Artificial intelligence techniques applied to automating meteor validation and trajectory quality control to direct the search for long-period comets

Marcelo De Cicco[1,2,3,4], Susana Zoghbi[1,8], Andres P. Stapper[1], Antonio J. Ordoñez[1], Jack Collison[1,6], Peter S. Gural[5], Sidha Ganju[1,7], José-Luis Galache[1], and Peter Jenniskens[1]

[1] **SETI Institute, 189 Bernardo Ave, Mountain View, CA 94043, USA**
decicco10@gmail.com, susana.zoghbi@cs.kuleuven.be, andresplatastapper@gmail.com, jack10@stanford.edu, siddhaganju@gmail.com, jlgalache@gmail.com, and petrus.m.jenniskens@nasa.gov

[2] INMETRO, Divisão de Metrologia de Materiais, Av. Nossa Senhora das Graças 50, Duque de Caxias, 25250-020, Brazil

[3] Observatório Nacional, Rua Gal. José Cristino 77, 20921-400, Brazil

[4] Exoss.org, Brazil

[5] Gural Software Development, 351 Samantha Drive, Sterling, VA 20164, USA
peter.s.gural@leidos.com

[6] Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

[7] Deep Vision Inc, 2627 Hanover Street, Palo Alto, CA 94304

[8] Computer Science Department, KU Leuven, Belgium

We describe an effort to automate the CAMS (Cameras for All-sky Meteor Surveillance) data reduction pipeline using artificial intelligence techniques to discriminate meteors from other types of detections and to determine correct solutions during triangulation. The effort will make it possible to have the results from a night of low-light video observations available to the observers the following day. As part of the data reduction pipeline, meteors are classified as real and assigned to showers. Results are presented in such a way that each shower can be identified, and new showers from the occasional encounter with the dust trails of long-period comets can be recognized. The detection of such rare showers will allow to direct the search for long-period comets in dedicated deep surveys.

1 Introduction

Long-period comets (abbreviated LPC), due to their potentially large size and fast impact speeds of up to 72 km/s, contribute to the impact hazard on planet Earth along with short-period comets and asteroids. These impacts can severely disrupt the ecosphere and entire human populations. Evidence indicates that the impact of a comet or asteroid with a diameter of about 10 km was responsible for the mass extinction of most species of dinosaurs. Impacts of future threatening asteroids could be mitigated if given sufficient warning time. However, any new long-period comet type on an impact trajectory with the Earth would likely be discovered only 6–12 months before impact, when it becomes visible as the Sun's heat and wind start sublimating its icy surface and ejecting rocky debris.

To provide extra warning time, the orbits of the comets debris could be used by researches to guide the search for comets while they are still far out, providing us years of extra warning time in case they can be detected along that orbit. Most suitable is debris ejected during the previous return of the comet to the inner Solar System, which will cause rare aperiodic meteor showers (outbursts). Therefore, detecting those showers requires a continuous and global search (Jenniskens, 1997).

Figure 1 shows how the debris evolves from a meteoroid coma into a meteoroid stream in one revolution as a result of differences in orbital period between grains. The intersection point at the Earth's orbit (the node) of those orbits is not constant. As described by Jenniskens (1997), LPC outbursts are due to gravitational perturbations on the individual meteoroid orbits, causing a periodic displacement of the stream relative to the Earth's orbit, which follows the Sun's reflex motion around the barycenter, with dominating contributions by Jupiter (12-year period) and Saturn (30-year).

Given that the long term hazard posed by long-period comets is statistically comparable to other naturally occurring events (Chapman and Morrison, 1994), it is important to design detection and prevention strategies in order to mitigate potential impacts.

To do so, the night sky needs to be monitored for an extended period of time (approximately 60 years) and from locations around the globe (Jenniskens et al., 1997;

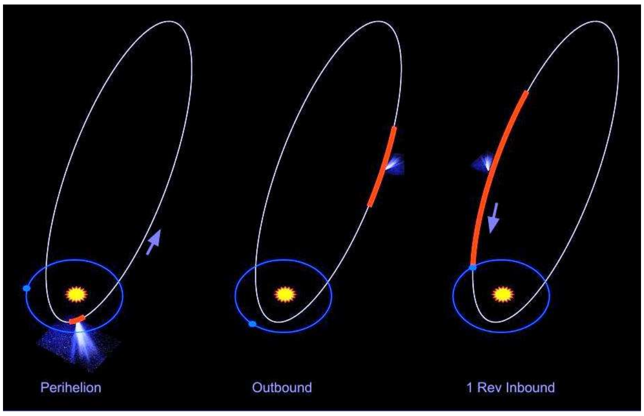


Figure 1 – A diagram showing debris trails formed during LPC travel to the inner Solar System. An LPC outburst happens when the Earth (the blue point) crosses that debris. Created by Peter S. Gural.

Lyytinen and Jenniskens, 2003). Low-light video camera surveillance of the night skies aimed at detecting previously observed and unobserved meteor showers has been demonstrated to produce meteoroid orbits that be used to identify where in the sky a parent long-period comet may reside when it still many years out.

In this study, we improved the data reduction pipeline for CAMS, the Cameras for All-sky Meteor Surveillance project (Jenniskens et al., 2016), to make it possible to operate this network into the future and create more of such networks globally. To achieve this goal, we set out to improve and automate the classification of meteors from non-meteors using machine learning and deep learning approaches and to improve the data visualization tools to recognize new meteor showers.

2 CAMS pipeline automation

The ability of CAMS to successfully identify meteor showers associated with potentially hazardous, long-period comets relies heavily on its time coverage, cadence, and total field-of-view. Meteoroids associated with these comets will appear sporadically in time, creating non-annual showers that last only between one and a few hours. When such showers have not been previously detected, they cannot be predicted. Observing these unpredictable meteor showers thus requires a coordinated, global effort amongst the meteor observing community.

For this effort to proceed efficiently, it is imperative that the data reduction involved be performed in an automated, systematic fashion. The overarching goal of this work is to automate the reduction pipeline for the CAMS network in order to perform the systematic identification and characterization of previously unknown meteor showers. Before this project, the CAMS reduction pipeline required an inordinate amount of human effort to calibrate, confirm, and triangulate meteor orbits. Thus, the major goal with this automation is to provide software tools that completely remove the human element from the CAMS data reduction from end to end. Ideally, we intend to proceed from image capture to orbit calculation without any human intervention without loss of fidelity to the results.

For this goal, a pipeline was implemented performing the following steps on CAMS data automatically: (1) meteor confirmation on operator end (i.e., on the local machines capturing the data at observing sites); (2) data retrieval; (3) data processing; (4) coincidence; and (5) data visualization.

The proposed automation involves the development of a series of software tools which perform these processes efficiently and cross-platform. Specifically, we developed Python scripts which implement this procedure from end to end. Step 1 provides code that will implement the automated confirmation learned on the observing machines. Step 2 requires automated FTP of detected object data from a server holding incoming data from CAMS sites around the world. Steps 3 and 4 interface with the existing CAMS software to process the incoming data automatically. Step 5 will combine the trajectory solution classifier trained on coincidence data (see Section 4) with the existing CAMS coincidence software to automatically find the best orbital solution for each incoming meteor. Finally, we provide tools for the public to visualize the CAMS meteor shower data on the web in an effort to ensure continued amateur engagement in the project.

3 Confirmation automation using Deep Learning

Current practice is for observers to visually inspect the CAMS detections to determine whether or not a detection is a meteor, or a non-meteor. This is a time consuming task that can be automated.

As a proof of concept for confirming meteor candidates using machine learning, we trained a random forest classifier (Breiman, 2001) on a training set of approximately 200 000 CAMS object detection data. This dataset consisted of time-series information for meteors, aircraft, clouds, and other objects identified in CAMS data (see Figure 2). For this classifier, we focused on the time-series of spatial and photometric data for each object (see Figure 2). Given the heterogeneity of the data (i.e., disparate numbers of data points), we opted to mine metrics from these time-series data that describe the trajectories and light curves of each object. For the

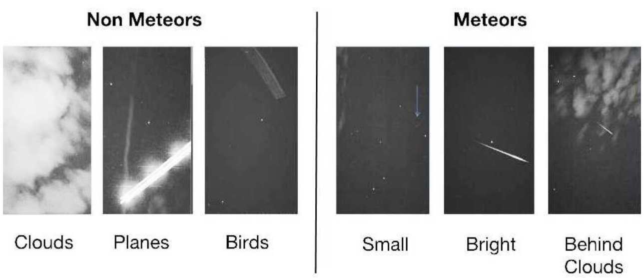


Figure 2 – Non-meteor and meteors images from CAMS.

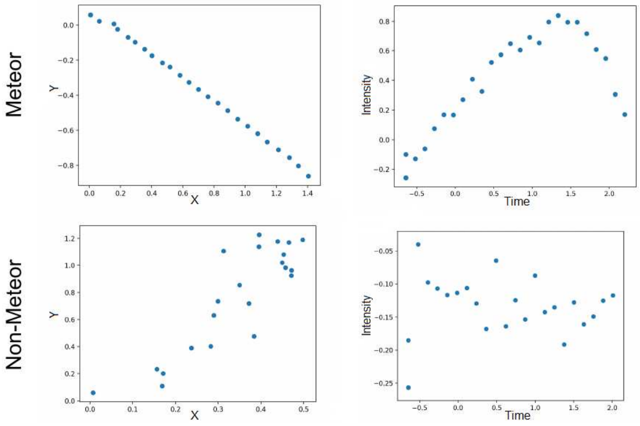


Figure 3 – Mining tracklets for features. In the trajectory graphs, distances are measured in pixels of images (*left*); light curves show intensity versus time (*right*).

trajectories, we calculated the coefficient of determination (R^2) and residual standard deviation of a best-fit line to the XY paths in the images, as well as the total distance traveled. Regarding the photometric variations, we extracted several statistical measures of the shape of their light curves including the mean intensity, median absolute deviation (MAD), skew, and kurtosis. We also included in this model two measures of the timescale of the event: the total time observed and the optimal period from a fast Fourier transform on the light curves.

In order to verify that these selected features do provide useful information for an object classifier, we examined the relationship of each feature to the object class. Figure 4 displays an example of such an investigation. This illustration shows that the light curve mean and MAD do seem to correlate with the different object classes.

Following this verification, we trained a Random Forest classifier to perform a binary classification of meteor vs. non-meteor. To account for the class imbalance in this training set (i.e., only 3% of the data are true meteors), we employed a class weighing scheme that accounts for the different class frequencies. Without any hyperparameter tuning, we achieved a meteor classification precision and recall on a test set of 90% and 81%, respectively.

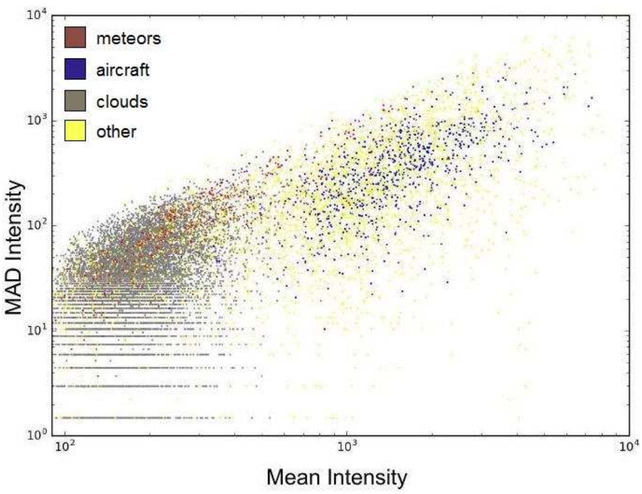


Figure 4 – An example distribution of light curve features (light curve mean and MAD) that appear to display correlations with different objects (maroon, meteor; blue, aircraft; grey, clouds; yellow, other).

Also, a classifier was implemented that takes as input a set of image frames and returns as output a probability score indicating whether or not the frames contain a meteor. Convolutional Neural Networks (CNNs)(Krizhevsky et al., 2012) provide an ideal architecture to tackle this effort since they are responsible for state-of-the-art performance in computer vision tasks such as object classification. Three main functional layers are used in their architecture: the convolutional, pooling (subsampling), and fully-connected layers (see Figure 4).

The convolutional layer consists of small, learnable filters. Each filter corresponds to a set of weights. During the forward pass, we convolve each filter with the input by successively computing dot products between small windows (also known as local receptive fields) of the input and the filter. As we slide the filter across the input, we produce an activation map that gives the responses of that filter at every spatial position. During backpropagation, we learn how to update the weights to minimize the errors the network makes on a given task. Intuitively, each filter learns to detect specific patterns, e.g., edges and color patches. The activation maps are then stacked together and become the input of the next layer.

The pooling layer down-samples the results from the previous layer by using a simple operation (e.g., the maximum or L2 norm) over a small window. The purpose of this pooling operation is to sort of summarize the findings of the activation map. It reveals whether a given feature is found anywhere in a region of the image and it discards the exact positional information. The rationale is that once a feature has been found, its exact location is not as important as its rough location relative to other features. The advantage is that it reduces the number of parameters needed in subsequent layers.

In a fully connected layer, neurons have full connections to all activations in the previous The output is simply computed by a matrix multiplication between the input and the weights, followed by a bias offset. The weights and the biases are learned through the backpropagation algorithm. In a classification task, the output of the last fully connected layer is usually passed through a softmax function to compute the probability of each class, which comprises the network predictions.

A typical CNN architecture is composed of a series of convolutional layers each followed by a pooling layer. Then a couple of fully-connected layers follow which feed into the final output. An example architecture is shown in Figure 5.

The network used in our study used five convolutional layers followed two fully connected layers and a binary softmax classifier. The network needs a large dataset

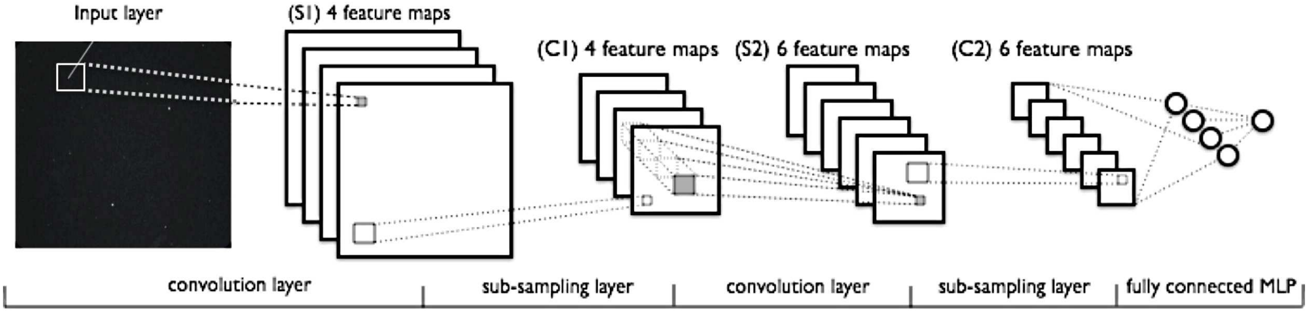


Figure 5 – Schematic diagram of a deep neural network.

to give reliable results. We performed standard data augmentation techniques on instances of the positive class, such as rotation and flipping.

Our CNN achieves precision and recall scores are 88.3% and 90.3%, respectively. Precision describes the percentage of objects assigned as "meteors" to be actual meteors in a pre-classified data set. The recall describes the percentage of actual meteors in that dataset that were detected as such. That means that a CNN would let about 12% of false detections through, while losing about 10% of all meteors in the data.

In addition, we trained a Long-Short Term Memory (LSTM) network that encodes the light curve tracklets into a latent space, and learns to predict whether or not the tracklet corresponds to a meteor. The LSTM achieves a precision of 90.0% and a recall of 89.1%. One key advantage of using Deep Learning is that we did not have to hand-engineer the meaningful features from both images and light curves. The models learned these on their own.

The advantage of using a machine learning approach is that it generally requires less training data and computation time when compared with deep learning. On the other hand, manually extracting and engineering features from the data will require more human time pre-training. In the end, we will explore combining this approach with the deep learning results in an ensemble classifying scheme.

4 Coincidence

Once a meteor has been confirmed the information from different cameras is used to generate a trajectory for each meteor. Deviations in the video capture can lead to erroneous solutions and improper trajectories and orbits. These deviations can be identified from the light curves, the latitude and longitude and the velocity of the observed meteor measured by each camera.

Current practice is for a data analyst to inspect each meteor trajectory calculated from meteor detections at two or more sites. The data are presented in the form of a light curve, a geographic projection of the trajectory and a side-view of the trajectory. The effort is labor intensive and time consuming and can be automated.

Considering the heterogeneity of the data, using the raw coincidence data in a machine learning classifier presents a significant challenge. As an alternative to this, we propose to proceed similarly as with the machine-learning-based approach for the meteor confirmation.

To summarize the method, we plan to extract singular metrics describing the goodness-of-fit for an individual coincidence solution. The two feature sets most important to this coincidence is the trajectory and light curve resulting from a solution. Thus, we propose to extract some descriptive features from these feature spaces for each object in an effort to define a parameter space conducive to discriminating between good and poor solutions. One example of such a metric is the collinearity of the combined, three-dimensional trajectory of a solution, incorporating data from all cameras that observed the object.

An optimal solution produces high collinearity between the trajectories from different cameras. Another example is the unit-lag autocorrelation of the light curve differences with respect to some reference light curve. A mismatching light curve should exhibit an anomalously high autocorrelation, indicating different light curve structures.

Taking these methods, we hope to train a machine learning classifier to distinguish good solutions from poor ones. The goal with this project is to implement this classifier in the automated CAMS pipeline. For this, it may be necessary to search through much of the camera combinatorial space when coinciding new datasets. We plan to use a hierarchical method which first scores the goodness of the initial solution using all N cameras for an object. If this solution is poor, we will then search through the next series combinations using $N-1$ cameras to search for an optimal solution. This iterative procedure will be repeated until an optimal solution is reached.

At the time of writing, we only implemented relatively simple classifiers in Coincidence based on the light curve shape (i.e., needing to have an F_{skew} shape between 0.05 and 0.95) and maximum errors in geographic positions. Plotting up such data as in Figure 4 showed meteors and non-meteors to separate to better than circa 70%.

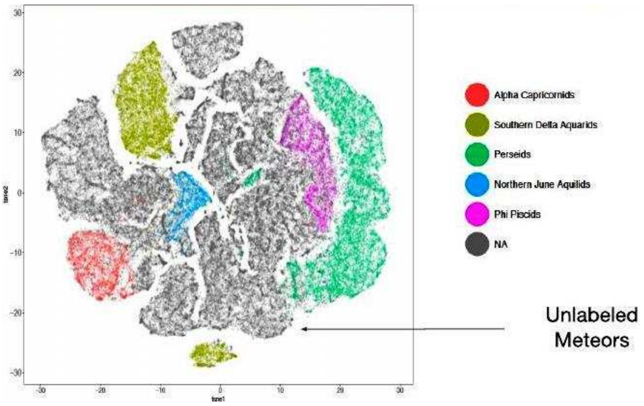


Figure 6 – Graphic showing groupings identified by t-SNE among meteor orbital parameters, compared to groupings (meteor showers) previously identified by visual inspection.

.

5 Data clustering: searching for outbursts and new meteor showers

Once the meteoroid orbits are calculated, the final step is to identify outbursts and new showers. Using a dataset of 122 295 CAMS meteoroid orbits preliminarily classified to showers, we tested the use of unsupervised machine learning (Hodeghatta Rao and Nayak, 2017) to identify newly recognized meteor showers and outbursts.

First, all parameters related to a meteoroid orbit were separated, then clustering methods were tested in order to identify the most effective ones for recognizing showers. The following methods were applied: dimension reduction via PCA and t-Stochastic Neighbor Embedding (t-SNE). We found that the IAU list shower classification was correctly clustered only by t-SNE, as shown in Figure 6.

A number of clusters are readily seen in Figure 6 that might also be as of yet unidentified meteor showers. In the next step, unsupervised machine learning with DBSCAN (density-based spatial clustering of applications with noise) was used to identify these groupings. Figure 7 shows some of the groups identified by DBSCAN. Some previously unidentified groups are indicated. They may represent previously unrecognized meteor showers. Apart from the diffuse groups associated with the main showers and with sporadic sources, there are also some compact clusters in these data. Those are potential meteor outbursts. DBSCAN can be used to identify such outbursts in an automated way.

6 Visualization

It is current practise to collect the meteoroid orbits over a period of time, then analyze for the presence of meteor showers. This provides insufficient feed-back to the observers. To keep up efficiency, it is important to present the results in near-real time.

To do so, we developed Java scripts to display the meteor radiant positions of each network. We chose to

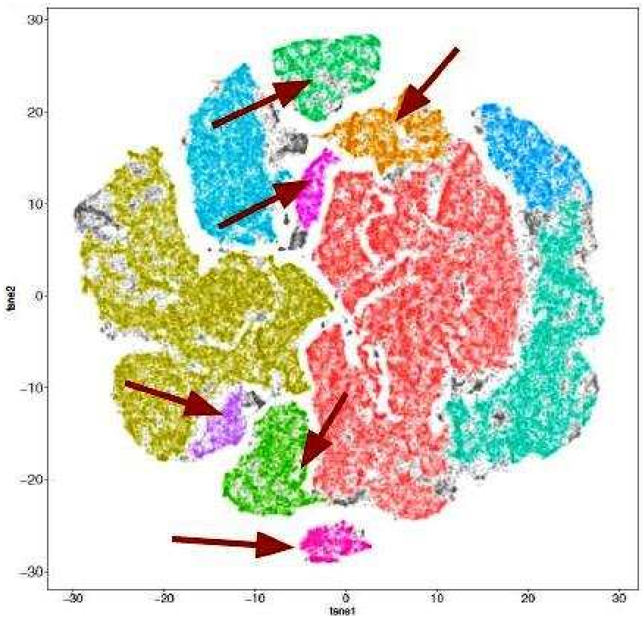


Figure 7 – In addition to established meteor showers, the Unsupervised Machine Learning with DBSCAN method yielded clusters corresponding to possible meteor showers not described before. These clusters are indicated by maroon arrows.

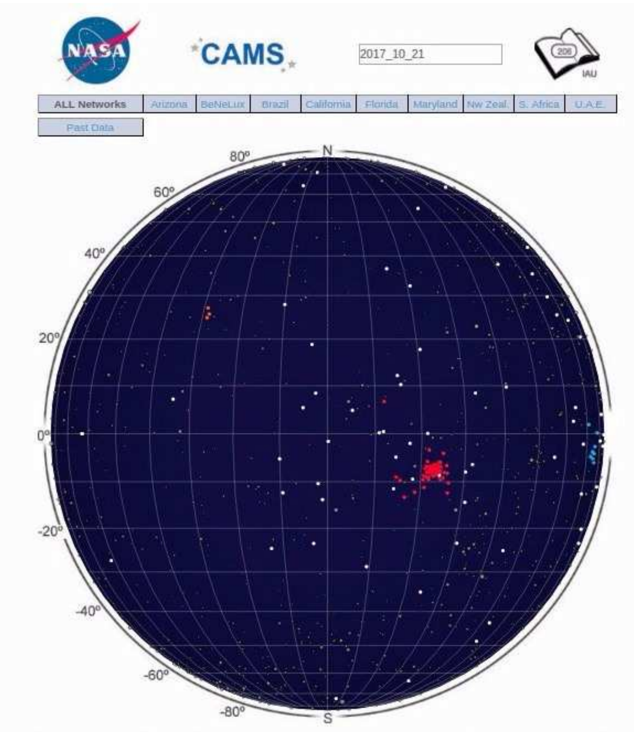
present the data on a sphere[1] that can be rotated so that at each point in time the whole sky can be examined and radiants near the poles are not spread out. The results are presented in Sun-centered ecliptic coordinates in order to provide a constant perspective (direction of Earth motion center of the graph when first displayed). The meteor showers are identified by comparing the new orbits to a look-up table of previously assigned meteor showers. The current look-up table is based on over 900 000 meteoroid orbits detected in all major video-based meteoroid orbit surveys (CAMS, SonotaCo, Edmond, and CMN).

The final display makes it possible to see the active meteor showers in any given night (Figure 8). Results from current data can be compared to past data (2010–2016) of shower-assigned CAMS data. That makes it possible to see unusual meteor shower activity and new showers. By hovering over a point produces the IAU meteor number and by clicking on a point brings up that shower in a planetarium program. This new method for visualizing meteor radiants will improve public interest in meteor science, as well as allowing a novel approach to studies of new meteor showers.

7 Conclusions

In order to fully automate meteor shower detection and classification from video observations, we applied Artificial Intelligence (AI) methods of Deep Learning and Machine Learning. Deep learning enabled a simple meteor/non-meteor classifier that took as input a set of max-pixel image frames and outputs a probability score that indicated whether or not the frames contained a

[1] `http://cams.seti.org/FDL`.



IAU Number	Code	Shower Name	Number of Meteors
0	SPO	sporadic meteors	[74]
2	STA	Southern Taurids	[7]
8	ORI	Orionids	[108]
17	NTA	Northern Taurids	[1]
22	LMI	Leonis Minorids	[3]
23	EGE	epsilon Geminids	[1]
[illegible]	[illegible]	Northern October delta Arietids	[1]
[illegible]	NUE	nu Eridanids	[2]

Figure 8 – A web-based interactive graphic showing meteor radiants plotted on the celestial sphere.

meteor. Machine Learning techniques were used to evaluate the triangulated trajectories to determine whether a solution was correct/incorrect.Machine Learning was also used to identify known and previously uncharacterized showers.

In doing so, we are able to identify meteoroid orbits potentially associated with long-period comets that pass close to Earth's orbit. When they were ejected in the previous return, the orbits of the meteoroids should directly trace that of their parent comet. Taking into consideration uncertainties in the orbital parameters of these meteoroids, the search space for these hazardous comets can be narrowed down. These regions can be then probed with dedicated, deep-sky surveys searches to attempt to locate these long-period comets.

And, finally, the interactive web-based tool enables an increase public commitment in meteor shower observations, which will hopefully result in expanded night sky surveillance coverage. The goal is to identify new, rare and non-episodic meteor showers and their potentially hazardous long-period comets.

Acknowledgements

This effort was executed as part of NASA's Frontier Development Lab, a research accelerator led by James Parr and Jason Kessler and by the SETI Institute's Bill Diamond. We acknowledge support of the NASA Portal (Robert B. Pittman and Daniel J. Rasky). The effort was sponsored by NASA (Victoria Friedensen and Darlene Weidemann) and by private parties, in particular IBM (Troy Hernandez and Graham Mackintosh) and NVIDIA (Alison Lowndes).

References

Breiman L. (2001). "Random forests". *Machine Learning*, **45**, 5–32.

Chapman C. R. and Morisson D. (1994) "Impacts on the Earth by asteroids and comets: assessing the hazard". *Nature*, **367**, 33–40.

Hodeghatta Rado U. and Nayak U. (2012). *Business Analytics Using R—A Practical Approach*. Apress, Chapter 7, pages 161–186.

Jenniskens P. (1997). "Meteor stream activity. IV. Meteor outbursts and the reflex motion of the Sun". *Astronomy & Astrophysics*, **317**, 953–961.

Jenniskens P., Betlem H., de Lignie M., and Langbroek M. (1997). "The detection of a dust trail in the orbit of an Earth-threatening long-period comet". *The Astrophysical Journal*, **479**, 441–447.

Jenniskens P., Nénon Q., Albers J., Gural P. S., Haberman B., Holman D., Morales R., Grigsby B. J., Samuels D., and Johannink C. (2016). "The established meteor showers as observed by CAMS". *Icarus*, **266**, 331–354.

Krizhevsky A., Sutskever I., and Hinton G. E. (2012). "ImageNet classification with deep convolutional neural networks". In Pereira F., Burges C. J. C., Bottou L., and Weinberger K. Q., editors, *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105.

Lyytinen E. and Jenniskens P. (2003). "Meteor outbursts from long-period comet dust trails". *Icarus*, **162**, 443–452.