12	Event Quality (Smallest Value in Columns 33-36)	
13	LSQ Residuals RMS	
14	Number of Online Stations	
15	Flash ID	
16	Flash Quality	
17-20	RLF Level 1 ID	
21-24	Uncorrected Time of Arrival	Seconds since beginning of day
25-28	Uncorrected Ch. 1 Peak Magnitude	T
29-32	Uncorrected Ch. 2 Peak Magnitude	T
33-36	Cross-Correlation Scores	

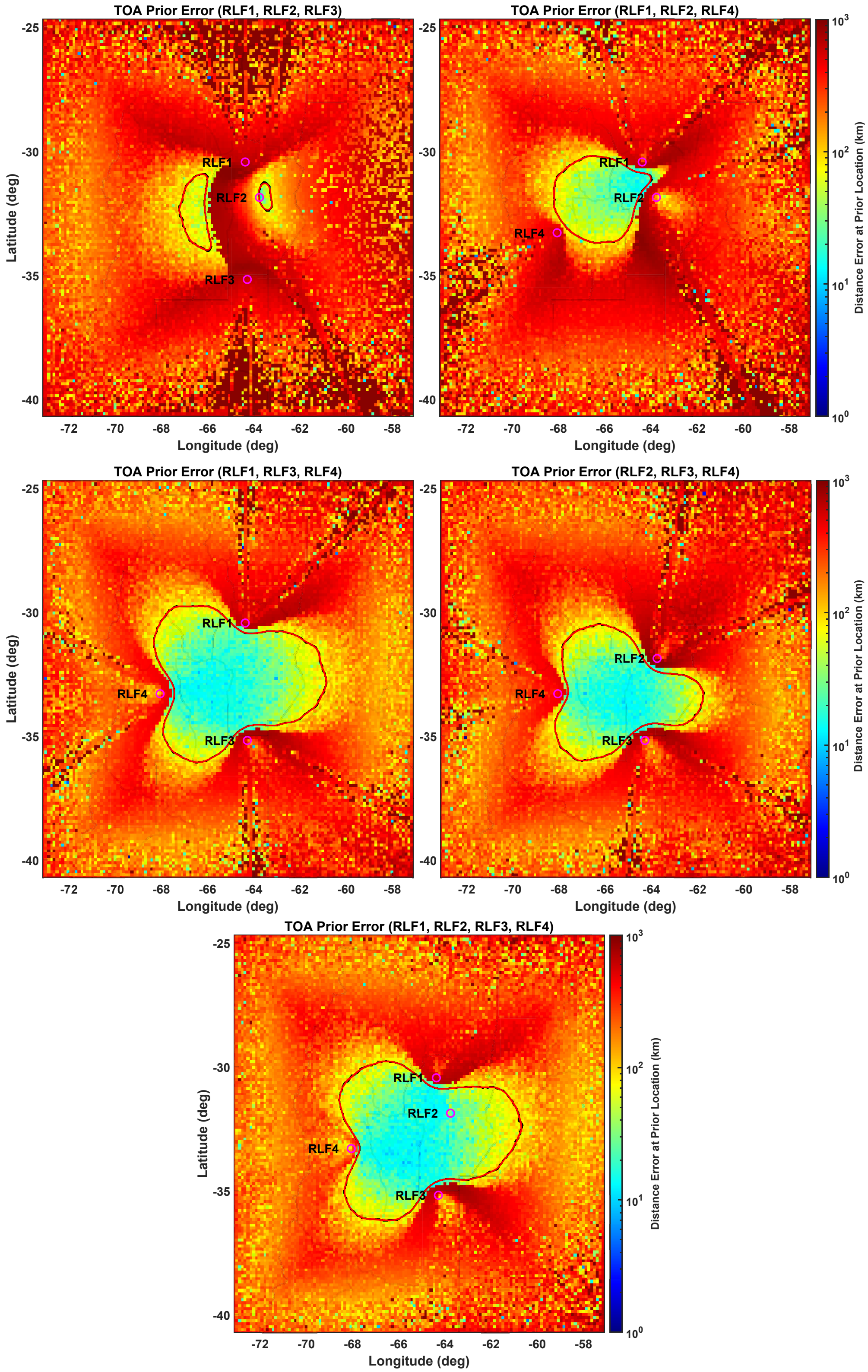

Figure 10: Maps of the *a priori* max error per location for simulated lightning events uniformly distributed throughout the maps. The red circle indicates the chosen array domain mask used to filter out unobservable events.

Table 4: Table describing the column variables for each flash row in the Level 2 flashes file. The last entry on RLF station missing from the observation point to the station that did not observe any of the events contained in the flash.

Column #	Variable	Units
1	UTC Time (Weighted Average)	Seconds since beginning of day
2	Latitude (Weighted Centroid)	Degrees
3	Longitude (Weighted Centroid)	Degrees
4	Max. Event Peak Current	kA
5	Time LSQ Uncertainty (Weighted Average)	Seconds
6	Latitude LSQ Uncertainty (Weighted Centroid)	Degrees
7	Longitude LSQ Uncertainty (Weighted Centroid)	Degrees
8	Max. Event Peak Current Uncertainty	kA
9	Flash Quality (Max. Event Quality)	
10	Flash Multiplicity (Number of Events)	
11	Flash Area (Events' Convex Hull)	km^2
12	Flash Duration	Seconds
13	RLF Station Missing Observations	

Table 5: Table describing the column variables for each entry row in the Level 2 supplementary information file, where each entry corresponding to an observation period between station resets.

Column #	Variable	Units
1	UTC Time Period Start	Seconds since beginning of day
2	UTC Time Period End	Seconds since beginning of day
3-6	RLF Clock Error Correction	ms
7-10	RLF Magnitude Gain Error Scaling Correction	