europlanet 2024 Research Infrastructure

H2020-INFRAIA-2019-1

Europlanet 2024 RI has received funding from the
European Union's Horizon 2020 Research and Innovation Programme under

Grant agreement no: 871149

Deliverable D10.9

Deliverable Title:	ML 4th Year Report
Due date of deliverable:	31st July 2023
Nature[1]:	R
Dissemination level[2]:	PU
Work package:	WP10
Lead beneficiary:	INAF
Contributing beneficiaries:	IWF-OeAW, KNOW, UNIPASSAU, ACRI-ST, IAP-CAS, AOP
Document status:	Final

Start date of project:	01 February 2020
Project Duration:	54 months
Co-ordinator:	Prof Nigel Mason

1. **Nature:** R = Report, P = Prototype, D = Demonstrator, O = Other
2. **Dissemination level:**

PU	PP	RE	CO
Public	Restricted to other programme participants (including the Commission Service)	Restricted to a group specified by the consortium (including the Commission Services)	Confidential, only for members of the consortium (excluding the Commission Services)

Executive Summary / Abstract:

This final report summarizes the work done in Work Package 10 'Machine Learning Solutions for Data Analysis and Exploitation in Planetary Sciences' during the four- year period of the Europlanet 2024 Research Infrastructure. The main aims of the work package are to foster wider use of machine learning technologies in data-driven space research and to provide open-source machine learning code developed for specific science cases. Work Package 10 is organized around six tasks that target management and coordination of the activities, the development of machine learning-based data analysis code and the dissemination of the tools, as well as integration of the results into VESPA, GMAP and SPIDER where appropriate. Despite delays in the development work due to the ongoing COVID-19 pandemic, work on all six tasks has been progressing. All proposed science cases (6 +1 extra in total) are completed and were uploaded to the dedicated repository: https://github.com/epn-ml/. The description of the scientific cases can be found at https://ml-portal.oeaw.ac.at/. A number of workshops have been conducted during the last three and a half years, mainly at the EPSC conferences, where the machine learning pipelines were presented. The tutorials of the cases have been implemented and uploaded on our Machine Learning Portal as well as on our public GitHub repository. A Jupyter book has been initiated with the ultimate goal of gathering all content developed in our work package. WP10 presented the achieved results through the three-year period at various scientific meetings. Some of them are listed here:

- In 2021, we conducted two workshops at EPSC 2021, introducing two of our machine learning pipelines. We put up first tutorials on our Machine Learning Portal as well as on our public GitHub repositories. ML organized machine learning sessions at EGU21 and EPSC2021, and had presentations at many conferences (LPSC2021, EGU21, EPSC2021, ESWW 2021, AGU Fall Meeting 2021). We started collaborations with national (FWF project at IWF) as well as international (EU Horizon 2020 EXPLORE project) research projects and started a series of fireballs workshops together with NA2. An EPN-TAP server was set up at the IWF, on which we started to integrate first data sets of our science cases into VESPA. Further, first steps were taken to include our pipelines in SPIDER.
- In 2022, ML organized machine learning sessions at EGU22 and EPSC2022, and had presentations at many conferences (EGU22, EPSC2022, ECML PKDD 2022, ESWW 2022, AGU Fall Meeting 2022). We have collaborations with national (FWF project at IWF) as well as international (EU Horizon 2020 EXPLORE project) research projects and started a series of fireballs workshops together with NA2. An EPN-TAP server was set up at the IWF, on which we started to integrate first data sets of our science cases into VESPA. Furthermore, next steps were undertaken to include our pipelines in SPIDER.

Since 2020, the project has developed ML tools to handle complex planetary data more efficiently and provide opportunities to combine and visualise multiple diverse datasets. This programme has been further enhanced through a collaboration with a second Horizon 2020 project, EXPLORE, which is developing applications for the exploitation of galactic, stellar and lunar data, and provides a platform for deploying and testing ML tools and services.

Table of Contents

Table of Abbreviation	
AGU	American Geophysical Union
ASPP	Atrous Spatial Pyramidal Pooling
BS	Bow Shock
CIR	Corotating Interaction Region
CME	Coronal Mass Ejection
CNN	Convolutional Neural Network
D	Deliverable
DMP	Data Management Plan
DTM	Digital Terrain Model
EGI	European Grid Infrastructure
EGU	European Geophysical Union
EOSC	European Open Science Cloud
EPN-2024-RI	Europlanet 2024 Research Infrastructure
EPSC	Europlanet Science Congress
ESA	European Space Agency
ESWW	European Space Weather Week
GAN	Generative Adversarial Network
GMAP	Geologic MApping of Planetary bodies
GPU	Graphics Processing Unit
HCS	Heliospheric Current Sheet
ICA	Independent Component Analysis
ICME	Interplanetary Coronal Mass Ejection
IMF	Interplanetary Magnetic Field
IoU	Intersection over Union
JpGU	Japan Geoscience Union
JRA	Joint Research Activity
LPSC	Lunar and Planetary Science Conference
LSTM	Long-Short-Term Memory Network
MASCS	Mercury Atmospheric and Surface Composition Spectrometer
MESSENGER	MErcury Surface, Space ENvironment, GEochemistry, and Ranging

ML	Machine Learning
MP	Magnetopause
MS	Milestone
NA	Networking Activity
PMC	Project Management Committee
SDA	Scientific Data Application
SPIDER	Sun Planet Interactions Digital Environment on Request
TRL	Technology Readiness Level
UMAP	Uniform Manifold Approximation and Projection
VA	Virtual Access
VESPA	Virtual European Solar and Planetary Access
WP	Work Package

1. Explanation of WP10 Work & Overview of Progress

a. Objectives and Description of Work

Traditionally, science starts with a hypothesis. We develop a theory that we test experimentally, producing data. We analyse the data and – hopefully – the process results in new knowledge.

The advent of ML has enabled a new approach, known as data-driven science. Using the wealth of datasets and streams available, ML can explore the data to find a pattern or commonality. Out of these initial steps comes a hypothesis that can be tested through data analysis, which again – hopefully – leads to new understanding. Clustering or fusing datasets, moreover, can reveal connections or knowledge that are not recognizable in the individual datasets.

All ML tools of this WP are open-source, ready-to-use, and highly customisable, enabling other researchers to freely deploy and adapt them for their own research scenarios.

The objectives and description of work for Work Package (WP) 10 'JRA4 ML - Machine Learning Solutions for Data Analysis and Exploitation in Planetary Sciences' are as follows, quoted from the proposal:

> JRA4 will develop Machine Learning (ML) powered data analysis and exploitation tools optimised for planetary science and integrate expert knowledge on ML into the planetary community. All tools can also be linked in a future project via the VA services of VESPA, GMAP and SPIDER (where appropriate).
>
> The main objectives are:
>
> - to develop ML tools, designed for and tested on planetary science cases submitted by the community, and to provide sustainable, open access to the resulting products, together with support documentation

- to foster wider use of ML technologies in data driven space research, demonstrate ML capabilities and generate a wider discussion on further possible applications of ML
- to identify scientific and commercial applications for the ML tools developed through the JRA tasks.

Description of work

This JRA was led by IWF-OEAW till 1 October 2022, co-led by the KNOW Centre. Since October it has been led by INAF and co-led by the KNOW centre. ML-powered data analysis and exploitation tools were developed targeting a set of representative scientific cases selected from about a dozen proposals for specific applications of ML in planetary science submitted by the scientific user community during proposal preparation. Software developed during the JRA is open source (Apache License 2.0), thoroughly documented and available via a Git service, so that all results can be used freely, and further developed and extended by the community.

A public GitHub account was set up (https://github.com/epn-ml). In the public GitHub repositories, we will place all the documents and files that are suitable to be made public, e.g., descriptions of the science cases, links to publications and other relevant information, final data sets, working code scripts, and presentations.

A website, our so-called ML Portal (ml-portal.oeaw.ac.at), was installed and serves as an access point to our activity. In the portal, among other content, we will provide:
- An introduction to the ML activity within Europlanet 2024 RI
- General information about ML
- ML tutorials
- Python Jupyter notebooks with different ML tools
- Downloadable ML tools and/or links to them, i.e., Python scripts
- Tutorials on how to use the tools and how to modify them for specific needs
- A list of presentations and publications with results of the ML activity
- Announcements of upcoming events (workshops, sessions at conferences, etc.).

The ML-powered tools focus on three types of planetary data: time-series data, where data mining can reveal the dynamical evolution of phenomena or time dependent events; imagery, where training algorithms to recognise features can support automated mapping and classification of common characteristics; and other kinds of data, for example spectral data, where characteristics like composition or surface ages can be identified.

Work Package Beneficiaries

Apart from the WP lead, IWF-OEAW, there are eight beneficiaries contributing to our WP. Table 1 lists the acronyms of the WP beneficiaries as used in the Europlanet 2024 Research Infrastructure (EPN2024-RI) proposal and their corresponding institutions.

Due to sanctions against Russia, the participation of LMSU has been terminated. One science case about the automatic detection of boundary crossings around the planet Mercury was proposed by LMSU. Since this science case was finished before the sanctions against Russia were installed, the termination of LMSU's participation in the Europlanet 2024 Research Infrastructure does not have any effect on WP10.

Table 1. Work package beneficiaries.	
Work Package Beneficiaries	
ACRI-ST	ACRI-ST, France
AOP	Armagh Observatory and Planetarium, Ireland
DLR	Deutsches Zentrum für Luft- und Raumfahrt, Germany
KNOW	Know-Center GmbH, Austria
IAP-CAS	Institute of Atmospheric Physics, Academy of Sciences of Czech Republic, Czech Republic
INAF	National Institute for Astrophysics, Italy
IWF-OEAW	Space Research Institute, Austrian Academy of Sciences, Austria
LMSU	M.V. Lomonosov Moscow State University, Russia
UNIPASSAU	University of Passau, Germany

Science Cases

The science cases proposed by the planetary science community during proposal preparation are listed in Table 2. The proposal by GMAP covers different cases dealing with the detection and classification of various planetary surface features, such as mounds and pits. In Figure 1 we have clustered the scientific cases according to the main type of data. The two cases listed in the cluster 'Other' use spectral data, which basically do not fall in one of the other two categories, but both may influence each other, and both may benefit from the codes developed for the other cases. The idea behind the clustering is that the science cases within one cluster can be tackled with similar approaches. Thus, the codes/tools developed for one of the cases can be used with (small) modifications for the other cases in the same cluster.

Table 2: list of science cases

Proposer	Science Case
IAP-CAS	Detection of plasma boundary crossings at planetary magnetospheres and solar wind
	Classification of plasma wave emissions in electromagnetic spectra
INAF	Mineral identification via reflectance spectra [possible applications foreseen in GMAP]
DLR	Classification of surface composition on the surface of Mercury [resulting data products can be used for GMAP]
AOP	Abundance of asteroids in Earth-like orbits from STEREO images
GMAP	Automatic recognition and analysis of planetary surface features
IWF-OEAW	Detection and classification of CMEs and CIRs in in-situ solar wind data
LMSU	Search for magnetopause/shockwave crossings on Mercury based on MESSENGER data

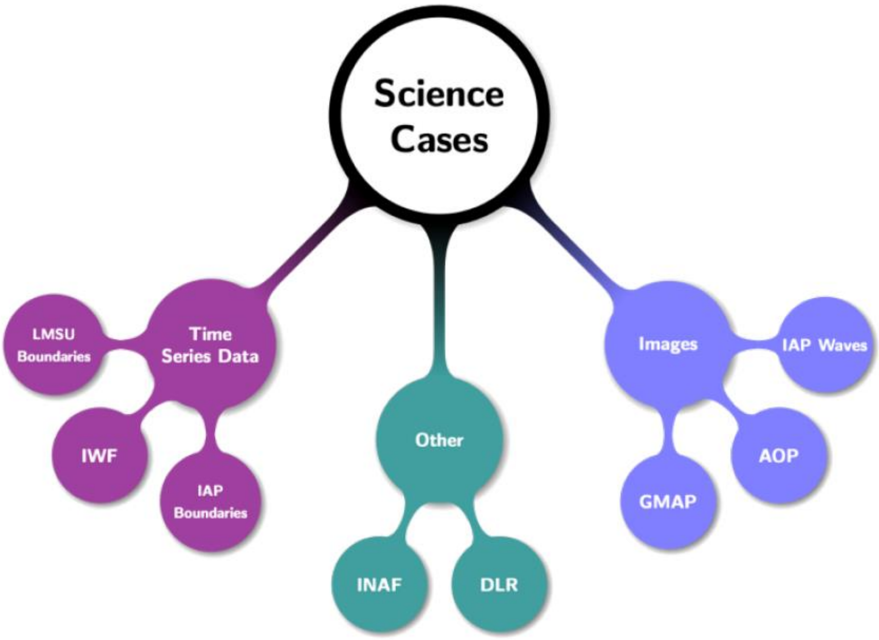

Figure 1. Clusters of science cases according to the corresponding data used

The science case by AOP needed to be re-formulated, since it was not do-able the way it was proposed. More details about the new science case can be found later in the according section.

Deliverables and Milestones

There are nine deliverables and three milestones for our WP, listed in Table 3.

Table 3: list of deliverables (D) and milestones (MS)

Abbreviations	Description	Month due	Finished
D10.1	Annual Report 1	M12	✓
D10.2	Annual Report 2	M24	✓
D10.3	Tutorial on Machine Learning and Basic How Tos (initial release)	M31	✓
D10.4	Demonstrator and Documentation of Data-Processing Techniques	M42	✓
D10.5	Demonstrator and Documentation of Time-based Signal Analysis and Automatic Classification Tool	M42	✓
D10.6	Demonstrator and Documentation of General Classification Toolset	M42	✓
D10.7	Annual Report 3	M36	✓

D10.8	Tutorial on Machine Learning and Basic How Tos (final release)	M42	✓
D10.9	Annual Report 4	M48	✓
MS11	Requirements for ML tools documented	M4	✓
MS51	ML Demonstrators implemented and tested	M30	✓
MS86	ML Demonstrators fully validated and integrated	M42	

b. Explanation of the work carried in WP

Task 1 - Management and Coordination

This task oversees the management of the ML JRA4, coordinates the activities within the WP and with the other WPs and reports to the PMC.

Task 2 - Requirements for Machine Learning, Tool Validation and Communication

Infrastructure
All information about our tools and implemented scientific cases are available on the ML Portal and our GitHub repository.It includes more information about the science cases, presentations, news regarding ML conferences, sessions and tutorials.

We put the draft version of a Jupyter book on our GitHub repository (https://github.com/epn-ml/europlanet-ml-book), which serves as a tutorial and reference book for the activity in our work package.

Presentations and Workshops
Fireball-tracking networks around the world are assisting in the recovery of fragments of fresh meteorites and understanding where in the solar system they originated. In collaboration with NA2, the ML WP organised the second and third workshops in this series of four, which were held on 4-5 February 2022 (virtual) and on 13-14 August 2022 (hybrid). These workshops bring together observers from different fireball networks, along with ML experts, to discuss how ML can support the fireballs community and to advise on handling the data collected.

Four ML pipelines have been presented in three workshops during EPSC2022 - the pipeline for the IAP-CAS boundaries science case, the pipeline for the GMAP mounds science case, as well as two pipelines for the GMAP pits science case. All of the pipelines are available on GitHub.

Presentations with results of the science cases are mentioned in the section about the individual science cases.

Collaborations
We continued our collaboration with two research projects at the IWF.

Further, we continued our collaboration with the EU Horizon 2020 project EXPLORE and we are further investigating the possibility to integrate our ML pipelines into the EXPLORE platform.

Task 3 - Data Pre-Processing, ETL and Feature Engineering

The aspects of data pre-processing and feature engineering are covered in the descriptions of the work for the individual science cases. Most science cases thereby utilize standard pre-processing methods or work on the raw data through end-to-end learning. However, we also explore new routes to automate pre-processing. For example, the GMAP Mounds science case utilizes data augmentation in the form of generative adversarial networks to overcome data sparsity. Details on the pre-processing conducted can be found below.

Task 4 - Time-based Signal Analysis and Automatic Classification

Tools for Time-Dependent Phenomena

The Sun emits not only heat and light, but a stream of electrically charged particles. This 'solar wind' interacts with objects in its path and can potentially strip away planetary atmospheres. Earth, and other planets with a global magnetic field, are largely shielded from the effects of the solar wind. However, solar activity can result in flares, emission of solar energetic particles and eruptions of material, called coronal mass ejections, that can interact with Earth's magnetic environment and, in severe cases, cause serious disruption to power grids, radio networks and satellites.

Europlanet 2024 RI has developed a suite of ML tools to support investigations of the solar wind and its effects on planetary environments over time. One tool supports forecasting the severity of a solar storm based on its magnetic orientation compared to the Earth's magnetic field. A second tool monitors the conditions controlling emissions by high-energy particles trapped in radiation belts. A third tool automatically identifies points in data streams when an orbiting spacecraft crosses over the boundary between a planet's protective magnetic field and the unshielded conditions of the solar wind. Collectively, this deployment of ML enhances our understanding of solar wind interactions and our ability to protect infrastructure both here on Earth and on the surface of, or in orbit around, other planets.

IWF ICME Science Case (Automatic Detection and Classification of Boundary Crossings in Spacecraft in situ Data)

Planetary magnetospheres create multiple sharp boundaries, such as the bow shock, where the solar wind plasma is decelerated and compressed, or the magnetopause, a transition between solar wind field and planetary field.

Interplanetary coronal mass ejections (ICMEs) are one of the main drivers for space weather disturbances. In the past, different machine learning approaches have been used to automatically detect events in existing time series resulting from solar wind in situ data. However, classification, early detection and ultimately forecasting still remain challenging when faced with the large amount of data from different instruments. While CNNs are often used to discover objects or patterns in images or data series, there are two main problems

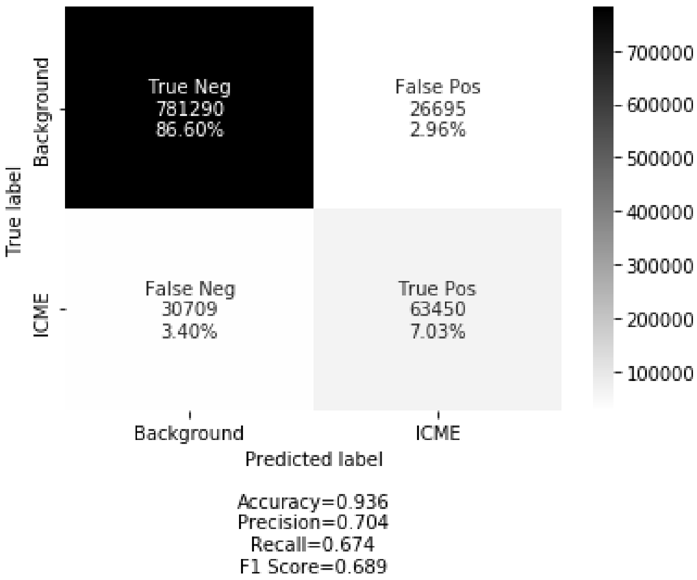

Figure 2: Timeline for the science cases. Also shown are the deadlines for the WP deliverables and milestones.

when facing our specific task: high duration variability and a rather ambiguous definition of start and end time.

After the reimplementation of a model proposed by Nguyen et al. (2019) in year 1 of this WP, the model was tested on STEREO-A and STEREO-B data as well as on WIND data. All three contain less variables than the original data set used by Nguyen et al. At a similar recall as for the original set, the precision for all three datasets was only around 30% and the accuracy in delivering start and end times was limited.

The next step was to align all three data sets in order to process more training data for a combined model. It was tested on held out datasets for WIND, STEREO-A and STEREO-B. Surprisingly, this did not sufficiently improve performance and lead us to explore other approaches.

Starting from the reimplementation a post processing step based on YOLO v5 (ultralytics) was investigated, in order to improve performance. Even though first results seemed promising, the idea was later discarded due to unsatisfactory results and the laborious pipeline. Since the ultimate goal is an explicit and widely applicable pipeline, it was decided to abandon the general approach of using multiple basic neural networks and the similarity measure used by Nguyen et al. (2019) completely and compose it as a segmentation problem instead.

We proposed a pipeline using a UNet (Ronneberger et al., 2015) including residual blocks, squeeze and excitation blocks, Atrous Spatial Pyramidal Pooling (ASPP) and attention blocks, similar to the ResUNet++ (Jha et al., 2019), for the automatic detection of ICMEs. Comparing it to last year's results, we find that our model outperforms the baseline regarding GPU usage, training time and robustness to missing features, thus making it more usable for other data sets, as well as the three aligned data sets. The confusion matrix is shown in Figure 2.

The relatively fast training allows straightforward tuning of hyperparameters. Our proposed pipeline can be used for any time series segmentation problem. The straightforward implementation allows a simple extension to a multiclass classification problem and paves the way to include corotating interaction regions into the range of detectable phenomena within our pipeline. Furthermore, we hope to apply our model to similar problems in the future.

There have been several quite successful attempts to automatically segment in situ time series. Labelling the different regions such as the magnetosphere, the magnetosheath and the background solar wind, the segmented maps were subsequently used to detect boundary crossings and build an according database. From an exploratory point of view, we were interested in whether it would be possible to train a convolutional neural network on a catalogue of bow shock crossings to obtain these directly, without the need for a fully segmented time series.

So far, we have developed a pipeline using only magnetic field and components, ion bulk-velocity and components, ion density, parallel ion temperature and perpendicular ion temperature from the Cluster 1 spacecraft, resampled to a 1-minute frequency. To account for the huge data imbalance, parts of the data, where no bow shock crossings are expected (for example when the spacecraft is in the night side part of the magnetopause or too far away in the solar wind) were removed. Since the temporal expansion of a bow shock crossing is quite limited, the labelling of the data had to be conducted thoughtfully. The labels are one-dimensional segmentation maps consisting of the values 0 or 1 for each point in time with a resolution of 10 min, indicating whether an ICME is taking place or not. We decided on a parameter between 0 and 1, which simultaneously defines if a given time
frame contains a bow shock crossing and how far from the centre it occurs.

The predicted label is clearly increasing for times when bow shock crossings occur. Thus, a peak detection algorithm can be used to extract a list of crossings. Even though Precision and Recall need to be improved, first results are promising and lead to the next steps:

• train on more data from different spacecraft
• use non-resampled datasets
• include additional features
• tune hyperparameters
• further experiment with model architecture
• cross-validation.

Metrics for a random split of the data can be seen in the table below.

Metric	Value
Precision	65 %
Recall	65%
True Positives	80
False Negatives	43
False Positives	43

The ML pipeline is available here on our GitHub repository.

Results of this science case were presented at the EGU21, at EPSC2021, at ESWW 2021, and at AGU21 (see presentations on the ML Portal and on GitHub). A further presentation was given in May 2021 at an international working group called 'CMEs, CIRs, HCS and large-scale structure' (led by, among others, Christian Möstl and Silvia Perri). This ML pipeline was presented in a workshop at EPSC2021 and is, together with a tutorial, available on our GitHub repository.

This science case was also presented at the EGU22 and the ESWW2022. Further, the results of this science case were published in the journal „Space Weather“:
Rüdisser, H. T., Windisch, A., Amerstorfer, U. V., Möstl, C., Amerstorfer, T., Bailey, R. L., & Reiss, M. A. (2022). **Automatic detection of interplanetary coronal mass ejections in solar wind in situ data.** *Space Weather*, 20, e2022SW003149. https://doi.org/10.1029/2022SW003149

References:

1. Nguyen, G., et al. (2019), Automatic Detection of Interplanetary Coronal Mass Ejections from In Situ Data: A Deep Learning Approach, *Astrophys. J. 874*, 145, doi:10.3847/1538-4357/ab0d24
2. Jha, D., et al. (2019), Resunet++: An advanced architecture for medical image segmentation, *arXiv e-prints,* arXiv:1911.07067

LMSU Boundaries Science Case

The goal of this case is to improve our understanding of Mercury's magnetosphere and its dynamics. We utilise the data recorded by the MESSENGER (MErcury Surface, Space ENvironment, GEochemistry, and Ranging) spacecraft, which collected vast amounts of heterogeneous data during its approximately 4000 orbit voyage, most interestingly the magnetic field data from the magnetometer. A typical orbit involved passing from the interplanetary magnetic field through the bow shock, the magnetosheath, the magnetopause, the magnetosphere of Mercury, and thereupon the same sequence in reverse. Since a mercurial year is about 88 Earth days, several years' worth of magnetometer data was recorded. This is nice because several variations in environmental configurations are recorded, which is useful to build automatic models for event recognition. The resulting data set of crossing times and positions is to be used in conjunction with the paraboloid magnetosphere model to compute the magnetic field lines in the magnetosphere; these can subsequently be used to perform modelling of trajectories of particles sputtered from the surface of the planet by space radiation.

Based on data from the mission, several global models of the magnetosphere were proposed (e.g., Winslow et al., 2013; Philpott et al., 2020). However, they could only describe an average shape of the bow shock and magnetopause crossings and can be prone to missing the statistical nuances in the data. Given large data, Neural Networks can be expected to approximate complex functions, which often surpass deterministic and rule-based methods, in a variety of time series tasks like classification (Fawaz et al., 2019), time series forecasting (Lim and Bohren, 2021), and rare time series event detection (Nguyen et al., 2018). We leverage these to develop a predictor, that can be used in real-time during orbit, to predict magnetic region for each step in a short window of observation. Figure 3 illustrates the different crossing labels for an exemplary orbit.

The use of statistical neural networks allows us to explore another aspect: with the help of active learning, it is possible to add samples to the training process incrementally. With this, we can examine how the model scales its predictive capacity with increasing data, and thus study how the variations such as changing solar wind and environmental conditions affects the manifestation of boundary signatures. To begin with, different orbits can be expected to have some element of similarity in the magnetic field structure, yet would have large variations in the same segments at different conditions. It is also interesting to study what the minimal amount is for the data needed to be able to generalize these phenomena for future missions such as BepiColombo.

The data set was manually labelled with the boundary crossings. To identify bow shocks, we first subtracted planetary dipole magnetic field components from the magnetometer measurements, computed the magnitude of the remainder attributed to external sources, applied the Savitzky-Golay filter to smooth the time profile of the remainder and computed its second derivative. The first and the last second derivative spikes as determined by z-score are assumed to be the enter and exit bow shock crossings respectively. Magnetopause boundaries were eyeballed using the cartesian components of the magnetic fields in the Mercury Solar Orbital coordinate system. During magnetopause crossings at least one of the components in the magnetogram experiences a sharp growth; the exact component depends on the spacecraft position. The beginning and ending points of this growth region are assumed to determine the magnetopause crossing edges. To supplement these, we also used the boundaries marked by Philpott et al. (2020) for a few orbits.

The distribution of the different magnetic regions, after annotation, is reported in Table 4. The

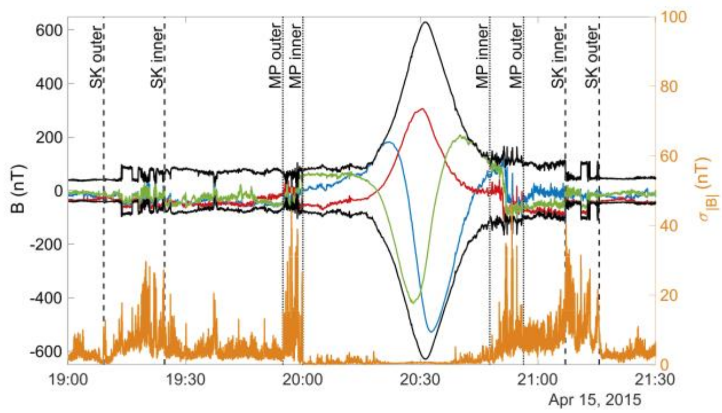

Figure 3: Exemplary labelled orbit from the work of Philpott et al. (2020).

boundaries of critical interest - bow shock and magnetopause - are minorities with only 3.7 and 2.3 % representation. The table highlights the data imbalance issue that requires investigating special techniques to ensure the predictor does not bias towards the overrepresented classes.

As a first step in pre-processing, feature selection was performed to assess the contribution of available features in the estimation of the output. Based on statistical correlations, the magnetic flux features (BX_MSO, BY_MSO, BZ_MSO), spacecraft position coordinates (X_MSO, Y_MSO, Z_MSO) and planetary velocity components (VX, VY, VZ) were found to be most informative. In addition, three meta features, namely EXTREMA, COSALPHA and RHO_DIPOLE, were selected.

In the feature preparation stage, a sliding window of variable sizes (3 seconds to 3 minutes) with a hop size of 1 second was computed on the time series signal to obtain feature vectors. Finally, the features were normalised to have mean of 0 and a standard deviation of 1. No other pre-processing or engineering was applied in order to allow the deep learning model to engineer features implicitly.

The windowed features are fed first into a block of 3 Convolutional layers with 1D filters, each followed by Batch Normalisation and ReLu activations. The activations obtained at the end of the CNN block are then passed to the Recurrent block with two layers of LSTMs. The final activations are then passed to a fully connected layer with softmax activations. The objective function used for training is Categorial cross entropy, with Adam optimizer.

The sample results in Figures 4 and 5 are from a model trained with two Mercury years of data, which is about 300 orbits.

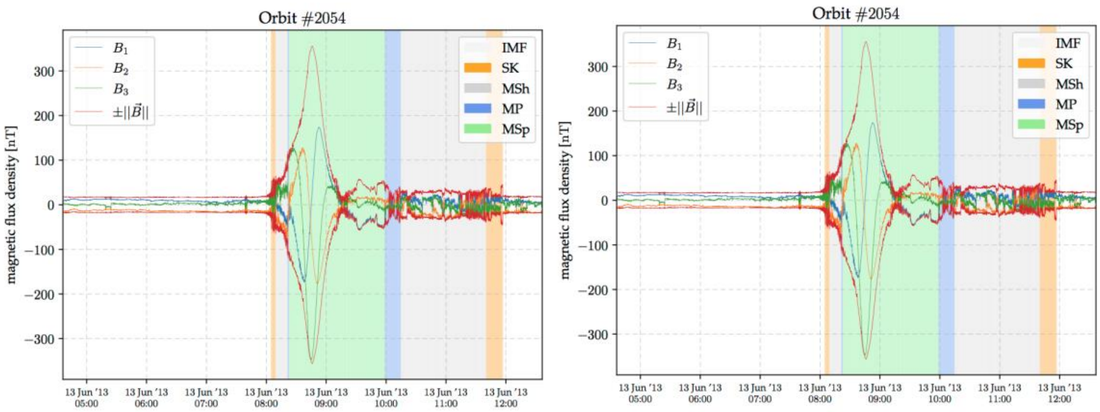

Figure 4: Left: Prediction; right: Ground truth.

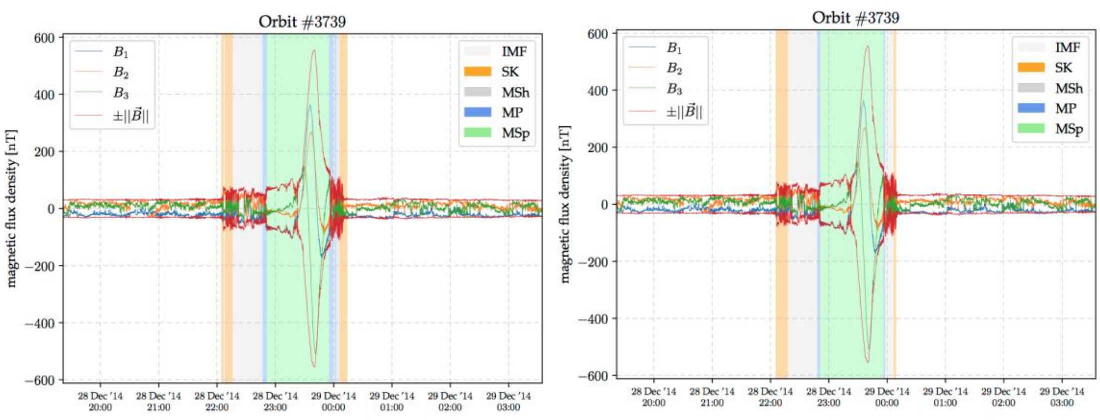

Figure 5: Left: Prediction; right: Ground truth.

Table 4. Class-wise distribution present in the data.

Label	Magnetic region	Statistical distribution
0	Interplanetary magnetic field (IMF)	65.4 %
1	Bow shock crossing (SK)	3.7 %
2	Magnetosheath (MSh)	14.5 %
3	Magnetopause crossing (MP)	2.3 %
4	Magnetosphere (MSp)	14.1 %

The window size used in these experiments is 30 seconds. Overall, the predictor achieves a macro F1 score of about 80% on the Bow-shock and the magnetopause crossings on a randomly sampled test of 300 orbits. None of the orbits overlap in the train and test sets.

Results of this science case were presented at the EGU21 as well as at the EPSC2021 (see presentations on the ML Portal and on GitHub). This ML pipeline was presented in a workshop at the EPSC2021 and is available on our GitHub repository.

We developed an efficient method to detect automatically the bow-shock and magnetopause boundary crossings using data from the MESSENGER magnetometer. To this end, we first prepared the data suited to Machine Learning. Next, we experimented with several ML models, specifically neural networks to find a usable baseline. Next, we devised an Active Learning (AL) approach to select only the most informative orbits in the training set by using an uncertainty criterion. Using this strategy, we were able to find a generalisable model with only 10% of the available data. A framework with these models was published and made available, open-source, for the community to experiment with on similar tasks.

We further extended this Active Learning approach by augmenting it with a Drift Detection strategy, such that the data sampler would first detect a distributional shift in the data, and then use the entropy-based criterion within the corresponding drift to order the most informative orbits. This further reduces the number of training orbits required, significantly outperforming random sampling. The results from this paper are in process of being assimilated into a paper to be submitted as an invited contribution in an AGU journal.

Results from this work were published in EGU, EPSC, ML-Helio 2022, and ECML-PKDD 2022.

The ML pipeline is available on our GitHub repository.

This science case was presented at the ECML PKDD 2022 and published in the proceedings of this conference:
Julka, S., Kirschstein, N., Granitzer, M., Lavrukhin, A. & Amerstorfer, U. V. (2022). Deep Active Learning for Detection of Mercury's Bow Shock and Magnetopause Crossings. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*.
https://2022.ecmlpkdd.org/wp-content/uploads/2022/09/sub_1177.pdf

References:

1. Philpott, L.C., et al. (2020), The Shape of Mercury's Magnetopause: The Picture From MESSENGER Magnetometer Observations and Future Prospects for BepiColombo, *J. Geophys. Res. (Space Physics) 125*, doi: 10.1029/2019JA027544
2. Winslow, R. M., et al. (2013), Mercurys magnetopause and bow shock from MESSENGER Magnetometer observations, *J. Geophys. Res. (Space Physics) 118*, 10.1002/jgra.50237
3. Fawaz, H.I., et al. (2019), Deep learning for time series classification: a review, *Data Mining and Knowledge Discovery 33,* doi: 10.1007/s10618-019-00619-1
4. Lim, B., and Zohren, S. (2021), Time-series forecasting with deep learn- ing: a survey, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379*, doi: 10.1098/rsta.2020.0209
5. Nguyen, V., et al. (2018), Applications of Anomaly Detection Using Deep Learning on Time Series Data. *In: 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), doi:10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00078.*

IAP wave emissions science case

This science case targets plasma wave identification from time-frequency spectrograms, specifically the electromagnetic whistler-mode "chorus" emission frequently observed in the inner magnetosphere of the Earth and other planets. The wave emissions typically occur as structured or unstructured features with visible boundaries in the time-frequency domain. We have created and delivered a training dataset of time-frequency spectrograms, 10 seconds each, generated from the data of the Wideband receiver on board four Cluster spacecraft. There are more than 4000 thousand events irregularly observed while spacecraft crossed the terrestrial magnetosphere and the nearby solar wind. We have visually checked the data and classified the intervals based on whether the chorus emission is present (82% of events) or not (18% of events). The value of local electron cyclotron frequency is included in the data for easier identification. The dataset is provided in the form of python data structures and can be used for both supervised and unsupervised machine learning.

Planetary magnetospheres create multiple sharp boundaries, such as the bow shock, where the solar wind plasma is decelerated and compressed, or the magnetopause, a transition between solar wind field and planetary field. The boundaries are identified by a discontinuity

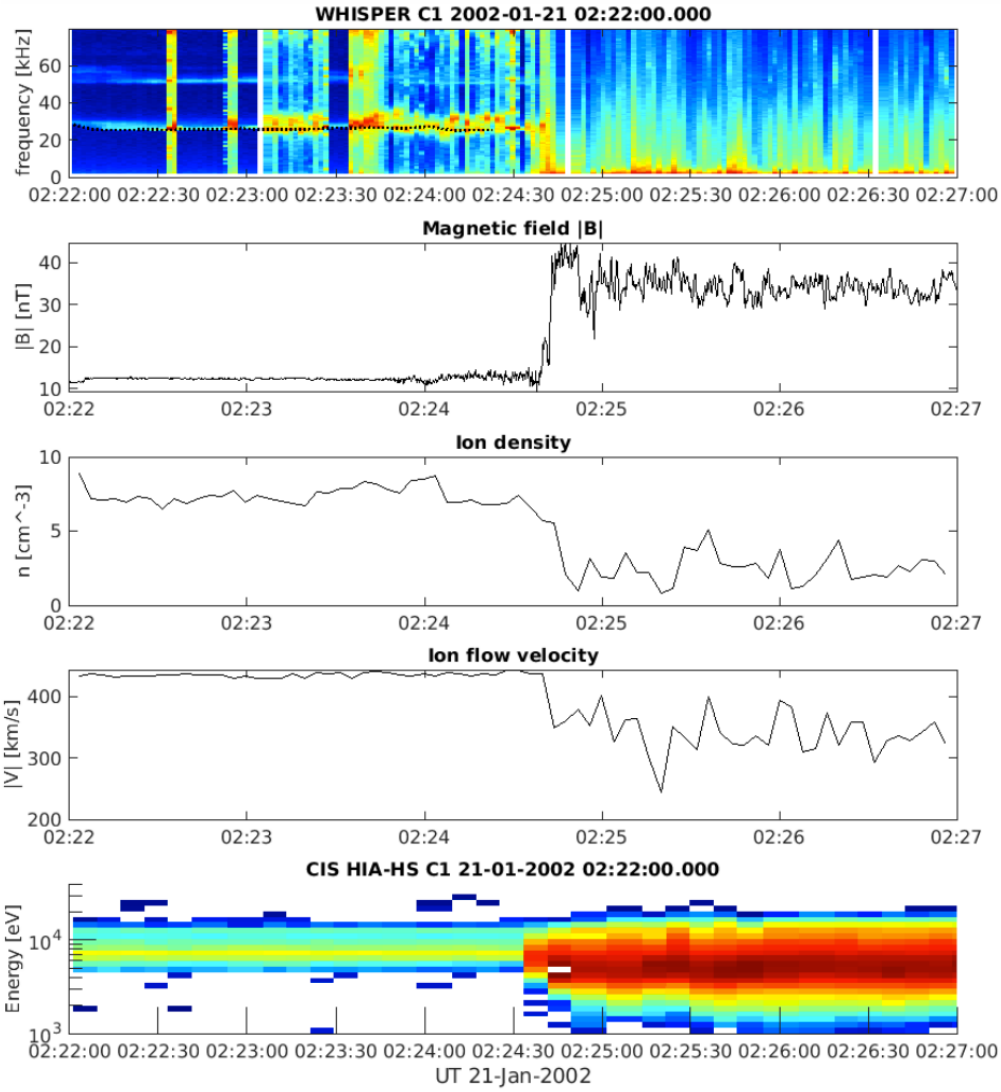

Figure 6: Example of a bow shock crossing.

in magnetic field, plasma density, and in the spectrum of high-frequency waves. These measurements are available on many planetary missions, such as Cluster or THEMIS (Figure 6). Due to the high amount of available data, a deep learning approach was found to be well-suited to automatically identify said boundaries. So far, the data has been pre-processed and the process of model development has been started. Code to process the original spacecraft data is available on GitHub.

Uni Passau received the data from the partners in Prague. Uni Passau pre-processed the data and conducted the basic explorative analysis. Since the task is to segment the pixels with the whistler waves in the spectrograms, and only have image-level labels are available, as a first task in feasibility Uni Passau built a classifier that classifies the image sample into a binary category of interest. This classifier provides an accuracy of about 70 %. As the segmentation task itself needs to be unsupervised, ie. without labels, Uni Passau is investigating an approach that might learn a similarity metric from the data and disentangle the representation space into relevant and irrelevant parts. This will be the major focus of this year’s development.

This science case was presented at the EPSC2021 and the AGU21 (see presentations on the ML Portal).

Task 5 - Images and Other (General) Classification Tools

A. Identifying Hazards and Resources

Every day, about 100 tonnes of rubble or dust from space enters the Earth's atmosphere. Most of this burns up without being seen or reaching the ground. However, larger meteors can cause fireballs that streak across the sky and some fragments can reach the ground as meteorites. Increasingly, camera networks dedicated to fireball tracking are being installed around the world, both to facilitate meteorite recovery for research purposes and to increase our understanding of impacts to support planetary defence initiatives. Europlanet 2024 RI has been working with the professional and amateur meteoritic research community to develop ML-powered tools that extract information from imagery of fireballs to help determine their characteristics, trajectories and potential origins.

Across the Solar System, images of planetary surfaces exhibit many common features, such as pits, mounds or craters. These features can reveal a wealth of information about the formation, history and potential useful resources of a planetary body. For example, 'skylights' or sinkholes on the Moon or Mars are of interest to geologists studying lava tubes and evidence of ancient volcanic activity; however, skylights are also potentially entrances to protected environments where underground habitats could be built for human explorers in the future. Craters can provide a detailed chronological record of the impact history of a planetary surface, potentially going back millions or billions of years, and may also trap water ice that could be used for life support and fuel. With a return of humans to the Moon planned within two-to-three years and international exploration strategies setting their sights on Mars, detailed and accurate mapping of surface features and resources at high resolution is essential. ML tools created by Europlanet 2024 RI and EXPLORE enable the automatic identification and labelling of mounts, pits, craters and other surface features. This not only enhances the speed of the mapping process but can also add in layers of information, such as the size, depth, composition and other characteristics of the features.

ML-based tools have also been developed to automatically calculate the depth of pits by detecting their shadows and measuring the width as it appears in satellite image. These will be primary targets for future space exploration and habitability since they are present on most rocky Solar System surfaces and, besides providing shelter from radiation, they have the potential to be entrances to sub-surface cavities which could, for instance in the case of Mars, harbour stable reservoirs of ice water.

GMAP Mounds Science Case

The GMAP Mounds identification science case aims to develop a generalised machine learning pipeline for the localisation and characterisation of specific geomorphological features (mounds) that are present on the surface of Mars. Mounds are positive relief features that can be ascribed to a variety of phenomena (e.g., De Toffoli et al., 2019). They can be related to monogenic edifices due to spring or mud volcanism, rootless cones on top of lava flows, pingos and so on. The focus of the investigation is related to the sedimentary/spring case of mud extrusion or sulphate oversaturated fluids. These objects usually are widespread regionally and/or contained in large complex craters (i.e., tens of km in diameter) often in populations of several hundreds/thousands. Previously, automatic detections were performed in some of these cases (Pozzobon et al., 2019) using topographic data in limited areas (i.e., Digital Terrain Models (DTMs) as rasters whose cells represent height values) in order to discriminate these objects in terms of pre-trained morphometric parameters and map them. Due to the scarcity of high-resolution DTMs and poor area coverage, the ML WP challenge is to reach the ability to detect such mound features by using simple grayscale panchromatic images at mid-high resolution with no need of topographic information.

The training set consists of two DTMs, one used for training and the other for testing. In the first step, the training DTM is tiled into several smaller fixed sized images. The label masks are created based on the available ground-truth shape files. The images are then scaled to be in range [-1,1]. The training set is then split further into train and validation sets with an 80/20 ratio. The train set is augmented in the next step with image manipulations such as flipping, rotation, rescaling and so on to create a large training set for the segmentation task.

For the initial image segmentation task, a standard UNet (Ronneberger et al., 2015) is trained using the training set. A mean IoU (Intersection over Union) of about 60 % on the validation set is obtained. This result is consistent with another GAN based model, indicating a saturation in information present in the training set.

Due to limited number of samples to train from, we learn a Generative model (Goodfellow et al., 2020) to approximate the true distribution of the landforms. We generate an augmented set using this approach and train the image segmentation again, observing an improvement of about 10% in the IoU. This is an interesting result, as it indicates that the model can be used to simulate the mound terrains. The approximated distribution space should be then factorizable into a set of independent mechanisms, which could control factors of variation.

A simulator of such likes can be used for controlled generation. Another advantage of latent space learning is that it can offer benefits in downstream tasks, which is an added advantage for storage and efficient searching. We have developed this simulator, and we plan to disseminate the method as a publication in the coming months.

Before the commencement of the previous year, we had already developed a basic segmentation pipeline using generative models (Image to Mask Autoencoders) to perform image segmentation and detect mound like features in the DTMs. As the data provided was severely limited, we initially attempted to generate simulated samples as an attempt to augment the training set. But that did not improve the segmentation. Finally, we augmented the training set with engineered features viz slope, aspect and hillshade and achieved a reasonable performance in the segmentation of mounds. Owing to a lack of proper validation set, it cannot be confirmed if the model will generalise well to unseen images.

In a parallel line of experiments, we investigated the representation space learnt by the generative models, wherein we devised a method to disentangle the mound specific representation from the non-mound representation, in order to perform controlled simulation. This would be useful to improve searchability in compressed representations, but it is yet to be determined if this approach can help directly in the main goal of segmentation.

Results from this work were presented in EPSC, EGU 2021, and the segmentation pipeline was demonstrated in a workshop in EPSC 2022.

The ML pipeline is available on our GitHub repository. The ML pipeline was presented during a workshop at the EPSC2022.

References:
6. Pozzobon, R., et al. (2019), Fluids mobilization in Arabia Terra, Mars: Depth of pressurized reservoir from mounds self-similar clustering, *Icarus 321*, 938, doi:10.1016/j.icarus.2018.12.023
7. De Toffoli, B., et al. (2019), Surface Expressions of Subsurface Sediment Mobilization Rooted into a Gas Hydrate-Rich Cryosphere on Mars, Scientific Reports 9, 8603, doi:10.1038/s41598-019-45057-7
8. Ronneberger, O., et al. (2015), U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Goodfellow, I., et al. (2020), Generative adversarial networks, *Communications of the ACM 63,* doi:10.1145/3422622

GMAP Pits Science Case

To improve the results obtained by the first tool for automated mapping of pits (DeepLandforms-YOLOv5, https://github.com/epn-ml/DeepLandforms-YOLOv5), a change of architecture was necessary. The results obtained by that tool, despite their goodness, need further processing since they are not immediately usable for proper mapping as they are composed only by a pair of coordinates that localize the centre of the detected features. Such detections still need to be properly mapped as polygonal shapes by users. Since this is a highly time-consuming and tedious task, it led to the development of a new tool, based on Deep Learning Instance Segmentation, to retrieve not only point coordinates of the detected features, but also a polygonal shape. The obtained results were then compared to the results obtained with the previous tool and with the MGC^3 database (Cushing et al. 2012, 2015) showing good results. A publication and this new tool will be released soon.

The next developing steps for *DeepLandforms* tool have been considered, for instance, we want to generalize it further by providing further baseline configuration. For instance, a configuration for PyTorch and another for Tensorflow python packages.
We are preparing a newer dataset to be tested with the updated tool.

Results of this science case were presented at the LPSC2021 and EGU21.

DeepLandforms has been presented with a live demo in a splinter-session at the Europlanet Science Congress 2022, held in Granada.
The paper presenting DeepLandforms, has been accepted in December 2022, and is available on https://doi.org/10.1029/2022EA002278.
The code is completely available on EPN-ML GitHub.
Constructor University (formerly Jacobs University) press release: https://www.jacobs-university.de/news/researchers-develop-ai-method-mapping-planets

References:
9. Cushing, G., et al. (2012), Candidate Cave Entrances on Mars, *J. Cave Karst Stud. 74*, 33–47, doi:10.4311/2010EX0167R
10. Cushing, G.E., et al. (2015), Atypical Pit Craters on Mars: New Insights from THEMIS, CTX, and HiRISE Observations, *J. Geophys. Res. (Planets) 120*, 1023–1043, doi:10.1002/2014JE004735
11. Nodjoumi, G., DeepLandforms-YOLOv5. Available online: https://zenodo.org/record/4430015 (accessed on 15 December 2021).

B. Tools for chemical characterisation

Many planetary missions carry spectrometers to gather information on the mineralogy of planetary surfaces. Remote sensing data suggest igneous rock-forming silicates on the surfaces of various bodies of our Solar System. This can help reveal the formation of rocky planets such as Mars and natural satellites such as the Moon. Minerals reflect certain parts of the electromagnetic spectrum more strongly than others depending on their chemical composition. To support these spectral studies, a Europlanet 2024 RI ML algorithm has been trained with reflectance spectra generated in laboratory experiments at the visible and infrared wavelengths that are key to unlocking surface composition. An additional bespoke tool has been trained for unsupervised classification of infrared observations by NASA’s MESSENGER spacecraft, which orbited Mercury between 2011 and 2015. With ESA’s BepiColombo mission set to arrive at the Solar System’s innermost planet in December 2025, this tool is helping BepiColombo’s spectrometer team to perform vital groundwork in identifying areas of particular interest and developing workflows for future data analysis.

DLR Surface Composition Science Case

In this science case, Mercury surface reflectance data from the MASCS instrument onboard the NASA/MESSENGER mission is analysed. First, NASA/PDS data is converted into a relational DB (PostgreSQL). Then the data is regridded with custom Postgis/PostgreSQL spatial queries. This produces a global hyperspectral data cube image of normalized MASCS visible (VIS) detector spectra, from the first Earth year of the orbital mission. The cube contains some anomalies, in regions of low coverage or from high levels of spectral variation within a single pixel. Thus, data artifacts, instrumental and photometric residual effects are all removed. The resulting data cube has several hundred features that are compressed via blind signal demixing with Independent Component Analysis (ICA). Initial results show that four components reconstruct the original dataset within the measurement estimated error. The four features were embedded in a two-dimensional space via Uniform Manifold Approximation and Projection (UMAP). No significant small-scale morphology was found after exploring UMAP hyperparameters. Finally, the 2D maps were partitioned with hierarchical agglomerative clustering. Dendrogram gap analysis shows a big gap between data partition in three and four clusters, and three clusters have been chosen as sig`nificant data segregation. At this initial stage, the existence of two large and spectrally distinct regions has been found, which have been designated the polar spectral unit and the equatorial spectral unit (see Figure 7).
The spatial extent of the polar unit in the northern hemisphere generally correlates well with that of the northern volcanic plains and partially to the surface highest temperature models in the equatorial region. This may indicate an interaction between mineral composition and structure and surface temperature, because Mercury reaches a diurnal temperature above 700 K. Chemical data spatial distribution from X-ray and Gamma ray spectrometers show no apparent correlation with the clusters. This could indicate that chemical composition produces no distinctive mineral phases for the instrument or that those phases were altered enough to be indistinguishable by the harsh space environment around Mercury. 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 characters.

First results of the science case were presented at EGU21 (see presentations on our ML Portal).

During the last year, DLR organized the code repository to be self-sufficient from data source to end result. The DLR ML team produced a complete report (PDF/HTML) present in the repository and published with Elsevier. The aim is to make the user able to reproduce the complete work just using the information contained in the repository. A video tutorial on DLR use case for GMAP Winter School has been released. The presentation is in the DLR repository as well, the video is available on GMAP server (https://www.planetarymapping.eu/).

INAF spectral analysis for planetary minerals case

(https://github.com/epn-ml/spectral-analysis-planetary-minerals)

We analyzed the best spectra data to provide to the project and implemented a procedure to format the data in a standard way using JSON format for data storage and transport; in this

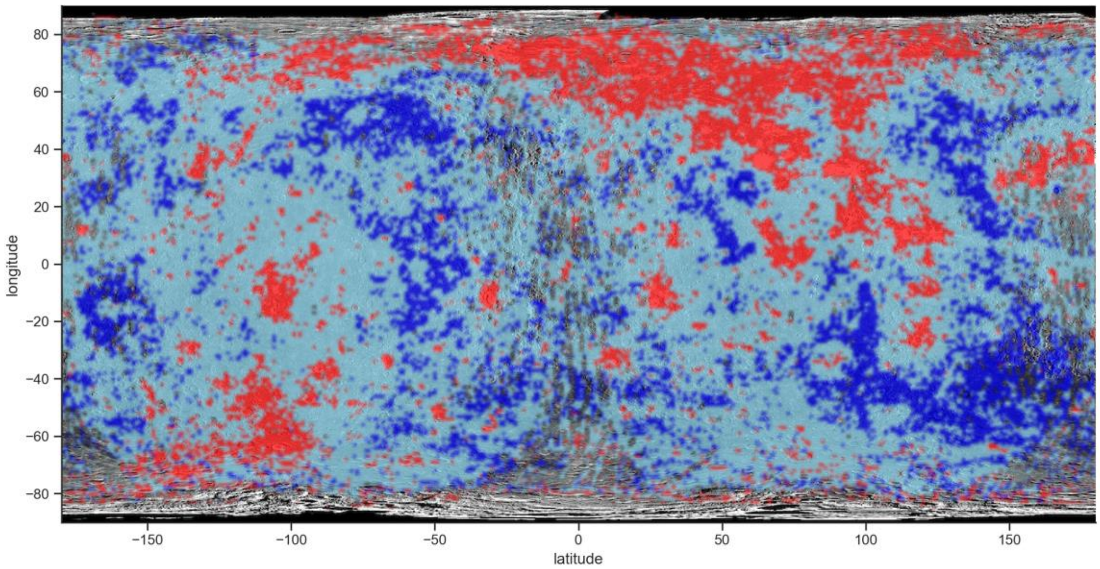

Figure 7: Agglomerative Clustering 3 classes.

way the communication of the methods of reading and managing the dataset will be simple, the metadata necessary for their understanding will also be stored in the dataset. In Fig. 8 the summary of the samples obtained by learning and test subsets are shown (from 683 samples in total, 152 samples are labelled indicating the element abundances). We test some ML algorithms on the selected dataset, to test if the dataset can be used for an ML analysis.

Figure 8. The number of samples which do not have consecutive wavelengths (and its reflectance). For more details see README.md

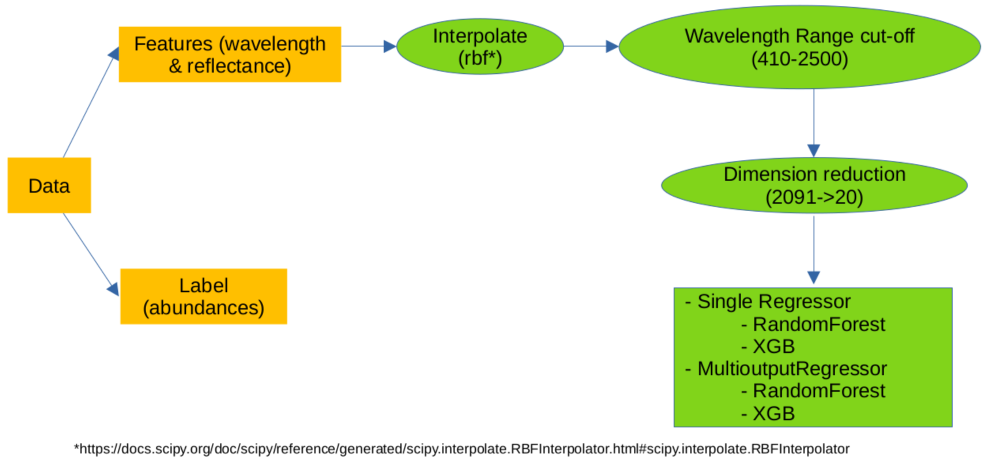

Figure 9: The Pipeline chart flow.

This is a multi-output classification or multi-output regression problem where the input is the obtained reflectance, and the output is a composition of different mineral phase names.

- In regression task, the output is a list of real numbers ranging from 0 to 100 whose sum must be 100.
- In classification task, the output is a list of binary values (0 or 1) which indicates that phase name exists (1) or not (0).

AOP Asteroid/meteor Science Case

The goal of this science case is to search for asteroids thought to exist along the Earth's orbit that may be leftover material from the formation of our planet. These asteroids always appear close to - or even behind - the Sun in the sky and are therefore difficult to detect from Earth. Images taken from the two STEREO probes which have been studying the Sun and its vicinity since 2009 will be used as the basis for the science case. The spacecraft have been slowly drifting along the Earth's orbit and are able to image the sky from different vantage points around the Sun. This enables the abundance of asteroids in Earth-like orbits to be constrained, including any large (hundred-metre to kilometre size) objects in unstable paths that are not picked up by surveys and present a long-term impact hazard to our planet.

During this reporting period, it was determined that Machine Learning would not be of benefit to the science cases originally proposed by AOP. Following some conferring within WP10, a re-defined science case was formulated aimed at the classification of meteor lightcurves. AC is presently in collaboration with Andreas Windisch (FH JOANNEUM) to take this forward to implementation. Work by AC in the last few months of 2022 focused on extracting the lightcurves from the raw data and already several thousand lightcurves have been made available to Dr Windisch and his group for pre-processing.

Task 6 - Virtual Access and Interfaces

The Machine Learning Portal provides the public point of entry to our ML activities. We continuously update the content on the portal. We also improved the ML Portal structure according to comments of the VA review board. A first draft of how JRA4 services can be onboarded into the EOSC has been provided in year 1 of the WP, including a description of the EOSC, EOSC portal and hub, and the European Grid Infrastructure (EGI). Onboarding a service into EOSC means that the service is listed in the portal of the EOSC site (like a shop window) but is hosted by the service provider. The EOSC expects mature services (TRL8/9) to be onboarded. Further possibilities to onboard ML demonstrator services on the EOSC are being explored. A preliminary list of requirements for onboarding has been identified.

Since 2022, we have been working on finalising and disseminating the APP (Analysing Planetary Pits) tool. APP is a Python framework for automatically deriving apparent depth profiles of Solar System pits by measuring the width of their shadows. It uses image segmentation to separate cropped satellite images (single- or multi-band) into shadow pixels or non-shadow background and calculates a profile of the apparent depth (h) of a pit – the depth at the edge of the shadow – along the entire length of the shadow. The testing of the shadow extraction is complete, proving that k-means clustering with silhouette analysis was the most accurate method. APP has been presented at a number of conferences and forums over the past year (EAS Annual Meeting, RAS National Astronomy Meeting, Congress on Geomorphology), including hands-on sessions at the Europlanet Science Congress in Granada where participants got to use APP in practice for the first time. A paper has been written describing APP which has now been submitted to the journal - Royal Astronomical Society's Techniques and Instruments (RASTI). The short-term plan is to make the tool publicly available in the ACRI-ST GitLab, the Europlanet 2024 RI GitHub and the EXPLORE platform.

The EXPLORE platform (https://explore-platform.eu) is a development platform whose main purpose is to validate, test and demonstrate the scientific data applications (SDAs) being delivered by the EXPLORE project. These SDAs will subsequently be deployed also on other platforms – when these are ready – such as ESA Datalabs and ESCAPE SAP. This portability is key to bringing the SDAs close to the data.

A joint effort between Europlanet RI 2024 and EXPLORE is now ongoing to update the EXPLORE platform to allow the deployment of JupyterLab-type applications (a technical update is necessary to run JupyterLab based docker images) which will be used to deploy the Jupyter notebooks. The LMSU boundaries ML pipeline was the first implemented in EXPLORE.

The following restriction are to be noted:

1. Only registered users can run SDAs on the platform, this is needed for resource management and also to attach the user's workspace to the running SDA. In this early phase of EXPLORE, the registration is upon invitation/request. In the longer-term self-registration may be added.
2. The EXPLORE platform is (currently) a development platform, which means that it has limited computing resources in the back end. In the longer term, it is foreseen to add elasticity to the infrastructure resources and evolve it into an operational service.

The deployment of tools in the ESA Datalabs is also being investigated.

c. Impact to date

In July 2023, the development phase of Europlanet 2024 RI's ML tools is largely complete, and training and dissemination are underway to support their adoption by the community. The synergistic collaboration between Europlanet 2024 RI and EXPLORE has also demonstrated the EXPLORE platform's usefulness to the planetary community in providing a test environment to deploy ML tools and other applications. The platform fills a previously unidentified gap in supporting the development of applications to the point of maturity needed for deployment on major data hubs, such as the European Science Open Cloud (EOSC) and ESA Datalabs.

Much of the development work for the ML tools has been performed by early career researchers, enabling them to build both their skills and experience and their professional profile within the scientific community.

The Europlanet 2024 RI ML tools were designed to be applicable for a large number of applications dealing with scientific databases. The structure of the ML repositories includes not only the description of the scientific cases with the corresponding ML technique but also the results that one can obtain by applying given technique. Therefore, the outcomes of the ML training can be compared and can be applied autonomously by the user in the spirit of independent standalone best practice to learn how to analyse a scientific problem. Repositories with documentation, numerical scripts and scientific graphs are available.
Overall, by developing ML tools tailored to planetary science, Europlanet 2024 RI has cemented collaborations, started to build new user communities and developed services that are already resulting in publications. While the planetary science community could be seen as late to the party in adopting ML, interest now is high. This couldn't be more timely – with flagship missions to Mercury and Jupiter soon adding to the deluge of data streams, the era of data-driven science is only just beginning.

In summary, at different occasions, e.g. conferences, we have presented results of our science cases as well as our ML activities in Europlanet 2024 RI. We have published a number of publications with ML contributions reporting our ML results of the scientific cases.

We have organized and convened four conference sessions specifically dedicated to ML in planetary sciences and heliophysics (and we will organize such sessions again in 2023) and we organized an ERIM session in June 2023.

Five workshops were conducted in the course of EPSC2021 and EPSC2022 to introduce our ML pipelines to the scientific community. All ML pipelines are available on our public GitHub repository. The session „Machine Learning in Planetary Sciences and Heliophysics“ at the EGU 2023 had 24 abstracts submitted.

An article describing the achievements of ML WP has been released on INNOVATION NEWSNETWORK portal(https://www.innovationnewsnetwork.com/machine-learning-for-a-new-era-of-data-driven-planetary-science/35810/).

d. Summary of plans for Year 4

Currently, we are working on feasible integration of first data sets of our science cases into VESPA. Further, we have integrated most of our ML pipelines into SPIDER and the EXPLORE platform.

Finally, we will finalize our work on the remaining science cases and publish the final version of our Jupyter book, containing documentation and tutorials about our work.

2. Update of data management plan

The data management plan will be updated in order to incorporate the comments raised by the VA review board.

3. Follow-up of recommendations & comments from previous review(s)

We have answered the issues raised in the VA review board report in a collected answer of all VAs.