EPSC2018

MD2/MTI4/LFI4 abstracts

13 Sep 2018 15:36:46

EPSC Abstracts
Vol. 12, EPSC2018-679-1, 2018
European Planetary Science Congress 2018

Automated detection of planetary craters: open and reproducible benchmark platform for the Martian surface

Joris Van den Bossche (1,3), Alexandre Boucaud (2,3), Frédéric Schmidt (4), Guillaume Lemaitre (1,3), Anthony Lagain (4), Alina Meresescu (4) and Balázs Kégl (2,3)
(1) Parietal team, Inria, CEA, Université Paris-Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France
(2) LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
(3) Paris-Saclay Center for Data Science, Université Paris-Saclay, Orsay, France
(4) GEOPS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, Rue du Belvédère, Bât. 504-509, 91405 Orsay, France

Abstract

We present an open framework to benchmark crater detection algorithms, based on infrared images of the Mars surface acquired from the THEMIS survey and a recently revised catalogue of all Martian craters with a diameter larger than $1\,\mathrm{km}$, representing $\sim$400 000 entities. Within this framework, we clearly define the problem and the model evaluation (i.e. data sets, metrics, and cross-validation). This framework is embedded in the Rapid Analytics and Model Prototyping (RAMP) computing platform which aims at providing a fair comparison of present and future methods. The platform has been deployed during a beta event to show a proof-of-concept.

1. Introduction

Impact craters are one of the most prominent geological features of telluric planets, yielding important information on their geological history [1]. Although any human with a short training can detect craters in an image, automating such process remains an open challenge. Consequently, crater detection algorithms have received a particular attention in recent years [2, 3, 4, 5, 1, 6]. As recently pointed out by Pedrosa *et al.* [7], it remains, however, difficult to make a fair comparison between those methods: the data sets and the metrics differ between studies.

To our knowledge, only Salamunićcar *et al.* [8] has proposed a framework addressing those issues. However, this framework, now a decade old, uses an outdated ground-truth catalogue, lacks standard metrics for object detection, and reproducing the evaluation is not easy. Herein, we propose an open framework integrated within a computing platform allowing for fair and reproducible comparison between crater detection methods.

2. Evaluation framework

In this regard, we define a common data set based on the full Martian surface, using the latest available crater catalogue as ground-truth as presented. We also define a testing methodology and metrics to compare the algorithms.

2.1. THEMIS data

We used THEMIS [10] day-time infrared image, a huge mosaic of the full Mars surface at 100m/pixel in cylindrical projection as the main dataset. We processed the THEMIS mosaic, one quadrangle at a time. Each quadrangle was first reprojected to their local stereographic projection. From these quadrangles images, we extracted all possible $224\,\mathrm{px} \times 224\,\mathrm{px}$ images - hereafter referred to as tiles - using an overlap of $56\,\mathrm{px}$, to make sure each crater would fit entirely in one tile at least. We also down-sampled the quadrangle images to cut additional tiles, using identical methodology, to include the craters too big to fit in a tile of the original image resolution.

2.2. Ground-truth catalogue

The ground-truth catalogue used for this benchmark is currently the most accurate database of Martian craters with a diameter larger than $1\,\mathrm{km}$ [9]. The catalogue is based on the work of Robbins *et al.* [11], cleaned up by human-operation [9]. It contains 376,439 verified impact structures larger than $1\,\mathrm{km}$ in diameter.

2.3. Testing methodology

The tiles dataset created is large enough to be split into independent training and testing sets. We selected tiles from the most heterogeneous quadrangles (position, crater population, etc.) to maximize the characteristics of the data while keeping a manageable volume.

Cross-validation was applied to the training data, using the quadrangles as folds. We used the hold-out testing data to perform the evaluation as well as checks for consistency with cross-validation scores.

2.4. Performance measures

A classic metric in object detection, the Intersection-over-Union (IoU), also known as the Jaccard index [12], is used as a similarity criterion between two objects that provides a good evaluation of both the location and size accuracy of a given prediction with respect to a crater.

As the final detection score, we compute the average precision, a performance metric based on the precision-recall curve. This metric aims at penalizing algorithms that lack of balance between precision (good predictions) and recall (completeness of the predictions).

3. Benchmark platform (RAMP)

The testing methodology presented in the previous section has been implemented within the Rapid Analytics and Model Prototyping (RAMP) platform[1]. The RAMP platform aims at the development of open source, collaborative, and reproducible solutions.

Concretely, the problem has been formulated as a detection problem in which the center and the radius of each crater has to be predicted. The dataset has been pre-processed and the evaluation has been fixed following the testing methodology. Therefore, the contributors willing to benchmark their algorithm are just required to provide the source code of the detector with predefined input and output format. Subsequently, each algorithm is tested on Amazon S3 on hardware with identical specification, ensuring consistency. All data and code described in this document are available online [2].

4. Preliminary results and conclusions

An initial version of this benchmark platform was used during a data science course at École Polytechnique (Palaiseau, France) in November 2017. Algorithms both using computer vision and deep learning methods were implemented. A next edition will occur at the end of July 2018 at the CIFAR summer school[3] in Toronto, Canada ; and a dedicated edition could be organized for the conference.

[1] `https://ramp.studio`
[2] `https://github.com/ramp-kits/mars-craters`
[3] `https://dlrlsummerschool.ca`

Acknowledgements

The authors thank USGS for providing the THEMIS mosaic. They also thank Olivier Grisel for useful comments.

References

[1] Urbach et al., *Planetary and Space Science*, 57, 7, 880–887, 2009.

[2] Salamunićcar et al., *IEEE Transactions on Geoscience and Remote Sensing,* , 48, 2317–2329, 2010.

[3] Bandeira et al., ,*IEEE Transactions on Geoscience and Remote Sensing*, 45, 12, 4008–4015, 2007.

[4] Martins et al., *IEEE Geoscience and Remote Sensing Letters*, 6, 1, 127–131, 2009.

[5] Kim et al., *Photogrammetric Engineering & Remote Sensing*, 71, 10, 1205–1217, 2005.

[6] Chen et al., *IEEE Transactions on Geoscience and Remote Sensing*, 1–12, 2018.

[7] Pedrosa, et al., *Geomatics, Natural Hazards and Risk*, Vol. 8, pp. 1306-1329, 2017

[8] Salamunićcar et al.,, *Advances in Space Research*, 42, 1, 6–19, 2008.

[9] Lagain, et al.,, *Journal of Geophysical Research*, 2018, submitted

[10] Edwards, C. S.,et al., *J. Geophys. Res.*, 116, E10008, 2011

[11] Robbins, et al., *Journal of Geophysical Research*, 117, E05004, 2012

[12] Jaccard., P., *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, (142), 547–579, 1901

EPSC Abstracts
Vol. 12, EPSC2018-706-1, 2018
European Planetary Science Congress 2018

Attempting rapid exoplanets' classification with neural networks

Sebastiano Padovan (1), Philipp Baumeister (2), Doris Breuer (1), Frank Sohl (1), Tilman Spohn (1), and Nicola Tosi (1,2)
(1) Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany, (2) Technische Universität, Berlin, Germany (sebastiano.padovan@dlr.de)

Abstract

Despite the number of confirmed exoplanets currently exceeds 3500, for only few hundreds of them both mass and radius have been determined (e.g., https://exoplanetarchive.ipac.caltech.edu). For these objects the mean density can be calculated to obtain a first handle on the interior structure. However, inferring the interior structure from just mass and radius is a highly non-unique problem (e.g., [1]). The forthcoming space-based telescopes TESS, CHEOPS, and PLATO will provide accurate determination of the radii of thousands of exoplanets (e.g., [2]). These improved instrumental capabilities along with the increasing temporal baseline of exoplanetary orbits' observations may also allow the inference of the fluid Love number k_2 (e.g., [3]), a parameter that depends on the concentration of matter in the interior and is akin to the moment of inertia (e.g., [4]). Exoplanetary masses will be determined through follow-up campaigns using ground-based telescopes. In the best case scenario, three constraints–mass, radius, and k_2–will be available to investigate the interior structure of exoplanets.

The goal of this work is to understand whether from the determination of the radius and k_2 only, it will possible to classify an exoplanet–e.g., earth-like versus neptune-like–and if any bounds on its mass can be placed in advance of follow-up campaigns that will determine its mass from radial velocity. We test this hypothesis using a neural network approach. We build a large (in excess of 10^5) set of interior structure models in the mass range between 1 and 20 earth masses, where the super-Earth and mini-Neptune classes overlap. The parameters entering the model (i.e., total mass, core mass fraction, silicate reference densities, water/ice mass fraction, etc.) represent the inputs for the forward model, while the resulting radius and k_2 represent the outputs. Since the forward problem is only mildly non-linear, we will use neural network architectures with one or two hidden layers with three to five neurons each. We will adopt the common training, validation, and test split of the data (in proportion 80:10:10).

If this approach at planets' classification proves successful, it will be possible to rapidly draw preliminary inferences on an exoplanet's interior in advance of mass determinations from follow-up radial velocity campaigns.

Acknowledgements

This work was supported by the DFG within the Research Unit FOR 2440 "Matter Under Planetary Interior Conditions" and through the priority program SPP 1992 "exploring the diversity of extrasolar planets" (project TO 704/03).

References

[1] Rogers, L.A. and Seager, S.: A framework for quantifying the degeneracies of exoplanet interior compositions, Astrophys. J., Vol. 712, 974-991, 2010.

[2] VV.AA.: PLATO definition study report, ESA-SCI(2017)1, 2017.

[3] Hellard, H. et al.: Measurability of the fluid Love number k_2 in WASP-121b, Abstract EPSC2018-310, 2018.

[4] Padovan, S. et al.: Fluid Love numbers with the matrix propagator method with an application to GJ436b, Abstract EPSC2018-684, 2018.

EPSC Abstracts
Vol. 12, EPSC2018-738, 2018
European Planetary Science Congress 2018

Semi-automated surface mapping via unsupervised classification Mercury'S Visible–Near-Infrared reflectance spectra

Mario D'Amore(1), Jörn Helbert(1), Alessandro Maturilli(1), Indhu Varatharajan (1).

(1) Institute for Planetary Research, DLR, Rutherfordstrasse 2, Berlin, Germany (mario.damore@dlr.de)

1. Abstract

The surface of Mercury has been mapped in the 400–1145 nm wavelength range by the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) instrument during orbital observations by the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft. Under the hypothesis that surface compositional information can be efficiently derived from spectral reflectance measurements with the use of statistical techniques, we have conducted unsupervised hierarchical clustering analyses to identify and characterize spectral units from MASCS observations. The results display a dichotomy, with two spectrally distinct groups: polar and equatorial units (see Fig. 1). The spatial extent of the polar unit in the northern hemisphere generally correlates well with that of the northern volcanic plains [1]. We extended our analysis on the latest MESENGER data delivery to PDS including the new spectral photometric correction ([2] in review, extension of [3]) , finding result consistent with our previous analysis based on our custom photometric effect removal.

2. Methods

2.1 Data managing: PostgreSQL

The most recent version of our data analysis procedure uses PostgreSQL, a type of database management that controls the creation, integrity, maintenance and use of a database. It embeds a high-level query language, which greatly simplifies database organization as well as retrieval and presentation of database information. We set up a data pipeline using the to update automatically the MASCS data, read them from the NASA Planetary Data System format, regrid the data to a common grid length, and store all information in the database. All data are then readily available to any authorized user in our network. We are working on a library to access the data directly from within our analysis software, and some preliminary functions have been implemented. As an example, the calculation of a parameter representing the database takes a few seconds even for the full dataset of ~5 million entries, if exploiting pre-indexed columns. It is thus straightforward to create and analyze rapidly the data, as for example the distribution of normalized radiance at a fixed wavelength. The new methodology provides facilities for controlling data access, enforcing data integrity, managing concurrency control, and recovering the database after a failure and restoring it from backup files, as well as maintaining database security.

2.1 Data retrieval : PostGIS

We use PostGIS that adds support for geographic objects in geographic information system and extends the database language with functions to create and manipulate geographic objects. A typical application is the definition of a large number of regions of interest (ROIs) and the search for all data points falling within each ROI. This facility may be used to extract spectral signatures specific to user-defined geological units in a few seconds and to explore the properties of the data from the different ROIs. A typical search for data from areas defined by a simple ROI, such as regions of impact melt associated with a given crater, takes less than 1 second. More elaborated query requires more computational power and result in longer response time. A typical example is the search for a MASCS measurement where no point outside a ROI were taken, that could increase the response time to > 60 seconds. We resample the whole dataset of ~5 Milion spectra on a planetary fixed grid or extract the information from collection of planetary region of interest in few minutes, allowing to quickly analyze the spectral characteristic of Mercury. We successfully tested remote access to the database using through a GIS visualization system, creating data visualization on the surface of Mercury that layers camera data and real-time-queried MASCS data.

solar lines in the calibrated spectra. The resultant hyperspectral map was then visually inspected to search for anomalies that originated mainly in regions of low coverage or from high levels of spectral variation within a single pixel. Our approach consist of 1. a data cleaning step, to remove data artifact, 2. Principal Component Analysis (PCA) feature compression and 3. K-means clustering (see scikit-learn python implementation in [4]). We found the existence of two large and spectrally distinct regions, which we call the polar spectral unit (PSU) and the equatorial spectral unit (ESU) (Fig. 1). Further analysis indicates the presence of smaller sub-units that lie near the boundaries of these large regions and may be transitional areas of intermediate spectral character.

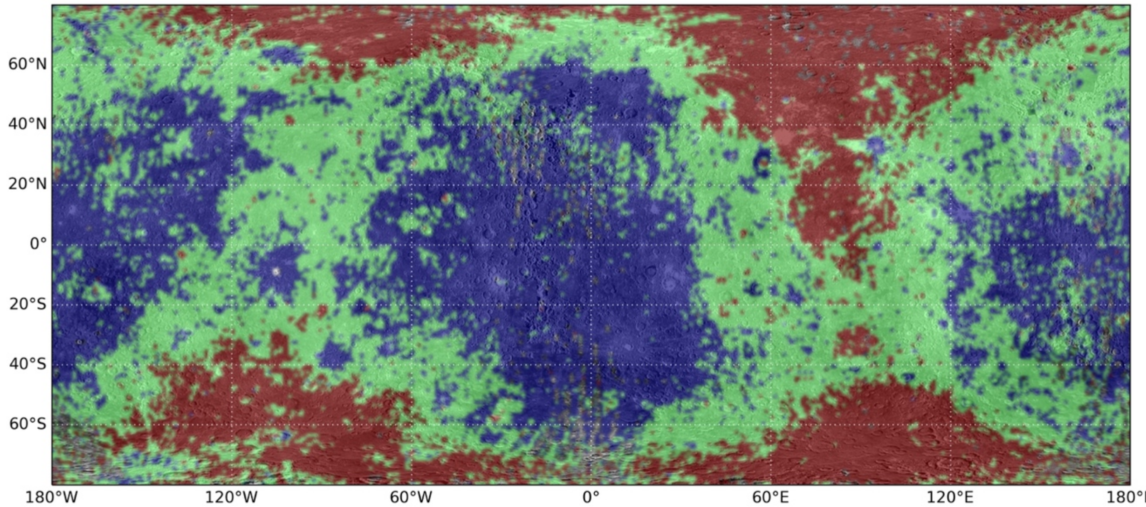

Figure 1. Spatial distribution of the unsupervised hierarchical clusters derived from normalized MASCS VIS spectra (overlaid on an MDIS base map). Red pixels indicate the polar spectral unit (PSU); blue pixels denote the equatorial spectral unit (ESU). Green is the intermediate transition region.

2.1 Machine Learning on multivariate data

We use a global hyperspectral data cube image of normalized MASCS visible (VIS) detector spectra, from the first Earth year of the orbital mission, to perform our unsupervised hierarchical clustering analysis. We reduced the grid resolution to 4°/pixel to improve the spatial coverage of the final map, but at the cost of increased sub-pixel variation. Data coverage varies from region to region, but global maps at 1°/pixel can be obtained with a high signal-to-noise ratio (SNR). With the absence of a formal global photometric correction for the MASCS data, we have corrected the dataset in an approximate fashion by normalizing all the spectra at 700 nm to account for large variations in observing geometry. We have excluded the most extreme observing geometries by limiting the incidence and emission angles to <85°, which means that latitudes poleward of 80° are excluded. For this analysis we used six spectral channels, each with a bandwidth of 10 nm, in order to focus on “interesting” spectral regions, e.g., bands at ~600 nm that are indicative of sulfides. By using wider spectral channels, we also avoided biases caused by artifacts in the spectra, e.g., the presence of

References

[1] Head, J. W. et al. (2011), Science, 333, 1853–1856.

[2] Domingue, D. L. et al, Icarus, in review.

[3] Domingue, D. L. et al, Icarus, 209, 2010, 101-124.

[4] Pedregosa, F., Journal of Machine Learning Research, 2013, 108-122

EPSC Abstracts
Vol. 12, EPSC2018-798-2, 2018
European Planetary Science Congress 2018

Deep Learning-Based Anomaly Detection to Find Changes over the Martian South Pole

Alfiah Rizky Diana Putri (1), Panagiotis Sidiropoulos (1,2) and Jan-Peter Muller (1)
(1) Imaging Group, Mullard Space Science Laboratory (MSSL), University College London, Department of Space & Climate Physics, Holmbury St Mary, Dorking, Surrey, RH5 6NT, United Kingdom (2) Cortexica Vision System, 30 Stamford Street, WeWork South Bank Central, London, SE1 9LQ (alfiah.putri.15@ucl.ac.uk)

Abstract

We propose a deep learning-based method to detect changes over the Martian surface, with a focus on the South Polar region. The method works by defining most of the images as "normal" and "anomalies" as candidates for changes.

1. Introduction

Although far from the Earth's surface, the Martian surface is not static. Some examples of changes are new impact craters, dust devil tracks, or dark slope streaks. With more than 40 years of orbital observations <100m, the amount of data available on Mars is enormous and too large to find changes manually. Because of this, we believe an automatic method will be useful for scientists to detect changes with a very large number of images.

Automatically detecting changes, especially over the polar region is a difficult problem if not done correct, as transient features and on-off changes are mixed with changes caused by seasons, one of which is the annual cycle of growth and recession of the polar cap, resulting in high number of false positive observations. Previously, Sidiropoulos and Muller [1] have developed a change detection algorithm that has successfully been tested with global Martian data. Their method was successfully trained on non-polar images and obtained reasonable results globally, although the method is far from perfect for the polar regions. Improvements can be made with more observed changes as training data, which we currently lack for Martian images.

2. Methods

To address the absence of reliable training data and to utilise the stack of overlapping data available, especially around the poles, we propose a deep-learning based method to detect anomalies on Martian images. One of the problems in creating a method based on deep learning is the computation and the amount of training samples needed to train the weights of the neurons in the network. Transfer learning is a method in machine learning in which networks which have been successfully trained to solve a specific problem are used to solve other similar problems.

In this research we started with AlexNet [2] Convolutional Neural Network (convnet). AlexNet has been trained to classify 1.2million ImageNet data and has been successfully used in planetary science research to classify features on HiRISE images, also by transfer learning [3]. In transfer learning for AlexNet, we take out the Fully Connected Network (FCN) Layers which works as an image classifier and replace them with a specific goal in mind.

In this research we are working with the assumption that most of the images don't change, and can act as "normal" data, while changes if they exist, are "anomalies". In this way, changes caused by "normal" processes, such as the appearance/ disappearance of ice cap or changes caused by differences in imaging condition are not picked up. To replace the FCN layers we used a OneClassSVM to detect "anomalies" from the other "normal" data.

We have isolated more than 20 regions with more than 30 overlapping ortho-rectified and coregistered (ACRO) [4] images over the south polar region [5], with most of them obtained from the Context Camera (CTX) as regions of interest. As inputs we use images co-registered and orthorectified to Digital Terrain Models from HRSC [6] scaled to neural network inputs. We randomly sampled 2.5km x 2.5km area with 100m overlap between samples, resized to input sizes required by AlexNet (227x227x3) instead of resizing pixel sizes to neural network inputs to ensure that the scales are similar between multi-instrument input images. This decision is made to increase the success rate for anomaly detection for multi-instrument image inputs.

Random image samples from the same region are used to define “normal” data for semi-unsupervised classifier. Distance obtained from OneClassSVM for testing data from the “normal” data are calculated and sorted. Carrying the assumption that most data are “normal” data, anomalies are then separated from normal data by adaptive binary thresholding to divide images into “normal” images and “anomalies”.

3. Results

We tested the method on several areas (areas overlapping B06_012028_0930_XN_87S273W and B02_010344_0985_XN_81S063W) from our regions of interest. Figure (1left) shows an example of an area which have been classified as “normal” by our network, with (1right) showing the same area in different date.

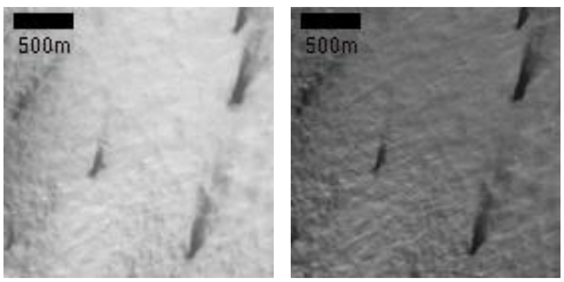

Figure 1(left) An example of detected “normal” data (P06_003206_0946_XI_85S277W, LS 212.06, MY 28) with 1(right) (P06_003562_0946_XI_85S276W, LS 229.25, MY 28) similar area in other image with different imaging condition

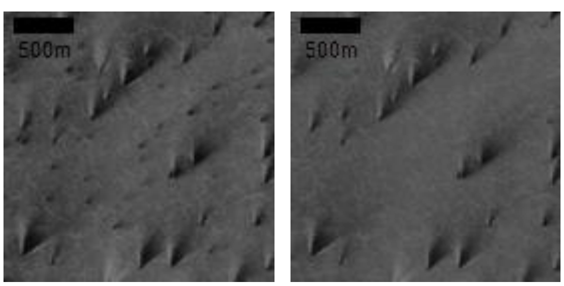

Figure 2(left) An example of detected “anomalies” data (P06_003206_0946_XI_85S277W, LS 212.06, MY 28) with 2(right) (P05_003074_0946_XI_85S277W, LS 205.81, MY 28) a similar area in another image

Figure (2) show an example of an area classified by our network as an “anomaly”, showing new seasonal fans appearing over the area.

4. Conclusions and Future Work

In this paper we have shown the potential of a semi-supervised deep learning method to do change detection research on Martian polar images by detecting anomalies over a region while ignoring expected appearance changes.

Currently we have only tested the method over specific areas out of our region of interest and only used ORIs and DTMs data from CTX. We are planning to test the method over the entire regions of interest for the south polar region. There are other data from different instruments (from MOC-NA until HiRISE) available to widen the dataset. Increase in accuracy and reduction of false positives can be obtained by building a more representative architecture for planetary/ Martian data as well as utilising available but yet unused information particular to Martian or polar data.

Acknowledgements

Part of the research leading to these results has received partial funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under iMars grant agreement n˚607379; MSSL STFC Consolidated grant no. ST/K000977/1 and the first author is supported by the Indonesian Endowment Fund for Education.

References

[1] Sidiropoulos & Muller: Automatic Detection of Changes on Mars Surface from High-Resolution Orbital Images. EGU Vol. 19. 2017. [2] Krizhevsky, et. al.: Imagenet classification with deep convolutional neural networks. Adv. in Neural Inf. Proc. Sys. 2012. [3] Wagstaff, et al.: Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas, 2018. [4] Sidiropoulos & Muller: A systematic solution to multi-instrument coregistration of high-resolution planetary images to an orthorectified baseline. IEEE Trans. on Geo and Remote Sens. 2018;56(1):78-92. [5] Putri, et. al.: Coregistration of Mars South Polar CTX Images to HRSC Base Images. EPSC Vol. 11. 2017. [6] Putri, et. al.: A New South Polar Digital Terrain Model of Mars from the High Resolution Stereo Camera (HRSC) onboard of the ESA Mars Express for Tracking Changes over the SPRC. PSS under-review, 2018.

EPSC Abstracts
Vol. 12, EPSC2018-829-1, 2018
European Planetary Science Congress 2018

Detection of early warning signals in paleoclimate data using a genetic time series segmentation algorithm

Athanasia Nikolaou (1,2), Pedro Antonio Gutiérrez (3), Antonio Durán Rosal (3), Francisco Fernández-Navarro (4), César Hervás-Martínez (3) and Maria Pérez-Ortiz (5)
(1) Institute for Planetary Research, German Aerospace Centre DLR, Berlin, Germany, (2) Centre of Astronomy and Astrophysics, Berlin Institute of Technology, Germany (3) Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain, (4) Department of Quantitative Methods, Universidad Loyola Andalucia, Seville, Spain (5) Computer Laboratory, University of Cambridge, United Kingdom (athanasia.nikolaou@dlr.de)

Abstract

We present a novel approach of analysing – visualising time series of a geophysical variable and we characterise its abrupt transitions in comparison to benchmark time series produced with model dynamical systems: a mathematical model (stochastic resonance) and a climate model of intermediate complexity (2D meridional ocean circulation with an atmospheric forcing) [1]. The method combines a genetic segmentation algorithm that uses ordinal regression and clusters the different segments of the time series around centroids located in a six-dimensional (6D) space of statistical metrics. After detecting statistical similarities it helps compare the type of transition observed in the time series to three separate studied types: a) noise transition, b) subcritical bifurcation crossing and c) transition to a limit cycle. The proposed method complements the causality analysis of a record of abrupt transitions in a geophysical system.

1. Introduction

The flexibility of the algorithms available in the machine learning range of methods should be combined with comprehensive physical systems in order to elucidate the exploration of a novel system. Here, we present a work [2] that uses the bio-inspired method of evolutionary algorithms in order to detect early warning signals (EWS) of an abrupt transition (tipping point) that is recorded in a geological time series. Statistical metrics have been used in the past [3] in order to detect EWS which we here extend to include: mean value, autocorrelation, standard deviation, kyrtosis, skewness and slope. An evolutionary algorithm is used to efficiently investigate this broad phase space. Prior knowledge of the Dansgaard-Oesger (DO) [4] transition points is not inserted into the algorithm.

2. Methods

Figure 1 shows an example "chromosome" for a sample data series. Six statistical metrics are calculated for each of the segments. The algorithm runs probabilistically, which means that it has random initialization of the segmentation and converges to a different segmentation pattern after each iteration until 5 clusters are formed in the 6D space. The results are given in percentage of encounter coherence over the total number of runs. If a segmentation pattern is encountered in all the runs, we assess that the detection of tipping point is 100% certain.

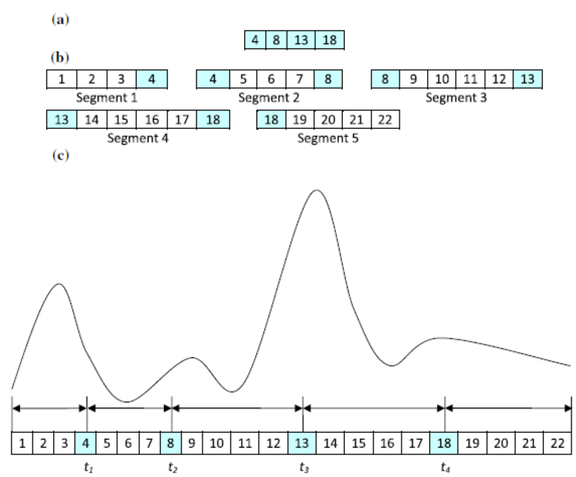

Figure 1: Chromosome representation of a sample time series composed of 22 data points. (a) Example chromosome. Each index corresponds to a position in the time series. (b) Segments that correspond to the chromosome. (c) The resulting segmentation of the time series. We obtain statistical characteristics for every segment. Image adopted with permission from [2].

3. Summary

The geophysical time series studied with this method are the ice cores NGRIP and GISP3 oxygen isotope data sets for the period spanning 50,000 yr before present (BP) until today. As a preliminary result, it is suggested that the DO events do not share the same classification and could be potentially attributed to different underlying dynamics. Further work has been conducted [5] in order to define a decision tree for the classification of the tipping points.

Acknowledgements

This work has been performed within the Ariadna project 13-9202 of the European Space Agency. A.N. acknowledges the contribution of Isabelle Dicaire and Sandro Calmanti to the study, and the project VH-NG-1017 of the Helmholtz association for enabling participation in the EPSC 2018 conference. The research work of P. A. Gutiérrez, A. Durán, and C. Hervás-Martínez is partially funded by the TIN2011-22794 project of the Spanish Ministerial Commission of Science and Technology (MICYT), FEDER funds and the P11-TIC-7508 project of the "Junta de Andalucia" (Spain).

References

[1] Artale V, Calmanti S, Sutera A, Thermohaline circulation sensitivity to intermediate-level anomalies, Tellus A, Vol. 54(2), pp. 159–174, 2002

[2] Nikolaou A, Gutierrez P A, Duran A, Dicaire I, Fernandez-Navarro F, Hervas-Martinez C: Detection of early warning signals in paleoclimate data using a genetic time series segmentation algorithm, Climate Dynamics, Vol. 44 , pp. 1919–1933, 2015.

[3] Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, Kefi S, Livina V, Seekell DA, Van Nes EH: Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PPLoS ONE 7(7): e41010, 2012

[4] Cimatoribus AA, Drijfhout SS, Livina V, van der Schrier G: Dansgaard-Oeschger events: bifurcation points in the climate system, Clim Past, Vol. 9, pp. 323–333, 2013

[5] Perez-Ortiz M., Duran-Rosal A.. Gutierrez P. A., Sanchez-Monedero J., Nikolaou A., Fernandez-Navarro F., Hervas-Martinez C.: On the use of evolutionary time series analysis for segmenting and predicting paleoclimate data, Neurocomputing, 2017.

EPSC Abstracts
Vol. 12, EPSC2018-844, 2018
European Planetary Science Congress 2018

Advanced Techniques for Signal Search and Automatic Classification of Observational Space Data

Al-Ubaidi Tarek (1), Khodachenko Maxim (1), Kern Roman (2), Granitzer Michael (3), Poedts Stefaan (4)
(1) Space Research Institute, Graz, Austria, (2) Know-Center, Graz University of Technology, Austria, (3) University of Passau, Germany, (4) Catholic University of Leuven, Belgium

Abstract

The presentation will outline various approaches in *machine learning* and *content based search* investigated by members of the former *IMPEx-FP7* (http://impex-fp7.oeaw.ac.at/) project consortium, in close cooperation with partners *Know-Center*, *Graz University of Technology*, and *University of Passau* and discuss some of the numerous possibilities that open up, using these or equivalent techniques in the emerging field of ***e-Science* in conjunction with space science**. In particular, the presentation will focus on applications that allow systems to **automatically classify and pre-select scientific data** and hence speed up scientific workflows significantly by supporting scientists with the cumbersome task of going through vast amounts of data manually, looking for specific patterns, signals and phenomena of interest prior to selecting specific data for closer examination and analysis.

Introduction

Due to extensive research in the field of information retrieval, search technologies are commonplace today and their algorithmic underpinnings are well understood and proven to scale up to massive amounts of data. While the capability to do complex searches for specific signals and phenomena would come in quite handy when analyzing heterogeneous scientific data, such methods are mostly limited to textual searches and thus will not (directly) apply to cases where the data at hand is time series from sensory data. This is the case for e.g. observational space data or data derived from simulations of various physical processes. In such instances the handling and processing of the data needs to be adapted first and the search paradigm needs to be redefined, since the actual search cannot directly be initiated by key terms entered by the user.

Powerful Tools for Science

A promising solution investigated by the team is a technique known as **query by example** or **content-based search** in the information retrieval community. Data is first transformed into a representation suitable to be managed using an inverted index by adopting and extending *symbolic representation techniques* which are designed to transform continuous data (discrete in the time domain) into a discrete or quantized representation, while keeping the associated information loss minimal (also see *wavelet* analysis). Further approaches from the field of *unsupervised machine learning* can then be applied to obtain temporal patterns of interest and identify a trade-off between frequent and surprising patterns. However, due to the unsupervised nature of the used techniques, the discovered patterns will not be tailored towards specific cases of scientific interest. In order to further limit the identified patterns to a set of candidates that are relevant in the context of specific questions in the realm of space science, techniques from the field of **supervised machine learning** can be established, using a procedure where human annotators provide labelled examples as references. Using these information retrieval and *machine learning* (as well as *deep learning*) methods a system can be built that automatically searches for specific phenomena in large quantities of (observational) data and most importantly also performs automatic classification of scientific data.

Content Based Search and Automatic Classification

Following a brief depiction of an example workflow for a simple application of the techniques outlined above, where a specific signal of interest is selected and then used as input for a comprehensive *content based search*. As a first step the user selects a portion of the time series data - a graphical interface allowing visual selection (and annotation) of patterns of interest is shown in **Figure 1**.

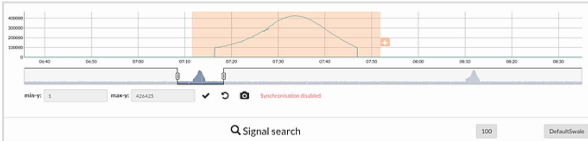

Figure 1*: The user provides a signal by manually selecting a small portion of data*

In this step the user could also provide detailed annotations that specify the signal and allow further enhancing the search capabilities. The system then responds with a ranked list of search results, i.e. signals in the investigated data that resemble the example given above with descending similarity. It should be noted here that the search technology used, scales up to hundreds of Gigabytes of data and beyond. See **Figure 2Error! Reference source not found.** for an example of a possible response generated by the system, given the input selected (**Figure 1**).

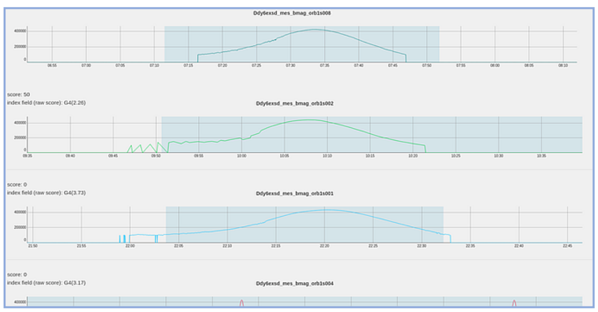

Figure 2*: The ranked list of signal search results*

The first response in particular shows almost identical characteristics as the input signal and is likely to originate from a similar physical environment.

Summary and Conclusions

With new missions leveraging up-to-data capabilities in *telematics* and thus producing ever increasing amounts of observational data, *content based search*, *machine learning* and related technologies can provide a powerful toolset to enhance data analysis and data driven investigations of any kind. Many time consuming tasks can already be sped up and in many instances improved by leveraging current approaches in *machine learning* and *artificial intelligence*. **Now is the time to start building prototype tools** and to carefully analyze scientific workflows in order to gather detailed and relevant requirements for the *e-Science* tools of the future. In this regard, it is crucial that experts in IT and machine learning are closely cooperating with (space) scientists, in order to gain a deep understanding of the problems at hand and to be able to build powerful solutions that will optimally support space science in the 21st century.

Acknowledgements

We would like to thank our colleagues at the ***Catholic University of Leuven***, the ***Skobeltsyn Institute of Nuclear Physics*** and the ***Ministry of Education and Science*** of the Russian Federation (Grant RFMEFI61617X0084) for their support and cooperation and the provision of simulation data as well as observational data that were crucial to build first prototypes and conduct tests of actual *real life* applications of the discussed technologies.

EPSC Abstracts
Vol. 12, EPSC2018-1068, 2018
European Planetary Science Congress 2018

Machine Learning Classification of Simulated Mars Observations

Justin Erwin, Loïc Trompet, Ann C Vandaele, Séverine Robert, Ian Thomas, Frank Daerden, Lori Neary, Sébastien Viscardy
Belgian Institute for Space Aeronomy (jt.erwin@aeronomie.be)

Abstract

In this fortuitous time in space exploration, by having multiple mission collecting a plethora of observations of terrestrial atmospheres we are burdened with the difficult task of extracting useful interpretations of the results. In this work we focus on column/number densities from spectral retrievals of Mars, and wish to find trends due the geography (i.e. Latitude, Longitude, Ls, local time, altitude, surface height) and external factors (i.e. solar activity), as well as correlation between the species. Considering the number of variables and large sample size, this problem is well suited for classification via machine learning. To explore the possible methods, we construct a simulated collection of retrievals using realistic orbit and observation parameters, along with interpolated profiles from the GEM-Mars GCM. We compare the different machine learning methods by their performance and their effectiveness in gaining insight into the trends and correlation.

1. Introduction

GEM-Mars is a General Circulation Model for the atmosphere of Mars with online atmospheric chemistry. The model is operated on a grid with a horizontal resolution of 4°x4° and with 103 vertical levels reaching from the surface to ~150 km. It calculates atmospheric heating and cooling rates by solar and IR radiation through atmospheric CO and dust and ice particles and solves the primitive equations of atmospheric dynamics. Geophysical boundary conditions are taken from observations. Physical parameterizations in the model include an interactive condensation/surface pressure cycle, a fully interactive water cycle including cloud radiative feedbacks, a thermal soil model including subsurface ice, interactive dust lifting schemes for saltation and dust devils, turbulent transport in the atmospheric surface layer and convective transport inside the planetary boundary layer (PBL), subgrid scale vertical mixing in the free troposphere, a low level blocking scheme, gravity wave drag, molecular diffusion, non-condensable gas enrichment, and atmospheric chemistry. A detailed description of the model, its formulation, grid, dynamical core and physical parameterizations, together with extensive validation against multiple datasets, was given in [1], and further details can be found in [2, 3, 4].

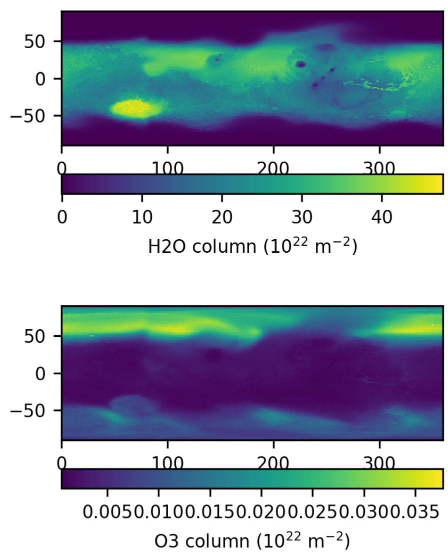

Figure 1: Simulated column densities for H2O and O3 from the GEM-Mars GCM (Ls=0, LST=15h).

In Figure 1, we present column densities for water and ozone from the GEM-Mars GCM that have been interpolated onto a high-resolution global grid. These images suggest an anti-correlation between water and ozone, which can be attributed to the chemistry module included in the GCM. We intend to classify this and other trends in the data using the techniques of Machine Learning.

2. Machine Learning Application

The variables in our data set include the geographic knowledge of the observation, possibly values for the solar activity, and column densities for one or many molecules and aerosols. Different observations necessarily include a small subset of the molecules in the Martian atmosphere (due to the order selection and sensitivity of the measurements), so our data set will have sparse coverage of the dependent variables.

To make use of most Machine Learning techniques, we first must normalize the data [5] (e.g. column densities). Due to the varied relative abundances of the species, we likely will need a per species normalization. We may also want to introduce additional features derived from the geographic variables that can sensibly be used along with our data (e.g. use Mars' orbital distance to adapt to solar insolation, or use surface height to adapt to large changes in surface pressure).

Next, we will explore the set of methods available in Scikit-learn [6, 7], a Python implementation of many machine learning algorithms. We will apply clustering on the data to search for clear separation in our dataset, and iteratively selecting new features derived from the dependent variable. Additionally, we will evaluate the effectiveness of support vector machines and principle component analysis in classifying the data, as well as the automated feature selection techniques available in Scikit-learn.

3. Summary and Conclusions

From previous, current and future spacecraft mission, we have an enormous supply of data which we would like to classify and search for trends and interesting features. We can then focus on these feature to see if they are supported by models or can highlight important physical processes. In this work, we create a simulated data set of spacecraft retrievals using the GEM-Mars GCM to determine the usefulness of Machine Learning for this purpose, and gain insight before applying the techniques to future data sets.

Acknowledgements

This project acknowledges funding by the Belgian Science Policy Office (BELSPO), with the financial and contractual coordination by the ESA Prodex Office (PEA 4000121493), and by the FNRS CRAMIC project under grant number T.0171.16. The research was performed as part of the "Excellence of Science" project "Evolution and Tracers of Habitability on Mars and the Earth" (FNRS 30442502).

References

[1] Neary, L., and F. Daerden (2018), The GEM-Mars General Circulation Model for Mars: Description and Evaluation, Icarus, 300, 458-476, https://doi.org/10.1016/j.icarus.2017.09.028

[2] Daerden, F., J. A. Whiteway, L. Neary, L. Komguem, M. T. Lemmon, N. G. Heavens, B. A. Cantor, E. Hébrard, and M. D. Smith (2015), A solar escalator on Mars: Self-lifting of dust layers by radiative heating. Geophys. Res. Lett., 42, 73197326.doi:10.1002/2015GL064892

[3] Viscardy, S., F. Daerden, and L. Neary (2016), Formation of layers of methane in the atmosphere of Mars after surface release, Geophys. Res. Lett., 43, 5, 1868-1875, doi:10.1002/2015GL067443.

[4] Smith, M., F. Daerden, L. Neary and S. Khayat (2018), The climatology of carbon monoxide and water vapor on Mars as observed by CRISM and modeled by the GEM-Mars general circulation model, Icarus, 301, 117-131, https://doi.org/10.1016/j.icarus.2017.09.027

[5] Müller, A. C., Gutdo, S. (2017), Introduction to Machine Learning with Python, O'Reilly Media Inc., Sebastopol, CA, USA.

[6] Pedregosa et al. (2011), Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research (12), pp. 2825-2830.

[7] Géron, A. (2017), Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly Media Inc., Sebastopol, CA, USA.

EPSC Abstracts
Vol. 12, EPSC2018-1089-2, 2018
European Planetary Science Congress 2018

Automatic Detection of Sub-km Craters on Mars for Equilibrium Population Statistics

A. M. Francis, J-P. Muller, P. Sidiropoulos and D. M. Persaud

Imaging Group, Mullard Space Science Laboratory, UCL

Abstract

Small (sub-km) crater size-frequency distributions are the standard metric for dating very young surfaces on the Martian surface, because of the lack of large, infrequent impact events and the unavailability of surface samples. However, small crater population statistics are poorly understood and make accurate absolute dating of young surfaces impossible. This is because several unknown factors which affect the crater production and erosion rates – such as atmospheric filtering, secondary cratering and partial resurfacing [1]. Constraining these factors, where possible, is important if we are to understand the recent history of the Martian surface. We present an algorithm capable of detecting small crater candidates in high-resolution visible imagery of the Martian surface. The algorithm classifies craters with a state-of-the-art F1-score (91%) when compared with other algorithms on the same dataset [2-4]. We use this alongside a mean-shift clustering algorithm to detect crater candidates in an extended HRSC image with near 100% recall and roughly 50% precision. The candidates can then be marked rapidly by a human expert, greatly increasing the speed of small crater counting exercises, when compared to traditional manual marking. The detection algorithm's performance is shown in both familiar (relative to the training set) and unfamiliar terrain, which we believe demonstrates that it is a viable tool for accurate and quick crater counting on Mars.

1. Introduction

Historically, CSFD's have been constructed manually by human experts [5]. We believe this is primarily due to two reasons: 1) human experts are thought to be the most accurate crater detector, given that we have no higher authority by which to check our answers; 2) Large craters have been shown to be of far more immediate use in age-dating, and are more easily countable by humans because there are many fewer of them than sub-km ones.

Small crater statistics are not well understood. This is because of various poorly constrained stochastic processes that effect both the production and erosion of small craters [1]. These small craters reach an equilibrium population distribution quickly, and therefore many surfaces have a stable number of small craters which cannot inform us of the surface age. With substantial amounts of data, the processes effecting production and erosion may be able to be isolated in these equilibrium populations, however a very large count of small Martian craters has never been conducted.

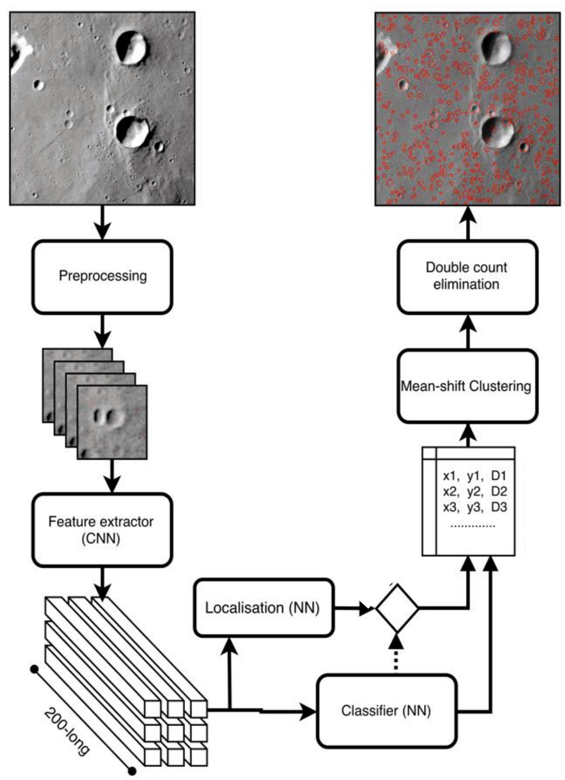

Figure 1: A flowchart of the algorithm, showing feature extraction, classification and clustering.

2. Method

Our algorithm comprises three distinct stages (see Figure 1). First, image patches are transformed by a 4-layer convolutional neural network into a set of features. Secondly, these features are used to classify the image patch as a crater or non-crater, by a neural network. A second neural network is then used on the crater candidates to estimate position and size within the image. Finally, many detections of the same crater in the extended scene are clustered using the mean-shift algorithm.

The convolutional network is initially trained in an unsupervised fashion, using an autoencoder architecture. The training data used is random patches of Martian terrain imagery from HRSC nd-4 products. After the unsupervised learning, both the convolutional network and the neural networks are trained using a dataset made available by Cohen et al. (link) in the Nanedi Valles region. We extend this dataset with additions from different terrain, and use data augmentation to increase the number of training examples.

3. Results

Our algorithm performs at the state-of-the-art when compared to other methods [2],[3],[4] using the same dataset. We perform with a 91% F1-score in a classification scenario, which will improve with more training data (Figure 2). In a detection scenario across an extended scene, the algorithm can be used to obtain crater candidates for expert marking. In this mode, the detection algorithm has a recall at or near to 100% and a precision of around 50%. This leads to a huge decrease in the time spent manually counting craters, given that errors of omission (the most time-consuming to correct) are negligible. Our detection algorithm shows robustness to a variety of terrain types, with reliable performance in areas that aren't represented in the training set. Using this tool, we aim to produce a large catalogue of small Martian craters, which will be used to constrain the effects of secondary cratering, erosion rates and partial resurfacing.

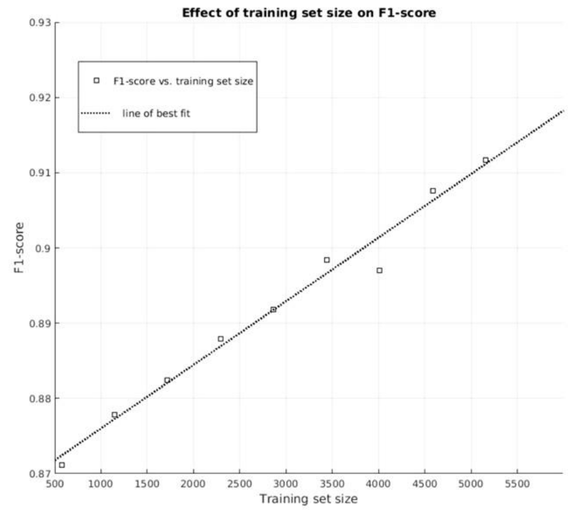

Figure 2: The classification performance (F1-score) of the algorithm, using different amounts of the available training set data. This is a clear indication that more data will increase performance.

4. Acknowledgements

Part of the research leading to these results has received partial funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under iMars grant agreement n˚607379; MSSL STFC Consolidated grant no. ST/K000977/1 and a STFC studentship for the first author.

References

[1] Williams, J., Bogert, C. H., Pathare, A. V., Michael, G. G., Kirchoff, M. R. and Hiesinger, H. (2018).. Meteorit Planet Schi, 53: pp. 554-582.

[2] Cohen, J.P., Lo, H.Z., Lu, T., Ding, W. (2016). Proceedings of the International Conference of Pattern Recognition, pp. 3642-3649

[3] Urbach, E.R. and Stepinski, T.F. (2009). Planetary and Space Science, vol. 57, no. 7.

[4] Ding, W. et al., (2011). ACM Transactions on Intelligent Systems and Technology, 2(4), pp.1–22.

[5] Hartmann, W.K. & Neukum, G., (2001). Space Science Reviews, 96, pp.165–194.