Deep neural networks for surface composition reconstruction from in situ exospheric measurements at Mercury

Adrian Kazakov[a], Anna Milillo[a], Francesco Lazzarotto[b], Stavro Ivanovski[c], Valeria Mangano[a], Miguel Escalona-Morán[d], Alessandro Mura[a], Stefano Massetti[a], Stefano Orsini[a], Elisabetta De Angelis[a], Rosanna Rispoli[a], Nello Vertolli[a], Alessandro Aronica[a], Roberto Sordini[a] and Raffaella Noschese[a]

[a] INAF-IAPS, via Fosso del Cavaliere 100, 00133, Rome, Italy

[b] INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122, Padova, Italy

[c] INAF-Osservatorio Astronomico di Trieste, via Giambattista Tiepolo 11, 34143, Trieste, Italy

[d] Augmented Intelligence Lab, Rua Lugo 2, 36470, Salceda de Caselas, Spain

E-mail: `adrian.kazakov@inaf.it`

Abstract. The surface information derived from the exospheric measurements at planetary bodies complement the surface mapping provided by devoted imagers. In fact, crucial knowledge about the surface release processes, the dynamics of the numerous interactions within the planetary environment, the erosion, space weathering and, eventually, about the evolution of the planet can be obtained. This study examines a tentative proxy method to derive the elemental and mineralogical composition of the regolith of Mercury from in situ measurements of its neutral exosphere through the use of Deep Neural Networks (DNNs). We present here three supervised feed-forward DNN architectures - a single network of fully-connected neural layers, a double network working in parallel, and a double network working in series, whose inputs are the exospheric densities and proton precipitation fluxes measured in a mock-up orbital run in view of the analysis of data from the SERENA (Search for Exospheric Refilling and Emitted Natural Abundances) instrument package on-board the Mercury Planetary Orbiter - part of

the BepiColombo space mission to Mercury, nominal phase starting in 2026. The tasks of the algorithms are to predict from those exospheric measurements the constitution of the surface regolith below in terms of chemical elements and mineralogy. Furthermore, this study presents a simplified model for generating the training datasets, from which the DNNs learn to recognize the relations between the neutral abundances of the exosphere and those on the surface. The processes considered here to act on the surface and to release neutral species to the exosphere are micrometeoroid impact vaporization (MIV), ion sputtering (SP), photon-stimulated desorption (PSD) and thermal desorption (TD). The uncertainty of their efficiencies in the models are explored to examine the range of accuracy of the methods with respect to different process parameters. An extensive testing campaign is carried out, which shows the adequacy of the DNNs to predict and reconstruct the surface composition maps from proxy exospheric measurements, and reveals an in-depth view of the capabilities of the algorithms. At the same time, the tests identify a broad potential for further development, where the method would aid the analysis of the complex surface-exosphere interaction and further reduce the uncertainties in the models for the generation of the planetary exosphere.

Keywords: Mercury, Exosphere, Surface composition, Deep Neural Networks

1. Introduction

The celestial bodies in our Solar System are constantly subjugated to the action of external agents, like the solar wind, solar radiation and micrometeoroids, which reshape them by adding, extracting, altering or moving material. These complex interactions with the surrounding environment are reverberating on both their surfaces and their atmospheres. In the case, where the atmospheres are very tenuous, as at Mercury, they are called exospheres - planetary envelopes where collisions between their constituent particles are so rare that the particles' trajectories can be considered ballistic [1], [2]. The exosphere of Mercury is the result of various processes and interactions of the surface with the planetary environment. The external factors acting on the planet, such as the dust particles, solar wind protons and heavy ions, solar radiation and intense heat are reflected on the composition and dynamics of the exosphere [3]. The investigation of the highly complex interaction between the outside environment and Mercury's magnetosphere, surface and exosphere is one of the main objectives of the BepiColombo mission to the innermost planet, because it could reveal to us not only knowledge about Mercury itself, but also another

view on the forces shaping up our Solar System [4], [5], [6].

By modeling the active processes and their effects on the surface we could simulate the generation of the exosphere. This is an extensively applied method that compares the results of the simulations to those measured from space ([7], [8], [9]) or from Earth ([10], [11], [12]). The models use mathematical relations about the physical processes in act to produce the exosphere to the extents of our knowledge coming from observations and/or laboratory experiments. However, there is an inherent complexity of the various interactions that includes electromagnetic, chemical, mechanical, thermal and other effects, some of which have not been exactly evaluated for each release process. This may give a very broad range of simulated results with a lot of uncertainty, depending on the assumptions made earlier in the modeling.

Particles that impact the planet may have the capability to release the surface material from it. The particles released into the exosphere can be subjected to other processes due to the interaction with the radiation pressure, photons and charged particles. These interactions can alter the charge and chemical status and the dynamics of the exospheric particles. However, in a first approximation in the sparse exosphere, the exospheric atomic and molecular abundances resulting from these impacts could be traced back to the planet connecting the surface properties, like composition, mineralogy and physical state to the different processes and the dynamics of matter around the planet [6], [13]. In the literature it is widely discussed the active processes in the formation and dynamics of Mercury's tenuous atmosphere [10], [14], [15], [16]. There are four predominant processes that act on the surface to release atoms and molecules into the exosphere. These are the micrometeoroid impact vaporization (MIV), the sputtering after solar wind and heavy ion impact (ion sputtering, SP), thermal desorption (TD) and photon-stimulated desorption (PSD). Two of them - the MIV and the SP could serve as valid indicators of the regolith composition below, as they tend to have higher energy transfers, which are able to uproot neutral species from their mineral. On the other hand, the TD and the PSD are usually less energetic, thus are able to release atoms and molecules bonded to the minerals with a weaker bond, like volatile elements, which in their main part fall back and are reabsorbed on the surface [3], [17], [18], [15].

Backtracking the conditions measured at an altitude above the surface to a previous state below is a complicated task to be done analytically. Machine learning algorithms can omit the detailed analytical description of all the processes at the surface and can be a good solution to

find the highly non-linear relationships between independent variables (inputs) and dependent variables (outputs) driven by those processes. By learning from example, artificial intelligence (AI) algorithms can resolve to some estimation the data generation mechanisms [19], [20], [21]. This would give another tool in our toolbox to explore in depth the relationships between the planetary environment's components. This work will show Deep Neural Networks (DNNs) within the data analysis of Mercury's exosphere to reconstruct the surface map underneath. DNNs, and especially the multilayer perceptrons, are suitable for the nonlinear regression tasks in hand, but are also better able to scale them with increasing training data and input parameters [22], [23], [24]. The basis of this method was presented in Kazakov et al. (2020) [25], where a highly simplified model for the generation of the exosphere by the action exclusively of MIV was traced back to the mineral constitution of the surface. The basic DNNs explored there have shown moderately accurate outcomes of the mineralogy reconstruction with the limited amount of tests performed in their study. Hereby, we are reporting a more deeply developed continuation of the method described in [25] with the current study performing a thorough investigation of improved algorithms on data generated through a more complex and plausible assumptions for the exospheric models. The development of this method would aid the constraining of the exopheric generation models and would also give insights to the interaction between the surface and the exosphere. This study is performed in view of the upcoming BepiColombo ESA/JAXA mission, which with its two spacecraft - the Mercury Planetary Orbiter (MPO) and the Mercury Magnetospheric Orbiter (MMO) - presents a unique opportunity to study the planet from two different vantage points with a variety of instruments and sensors [6]. The development of the Deep Neural Networks and the subsequent scientific analysis are envisioned as a part of the Ground Segment of the suite of neutral and charged particle detectors SERENA (Search for Exospheric Refilling and Emitted Natural Abundances) on-board the MPO. In particular, the study is directed towards the future utilization of measurements from the SERENA sensors: the mass spectrometer STROFIO (STart from a ROtating Field mass spectrOmeter), which will measure the exospheric gas composition, MIPA (Miniature Ion Precipitation Analyser) and PICAM (Planetary Ion CAMera), which would detect plasma fluxes from the Solar Wind and magnetosphere toward the surface, and ELENA (Emitted Low Energy Neutral Atoms) providing the ion precipitation map onto the surface through back-scattered neutral detection [26], [27], [28].

In the next **Section 2** we establish the multivariate regression tasks for surface reconstruction. The algorithms to accomplish these tasks are further characterized component by component, and the recipe for building working neural networks architectures are defined. The architectures' hyper-parameters choice assisted by a Bayesian search is portrayed later in the section. **Section 3** gives insights on the simulated data generation, including the models of the environment and exospheric measurement simulation, dataset definition and feature selection for the input of the DNNs. The results of a broad training and test campaign, as well as the appropriate metric to compare the algorithms, are discussed in **Section 4**. The reconstruction of the surface mineralogy and elemental composition is also demonstrated in a visual way. Some ranges of the algorithms' predictions in terms of the variation of the exospheric model parameters are explored. The study concludes with **Section 5**, where a view of future improvements and applications is outlined.

2. The Methods

Kazakov et al. (2020) [25] have explored three different Deep Neural Network algorithms for the task of mineral composition prediction from exospheric composition measurements in a supervised learning setting. In their dataset generation they have assumed only the action of the micrometeoroids on the surface and the subsequent vaporization (MIV), which is expected to be the dominant release process on the night side of the planet. In this study we have continued and fortified their investigation on the application of Deep Neural Networks for the supervised multivariate regression (MVR) on the input data. We have taken their fully-connected MVR feed-forward network (here referred to as simply the MVR) working with inputs from a single exospheric volume, which has shown moderately accurate results in the reconstruction of the mineralogy of the regolith surface - 85.95 % test set similarity (ES1, see Section 4.1 for explanation of this metric). Improvements on the method are made in the following directions:

- extension of the application of the algorithms to the elemental composition prediction, and to the sequential prediction of elemental and mineral fractions;
- development of two more branches of the MVR DNN architecture with the aim to employ domain knowledge splitting the prediction at the borders of the data generation mechanisms;

- utilization of hyper-parameter tuning algorithm to find the optimal design parameters of the Deep Neural Networks;
- employment of enhanced features to improve the algorithms' inference at different scales (such as measurements at a wider range of altitudes);
- advent of DNNs capable to make predictions in a multi-process environment for the exospheric generation using a model with four active processes for the expelling of neutral species - MIV, SP, TD, PSD.

2.1. Prediction Tasks

The Deep Neural Networks developed in this work are applied to the following three tasks for the reconstruction of the regolith source material: (1) *"elements"* - prediction of the elemental surface composition fractions from the exospheric density measurements above; (2) *"minerals"* - direct prediction of the mineral composition fractions from the exospheric density measurements above; (3) *"e2m"* - prediction of the elemental surface composition fractions from the exospheric density measurements above, followed by prediction of the mineral surface composition fractions from the elemental surface composition fractions. The three different tasks are summarized in the network depictions shown on Figure 1.a, 1.b and 1.d.

In their essence, those are all multi-variate regression tasks, which would output the values of either the elemental or mineral composition fractions in the form:

$$\sum_{i=1}^{n} Fraction_i = 1 \tag{1}$$

where n is either the total number of elements or the total number of minerals taken into account in the prediction task.

The algorithms trained for these prediction tasks encode in themselves the relationships between the various processes that act on the surface in order to generate the exosphere. When viewed from the neural networks' process estimation point of view, the convention of our study establishes a split of the overall exospheric generation to three steps: (1) decomposition of the regolith source mineralogy to its constituent elements - atoms and molecules; (2) ejection of the elements from the surface according to process efficiency and energetic distributions; (3) behavior of the elemental release under various actions to create the state of the exosphere at a

given time.

Then, the *"elements"* task "reverse engineers" the second and third steps of the exospheric generation - the processes yields (efficiencies to extract the atoms) and chemistry, and the energy and spread distributions - and learns how to reconstruct them passing through a single DNN estimator. The *"minerals"* task also incorporates into a single DNN the first step to predict the mineralogy, while the *"e2m"* task has the same objective, but step (1) is separated from steps (2) and (3) in a second DNN. By doing this split, the networks' internal non-linear relations, which model the reconstruction process, are divided into two presumably simpler data representation estimators.

2.2. The Deep Neural Network Algorithms

We developed three neural network architectures capable to carry out the tasks described above. (1) *MVR* - a single feed-forward fully-connected Deep Neural Network for the tasks "elements" and "minerals" (shown on Figure 1.a and 1.b), essentially a multilayer perceptron [22]. (2) *MVRx2* - a double feed-forward fully-connected DNN consisting of two identical neural networks working in parallel for the tasks "elements" and "minerals" (Figure 1.c), where each input example is routed to only one of the two DNNs based on a certain condition. There are three subtypes of this neural network, in which the condition for the routing is respectively the altitude of measurement below or above 1000 km, measurement on the dayside or on the nightside, or measurement with or without the presence of an incoming proton flux. Each DNN branch (total of six, or two for each subtype) is trained only on examples, which satisfy its condition. The three different subtypes are examined as a step towards a single DNN architecture, which would incorporate all the conditions for the split of the data distribution estimators.

(3) *MVRe2m* - a double feed-forward fully-connected DNN consisting of two identical neural networks working in series for the task "e2m" (Figure 1.d), where each input example passes sequentially through both networks.

The three architectures and their subtypes, combined with the prediction tasks, form 9 DNN/task combination nomenclatures (essentially 9 different algorithms) and are described in Table 1. All the algorithms are written in Python and use the Keras programming framework

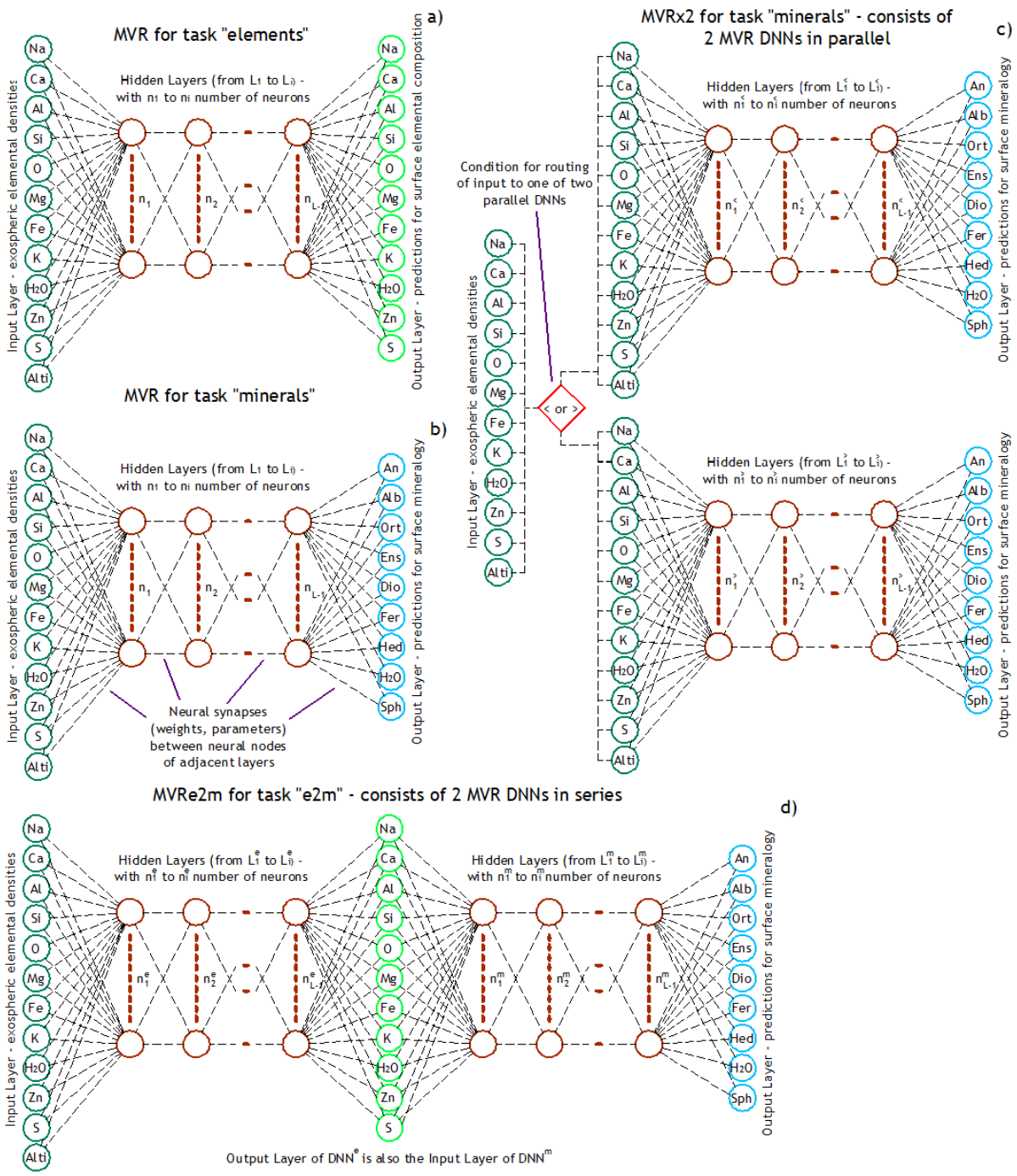

Figure 1. DNN algorithms schematics: (a) MVR for task *"elements"*, (b) MVR for task *"minerals"*, (c) MVRx2 for task *"minerals"*, (d) MVRe2m for task *"e2m"*. The input layer for all the networks shown are the exospheric densities + the altitude of the measurement (feature set F01, see §3.3.3). The hidden layer structure consists of L number of layers, each connected to its adjacent by synapses (the weights of the network). On each hidden neuron is performed a ReLU activation function. The output is the relative surface composition: mineralogy for tasks "minerals" and "e2m" and elemental composition for task "elements".

with the TensorFlow backend [29]. While, the three different DNN architectures and their subtypes are distinct on a macro level, their constituting parts and their working principles are very similar.

DNN Algorithms		
Algorithm nomenclature	**DNN Architecture**	**Prediction Task**
MVR/"elements"	Single DNN	"elements"
MVR/"minerals"	Single DNN	"minerals"
MVRx2-1000/"elements"	Parallel DNNs, split at 1000 km altitude	"elements"
MVRx2-1000/"minerals"	Parallel DNNs, split at 1000 km altitude	"minerals"
MVRx2-DN/"elements"	Parallel DNNs, split for dayside and nightside	"elements"
MVRx2-DN/"minerals"	Parallel DNNs, split for dayside and nightside	"minerals"
MVRx2-H+/"elements"	Parallel DNNs, split for presence of H+ flux	"elements"
MVRx2-H+/"minerals"	Parallel DNNs, split for presence of H+ flux	"minerals"
MVRe2m/"e2m"	Serial DNNs	"e2m"

Table 1. The ten different prediction algorithms (DNN/task combinations).

In this sense, we create our algorithms by combining the following components of the neural networks:

- The input layer - consists of the input neurons of the network, which enumerate the so called features of each data example. Additionally, each feature is standardized by subtracting the mean and dividing by the standard deviation, to give a better performance to the learning algorithm.
- Hidden layers - each layer constitutes a collection of hidden neurons, which are connected to all the neurons of the previous layer (hidden or input) and to all the neurons of the next layer (hidden or output) by neural synapses to form a multi-layer perceptron (MLP) [22], [23]. The full network of synapses is represented by combination (weight) matrices, while on each hidden neuron is further performed a non-linear operation - the activation function. In all of our networks, the hidden layer activation function is the rectified linear unit (ReLU), which prevents the multilayer perceptron networks from having vanishing gradients [30]. When properly trained the parameters of the weight matrices and the activation functions form a representation (estimation) of the relationships among the features of the input layer (the independent variables), which when combined give an accurate numerical prediction on its regression task [21]. This estimation for each layer is given with the equation:

$$\boldsymbol{h} = a(\boldsymbol{W}^T\boldsymbol{x} + \boldsymbol{b}), \tag{2}$$

  where $\boldsymbol{W}^T$ is the weight matrix, $\boldsymbol{x}$ is the input vector for the hidden layer (input features or activations from a previous hidden layer), $\boldsymbol{b}$ is the bias vector of the affine transformation, and the function a is the ReLU ($a(\boldsymbol{z}) = \max(0, \boldsymbol{z})$).
- The output layer - in our networks there is one output neuron for each element or mineral, which represents the fraction of this element or mineral in the analyzed surface area. A softmax function is used to exponentiate and normalize the output in order to produce the relative surface composition with values between 0 and 1 and the total of all output neurons to sum to 1:

$$\hat{y}_i = \text{softmax}(\boldsymbol{z})_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}. \tag{3}$$

- Loss function - used to evaluate the difference between the output of the neural network

(prediction) and the ground truth given in the specification of the dataset. In all our DNNs we utilize the mean squared error to evaluate the error on each example and sum for the total loss in each learning iteration:

$$J(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{m} \sum_{i=1}^{m} \|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{\boldsymbol{i}}\|^2 . \tag{4}$$

- Regularizer - a set of constraints or penalties added to the loss function. This is necessary in order to reduce the variance of the trained algorithms on the inference for the examples it has not seen during its training. We employ an L2 regularization on the weights of each hidden layer (also known as weight decay) - a particular constraint that does not allow them to become very large and overfit the learned parameters to the particular dataset used for training [31]:

$$\hat{J}(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{y}) = J(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{y}) + \lambda \sum_{i=1}^{m} \|\boldsymbol{\theta}_i\|^2, \tag{5}$$

  where λ is the regularization coefficient and $\boldsymbol{\theta}$ is the vector of all weight parameters (unfolded $\boldsymbol{W}$ matrices).

- Optimization (learning) algorithm - after one iteration of predictions the weights of the combination matrices for each hidden layer should be updated in order to reduce the total error of the prediction. We perform a backpropagation optimization ([32], [33]) to update the weights of the matrices by following their gradients with respect to the loss function:

$$\boldsymbol{\theta} := \boldsymbol{\theta} + \alpha \frac{1}{m} \bigtriangledown_{\boldsymbol{\theta}} \sum_{i=1}^{m} L_i(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{y}), \tag{6}$$

  where α is the learning rate and $L_i = \|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{\boldsymbol{i}}\|^2$ is the loss of each example. And in particular we use the Adam algorithm - a variation of the stochastic gradient descent, which employs momentum and an adaptive learning rate [34]. The stochastic variant of the gradient descent implies that the learning iterations are not performed on the full example set, but only on a random portion of it called a mini-batch. Hence, in equation (6) the number m is the number of examples in the mini-batch.

2.3. Hyper-Parameter Optimization

A necessary part of training the DNN networks is the fine tuning of their hyper-parameters. As hyper-parameters in this case we define the following:

- Learning rate of the backpropagation optimization algorithm - sets how big of a step the algorithm takes with each optimization pass along the weight gradients of the loss function. This is a particularly important hyper-parameter, which has to be balanced to be large enough to overcome plateaus in the parameter space of the loss function, but at the same time small enough in order to reach the minimum of the error (or at least stay in the vicinity of the minimum).
- Mini-batch size - the size of the random example subset, on which each training iteration is performed;
- Number of hidden layers of the neural network;
- Number of neurons in each hidden layer;
- L2 regularization coefficient - defines the significance of the penalty on the weights becoming too large.

We have used a Bayesian optimization with the Gaussian Process (GP) approach [35]. The hyper-parameter tuning algorithm is based on the *scikit-optimize* library [36]. The algorithm uses prior probability distribution function to find the hyper-parameters that give the minimum total loss on the cross-validation dataset. The optimal hyper-parameters found after 50 iterations of the GP optimization on the *MVR/"minerals"* algorithm are shown in Table 2. Those were found to be adequate also for the other 9 network/task combinations, where the further fine-tuned hyper-parameter set has not given distinctively better results during training and hold-out validation.

3. The Datasets

Since the BepiColombo nominal mission will commence in 2026, we cannot use real data from SERENA. At the same time, as the application of the methods would be to reverse engineer a particular model for the reconstruction of the surface composition, we are compelled to create our datasets through simulations of this model.

Hyperparameter Tuning					
Parameter	**Minimum search value**	**Maximum search value**	**Prior**	**Optimal value after 50 iterations**	**Chosen value**
Learning Rate	10^{-6}	10^{-2}	log uniform	0.000468	**0.0005**
Regularization Coef.	10^{-8}	10^{-2}	log-uniform	0.00000895	**0.00001**
Mini-batch Size	32	Full dataset	normal	648	**512**
# of Layers	1	10	normal	4	**4**
# of neurons, Lay.1	50	1000	normal	589	**600**
# of neurons, Lay.2	50	1000	normal	511	**500**
# of neurons, Lay.3	50	1000	normal	344	**350**
# of neurons, Lay.4	50	1000	normal	249	**250**
# of neurons, Lay.5	50	1000	normal	-	**-**
# of neurons, Lay.6	50	1000	normal	-	**-**
# of neurons, Lay.7	50	1000	normal	-	**-**
# of neurons, Lay.8	50	1000	normal	-	**-**
# of neurons, Lay.9	50	1000	normal	-	**-**
# of neurons, Lay.10	50	1000	normal	-	**-**

Table 2. Bayesian hyper-parameter optimization for the MVR network

There are 2 types of datasets created for evaluating the algorithms: training datasets and test datasets. The generation of all datasets is as per the following procedure: (1) The surface mineralogy is generated for random mineral compositions; (2) The minerals are decomposed by mol% to their constituent elements; (3) The planet position and orientation along the orbit are calculated given the observation time; (4) The effects acting on the surface are applied by adhering to a particular model for each desorption process; (5) The released particles for each process and each species are calculated; (6) The energy distribution of the ejecta for each process and each species is modeled; (7) The exospheric density is computed for each considered altitude for each species; (8) A mock-up orbit representing the MPO spacecraft is initiated; (9) A scan of the exosphere is performed at the altitudes of the mock-up orbit, giving out directly the exospheric density above each surface area (without taking into account instrument response functions). In order to generate more examples for the algorithms to learn from, this procedure is repeated on a large number of different randomly created surfaces.

3.1. Models of the Environment

A variety of models of the environment have been used in the past to simulate the exosphere at the times of the measurements of orbiting spacecraft. In the forward analysis methods those simulated measurements are compared to the real observations of either column densities

or surface densities [7], [8], [9], [10], [11], [12]. The hereby developed method is reverse-engineering the processes that create the exosphere and is predicting the surface composition by reconstructing a particular model used in training the DNN algorithms. The DNNs are additionally tested on a subset of variations of the same model with some process parameters changed with respect to the original one. However, the method is not constrained to be trained on only this particular model, which is specially defined for the present algorithms' development, but is envisioned to be capable to reconstruct any exosphere generation model and predict the surface composition with respect to it, following the construction of suitable interfaces. Those models, on whose generated datasets we obtain successfully trained Deep Neural Networks, can be termed *base nodes*, around which would be constrained the reconstructed mineral and elemental composition in the hyperspace of the environmental model parameters. This study focuses on one *base node* model and the vicinity of its parameter space. The following subsections will describe it and its variations, and how they are applied to create the training and testing datasets.

3.1.1. Surface Models and Creation

The grid of the surface taken in our construction is relatively coarse comprised of 36×18 surface tiles in a modified Mercator projection. Each tile is 10°×10°, which at the equator translates to ≈425 km×425 km. As a first step in our dataset generation we create a regolith mineral composition from the minerals shown in Table 3, which represents a mineral set that is assumed to feature on the surface of Mercury [16]. However, in order not to overfit the training towards a particular mineral composition and to have a good enough size of training examples, we generate 40 different surfaces randomly for the training datasets, where different regions on the map have also different mineral compositions. In this manner we acquire a wide range of diverse mineral fractions across our data. For the test datasets we create 3 other surfaces, which represent more probable mineral compositions for Mercury (described in Table 3).

The algorithm for creating a random surface mineralogy is the following: (1) For each mineral create a random number of patches (from 6 to 10) on the map grid, where each patch can be as small as a surface tile and can reach several tiles in size; (2) Set the rare minerals, which cannot be chosen as primary or secondary (such as water, which in this case is taken as a source/mineral; (3) Randomly choose one mineral to be primary, and one mineral to be secondary

Train and Test Datasets Mineralogy						
Mineral Name	Chemical Formula	Test Subsets Mineralogy			Train Subsets Mineralogy	
		Fraction in Test Subset 1	Fraction in Test Subset 2	Fraction in Test Subset 3	# of subsets as Primary	# of subsets as Secondary
Anorthite	$CaAl_2Si_2O_8$	0.316	0.122	0.223	9	6
Albite	$NaAlSi_3O_8$	0.174	0.317	0.131	6	7
Orthoclase	$KAlSi_3O_8$	0.086	0.108	0.111	6	6
Enstatite	$Mg_2Si_2O_6$	0.061	0.174	0.294	7	7
Diopside	$MgCaSi_2O_6$	0.114	0.126	0.019	6	6
Ferrosilite	$Fe_2Si_2O_6$	0.088	0.076	0.167	6	5
Hedenbergite	$FeCaSi_2O_6$	0.039	0.044	0.018	0	1
Sphalerite	ZnS	0.084	0.021	0.014	0	1
Water Ice	H_2O	0.038	0.012	0.023	0	1

Table 3. Mineral composition of the test and train datasets. The fractions given for the test sets are averages across the surface tiles. The numbers given for the train sets are number of maps created with such a mineral being primary or secondary.

and add everywhere on the map these kinds of minerals; (4) Set randomly other ever-present minerals, which will have lower fractions than the primary or secondary mineral; (5) Calculate the fractions of the minerals on each tile of the surface layer grid of the regolith. The resulting mineral composition is the "actual", or the "ground truth", fed to the algorithms ("minerals" and "e2m" tasks).

After the mineralogy of each surface from the datasets is established, it is decomposed into its 11 different constituent elements (atoms or molecules) by the classical additive method. The resulting elemental composition is the "actual", or the "ground truth", fed for the "element" task of the algorithms.

Above was described the model for the creation of the source materials, which is the target for the reconstruction algorithms. However, there is also the ambient population of elements and molecules, which is fallen back to the surface and adhered to the regolith grains. It is the population of Na, K, H_2O, S and O_2, on which the thermal and photon-stimulated desorption processes are acting to release again to the exosphere. We've chosen to model this population as the average between the source elemental fraction on a particular tile of the surface grid and the mean elemental fraction over the whole planet surface.

3.1.2. Active Effects and Processes for Generation of the Exosphere

The active effects considered here are the four main surface release processes described in Section 1, namely the MIV, SP, TD and PSD. However, there are quite a few existing gaps in the understanding of these processes, which make the problem not fully constrained in terms of what is observed and what model parameters correspond to the observations. In this study we consider some simplifications of the parameters describing each surface release process, those are the yield (number of released particles after the impact of a particle or a photon), the released particles energy distributions and the spread of the elemental populations through the exosphere. The simplifications made in the modeling are justified in the current state of development of the algorithms by the capability of still generating plausible exospheres, while the time to test their feasibility for the application is reduced. Although in the definition of the models valid assumptions on the generation processes from the literature are used, we hereby define our own nomenclature for the model amalgamation used in this study. The basic model name is PTF-M01 coming from the initials of the PipeTheFirst prototype analysis software and model variation M01 - the base node of the model parameters.

Micrometeorite impact vaporization (MIV)

The MIV is described in the model as an incoming flux of interplanetary dust, which expels matter from the surface regolith relative to the incoming flux intensity and velocity [37]. We have taken a value of the flux at Mercury perihelion of 2.8x10^{-16} g/cm^{2}s. The incoming flux is increased on the ram side of the planet and decreased on the trailing side [38]. In our model, for simulating this dawn-dusk asymmetry in the in-fall of micrometeorites, this relation is simplified to a linear law starting from the 2.8x10^{-16} g/cm^{2}s at the leading side and falling down to half of this value on the trailing side. At the same time, there is a linear relation considered between the planet distance to the Sun and the velocity of the bombarding micrometeorites resulting in a decrease of the flux at aphelion and a temporal asymmetry. Similarly, for the impactor velocity, we have taken the modal impact velocity from [37] of 15.8 km/s at perihelion, which is increased on the leading side, and decreased on the trailing side and at aphelion. Thus, the range of the incoming flux in our MIV model is from 0.7x10^{-16} to 2.8x10^{-16} g/cm^{2}s and the range of the incoming micrometeorites velocity is from 9 to 28.5 km/s. The vapor release from the MIV is taken as 5 times that of the incoming flux as per [37] (efficiency of 5). This results in outgoing

fluxes of vapor from 3.5x10^{-16} to 14x10^{-16} g/cm^{2}s. It should be noted that the vaporized species contain also bigger molecules, such as CaO, NaOH, NaO and others from the complex chemistry in the impact-produced cloud [39], [40], but for our first iteration of the DNN analysis we have considered a very short photolysis lifetime of these species, which are quickly destroyed to their constituent elements with no further energization. In this model we use Maxwellian (thermal) distributions for the velocities of the excited particles. The temperature of the vapor depends on the in-falling flux intensity and velocity and ranges from 3000 K to 5000 K (average 4000 K as in [12]).

Ion sputtering (SP)

The SP is a flux of bombarding ions, which expels with some efficiency atoms/molecules from the surface regolith [16], [3]. We are taking an intermediate flux of protons equal to 5x10^8 1/cm^{2}s ([3]). This value is increased in the current model with decreasing distance to the Sun (by 50 % at perihelion to 7.5x10^8 1/cm^{2}s). The magnetospheric cusp shapes are assumed to widen in area when the solar wind dynamic pressure increases as shown on Figure 2. A more dynamic dependence on the magnetic reconnection rate driven by the Interplanetary Magnetic Field (IMF) strength and orientation is neglected, as well as the northward shift of the magnetic dipole of Mercury. This simplification is considered satisfactory for the development stage of the DNNs. In the main variation PTF-M01 the yield efficiency of this process is taken constant for each species of outgoing atoms/molecules and is given a fairly high value of 0.1 as to what is in the literature [41]. This is justified in order to have a more pronounced effect coming from the SP in the analysis of the capabilities of the algorithms in the high latitude zones of solar wind precipitation. The resulting outgoing total flux is then from 0.5 to 0.75x10^8 1/cm^{2}s. Again, a thermal distribution is assumed for the released particles ranging from 7000 K to 14000 K (in the literature, 6500 K for Na and 12000 K for Ca, [12]).

Thermal desorption (TD)

The TD in this model is acting only on the ambient population of molecules (Na, K, H_2O, S, O_2), on the surface as in [15]. The model assumes a constant flux of desorbed neutral species limited by the diffusion rate of 1x10^7 1/cm^{2}s and a $\approx$ 1% loss, which for a surface-exosphere

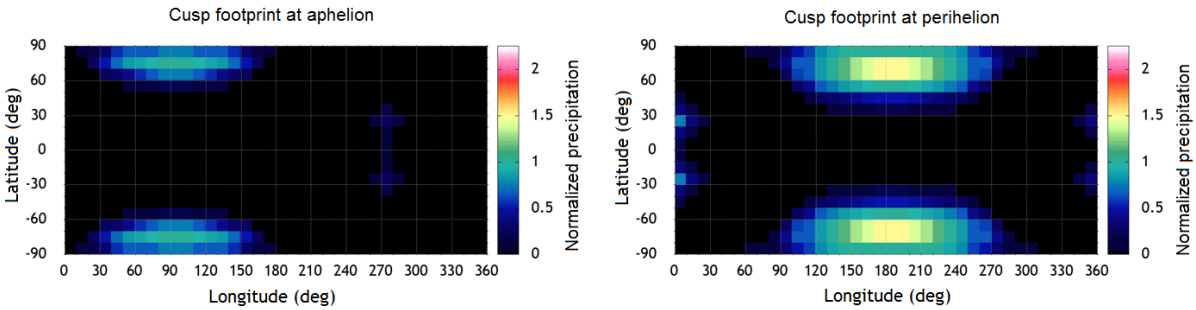

Figure 2. Magnetospheric cusp footprint shapes at aphelion (a) and perihelion (b). The incoming flux is normalized to the maximum value at aphelion (5x10^8 1/cm^{2}s). Note: Longitude meridians are fixed on the surface.

system at equilibrium yields about 1x10^9 1/cm^{2}s at 550 K - the sub-solar point temperature at aphelion (from [15]). The temperature at perihelion increases to about 750 K, which in our current model results in an increase of the released particles to 1.32x10^9 1/cm^{2}s for the volatile species Na, K, H_2O, S, O_2. It is a simplified linear increase in the thermal desorption sufficient for the purposes of the current study. The thermal map on the surface is following a "one-fourth" law between a minimum temperature of 150 K and the maximum temperature depending on the distance from the Sun:

$$T_s(\phi, \theta) = T_{\min} + (T_{\max} - T_{\min})(\cos\phi\cos\theta)^{1/4}, \tag{7}$$

where ϕ is the latitude and θ is the longitude [15]. The released particles are following a Maxwellian energy distribution from 350 K to 750 K.

Photon-stimulated desorption (PSD)

The PSD is modeled as a flux of incoming photons each able to expel atoms/molecules from the ambient population of volatile species with some efficiency depending on the cross-section for the impact of the photon [12], [42], [16]. The flux of incoming photons is taken from 1x10^{16} 1/cm^{2}s at aphelion to 2.25x10^{16} 1/cm^{2}s at perihelion [12], which would yield for a Na population ($f_{\mathrm{Na}} = 1$), for example, an outgoing flux from 2.5x10^9 1/cm^{2}s to 5.6x10^9 1/cm^{2}s following the equation:

$$\phi_{\mathrm{Na}}^{\mathrm{PSD}} = \frac{1}{4}\phi_{\mathrm{ph}} Q_{\mathrm{Na}} f_{\mathrm{Na}} N_{\mathrm{S}}, \tag{8}$$

where $\phi_{\mathrm{Na}}^{\mathrm{PSD}}$ is the outgoing flux of sodium particles due to PSD, ϕ_{ph} is the incoming flux of photons, Q_{Na} is the PSD cross-section for Na (taken as $Q_{\mathrm{Na}} = 1.4\mathrm{x}10^{-21}$ cm^2 [42]), f_{Na} is the fraction of sodium present in the regolith, and N_{S} is the regolith surface density (assumed to be $N_{\mathrm{S}} = 7.5\mathrm{x}10^{14}$ cm^{-2} [42]). Here we use again the thermal distribution for the energies of the outgoing particles from 1050 K to 1400 K [3], [15].

Energy distribution curves and trajectories

In the following step of the exospheric simulation the thermal distributions for the velocities of the excited particles are translated to distributions with altitude following ballistic trajectories. We have applied a uniform distribution in azimuth angles and a cosine law for the elevation angles. These two types of particle distributions (energy and angle) in the released particles result in the spread of the elemental species ejected from one surface tile to the exospheric volumes above that tile, but also to adjacent volumes with increasing altitude [25], [43] (See Figure 3). We have not taken into account in our model photoionization, charge exchange, dissociation or radiation pressure on the particles, which for some of the volatile species (Na, K, etc.) could be moderate in magnitude, but is a second order effect at the relatively low altitudes of the Mercury Planetary Orbiter's orbit passes.

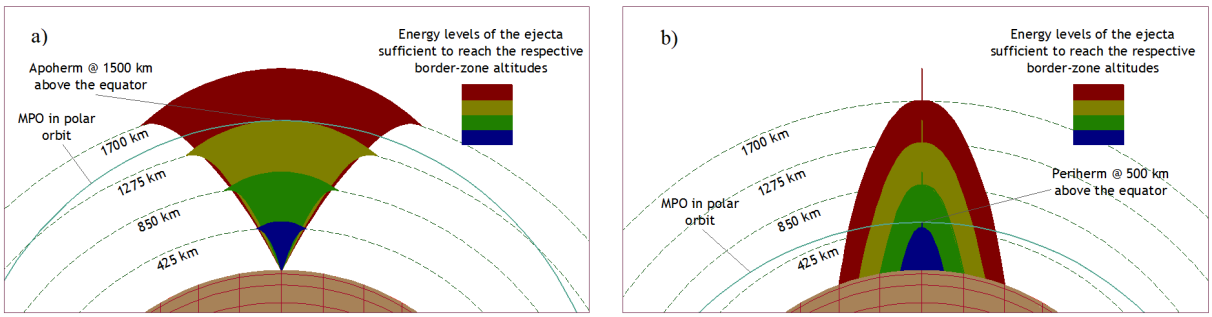

Figure 3. Exospheric particle collection from the surface of the planet: (a) Spread of the vapor from a MIV event on the surface of Mercury showing particles of different energies reaching different altitudes. (b) Contribution of areas on the surface of Mercury to the composition of the exosphere at different altitudes. (Figure taken with permission from [25])

The model utilized in this study - PTF-S011-M01 - is thus much more closer to the actual processes, thought to be acting on the surface of the planet to release the neutral particles and create its exosphere, than the one used for the initial development of the MVR Deep Neural Network from Kazakov et al. (2020) - the PTF-S005. However, we remain cautious about

the many simplifications assumed in our model for the exosphere generation, but we deem the complexity sufficient to assess the capabilities of the DNN algorithms to characterize the surface composition in a trace back from the exospheric density measurements. We have identified steps to go further in the roadmap for the application of the method to real data from measurements around Mercury (see §5).

3.1.3. Model Variations

The training datasets do not have varying basic model parameters for the exospheric release processes. Therefore, the data representation learned from the algorithms is the one resulting from the models described in the previous subsection - the M01. However, in order to test the capabilities of the algorithms to accurately reconstruct the surface, and inspect the confidence intervals of their predictions in the parameter-space vicinity of M01, we set for some of the test datasets different parameters for the processes in the model (see Table 4). The model parameter variations are divided into four groups - variations on the species release temperature for the MIV and SP processes (from M02 to M07), variations in the process efficiency for the MIV, SP and PSD processes (from M08 to M15 and from M20 to M23), variations in the ambient population composition at equilibrium (M16 and M17, and other added effects - namely, a coronal mass ejection (CME, M18), which increases both the ion precipitation area and the proton flux (by a factor of 1.5) and an increase of micrometeoroids by cometary dust streams (M19), which increases the incoming micrometeoroid flux by a factor of 1.5. The M16 model variation uses an ambient population, which is closer in composition to the source target population, while M17's ambient population is closer to the mean elemental composition of the entire planet surface. Additionaly, we use a different ion sputtering yields for the variations M21 (refractory species at 0.05 yield, volatile species at 0.10 yield), M22 (refractory species at 0.10 yield, volatile species at 0.05 yield) and M23 (the species yields are extrapolated according to their surface binding energies taken from [16], K with a low binding energy was kept at yield of 0.10, while all other species are given a lower ion sputtering yield, e.g. Si with a high binding energy has yield of 0.021). The model variations with the largest differences to M01 are the M24 and M25, where a combination of parameters have been changed with respect to M01.

3.2. Measurement Simulations

Model Parameters Variations				
Model Number	Process Temperature	Process Efficiency	Ambient Population	Other Effects
M01	MIV 4000 K SP 10500 K TD 600 K PSD 1050 K	MIV 5 SP 0.1 TD N/A PSD 18	(source + mean) /2	-
M02	MIV 3000 K	-	-	-
M03	MIV 5000 K	-	-	-
M04	SP 7000 K	-	-	-
M05	SP 14000 K	-	-	-
M06	MIV, SP low	-	-	-
M07	MIV, SP high	-	-	-
M08	-	MIV 3.5	-	-
M09	-	MIV 6.5	-	-
M10	-	SP 0.035	-	-
M11	-	SP 0.07	-	-
M12	-	PSD 10	-	-
M13	-	PSD 40	-	-
M14	-	all low	-	-
M15	-	all high	-	-
M16	-	-	source	-
M17	-	-	mean	-
M18	-	-	-	CME
M19	-	-	-	comet
M20	-	Only MIV	-	-
M21	-	mix SP 1	-	-
M22	-	mix SP 2	-	-
M23	-	mix SP 3	-	-
M24	MIV, SP high	all low	mean	-
M25	MIV, SP low	all high	source	-

Table 4. Model Variations

In expectation of the arrival of BepiColombo at Mercury and the analysis of the exosphere with its unique set of instruments, namely the SERENA suite of particles detectors, this study utilizes for its datasets mock-up measurements of the exosphere densities and the proton incoming fluxes as if coming from orbits at an altitude range similar to the Mercury Planetary Orbiter. The virtual measurements are formed by combining the orbital motion with the observation properties over the generated exosphere as described in the previous subsection. For the datasets used here we have defined three different orbits:

- MPO-P/A - an elliptical mock-up of the MPO polar orbit with a periherm at 500 km altitude and an apoherm at 1500 km altitude - all the training sets are created with this

orbit, and most of the test sets;

- C500 - a circular orbit of 500 km altitude;
- C1500 - a circular orbit of 1500 km altitude.

Moreover, these orbits are combined with one of two sets of observational scenarios to define a measurement:

- Mercury year - this is the observational scenario performed for the training datasets. It is combined with the MPO-P/A orbit to scan the exosphere for a full Mercury year around the Sun. In this time Mercury completes $1\frac{1}{2}$ rotations and the virtual spacecraft completes three scans of the exosphere above each area of the surface. Only one measurement is taken above each surface tile per scan resulting in 648 measurements per scan. For one Mercury year this gives us a total of 1944 example measurements per training set.
- Static swipe - this is the observational scenario performed for the test datasets. It freezes the orbit of Mercury at a given TAA-M and performs a single scan of the exosphere along one of the orbits MPO-P/A, C500 or C1500. This would result in a total of 648 example measurements per observation (test set). Note that for the MPO-P/A orbit the altitudes are different along the orbit, resulting in half the exosphere sampled at lower and half at higher altitudes.

3.3. Dataset Generation

The training and testing datasets are generated with the PipeTheFirst prototype AI data analysis software. Both types of datasets are created from the base model M01 or a variant of it in case of some test sets for the model variations test campaign (See §4.2.3).

3.3.1. Training Sets Generation

There are 41 training sets generated - 40 for training and 1 control set used for hold-out validation during the hyper-parameter optimization. Each of them has 1944 example measurements of the exosphere and is representing an observation during one Mercury year starting at perihelion. The model for the exosphere generation during the pseudo-dynamic simulation is M01, hence no variations in the process parameters have been enforced to any of the training exospheres. However, the base regoliths for the exospheres are different for each

dataset. The distribution of primary and secondary minerals among the 40 datasets is shown in Table 3.

3.3.2. Test Sets Generation

The total number of test datasets is 90, but unlike the training sets, they are created from only the three distinct regolith bases (Table 3). The differences between each group of three test sets is realized by varying either the orbit, the true anomaly angle of Mercury or the parameters of the model for the generation of the exosphere. The first 18 test datasets are used in the algorithm testing campaign without variation of the process parameters enforced on them (so using model M01) (see in Table 5). The last 72 test datasets are created by enforcing the model variations from M02 to M25 on each of the three base regoliths.

Test Sets							
Test Set Number	Test Examples	Regolith Name	Regolith		Orbit	TAA_M	Model Number
			Primary	Secondary			
Map001	648	R01-An	Anortite	Albite	P/A	0	M01
Map002	648	R01-An	Anortite	Albite	C500	0	M01
Map003	648	R01-An	Anortite	Albite	C1500	0	M01
Map004	648	R01-An	Anortite	Albite	P/A	180	M01
Map005	648	R01-An	Anortite	Albite	C500	180	M01
Map006	648	R01-An	Anortite	Albite	C1500	180	M01
Map007	648	R02-Al	Albite	Enstatite	P/A	0	M01
Map008	648	R02-Al	Albite	Enstatite	C500	0	M01
Map009	648	R02-Al	Albite	Enstatite	C1500	0	M01
Map010	648	R02-Al	Albite	Enstatite	P/A	180	M01
Map011	648	R02-Al	Albite	Enstatite	C500	180	M01
Map012	648	R02-Al	Albite	Enstatite	C1500	180	M01
Map013	648	R03-En	Enstatite	Anortite	P/A	0	M01
Map014	648	R03-En	Enstatite	Anortite	C500	0	M01
Map015	648	R03-En	Enstatite	Anortite	C1500	0	M01
Map016	648	R03-En	Enstatite	Anortite	P/A	180	M01
Map017	648	R03-En	Enstatite	Anortite	C500	180	M01
Map018	648	R03-En	Enstatite	Anortite	C1500	180	M01

Table 5. Test sets subset used in the algorithm testing campaign

3.3.3. Feature Selection

The DNN algorithms estimate the relationships between the input features from a set of training examples. However, the exact features selected to describe each example could influence the performance of the algorithms to find an accurate representation of the data. Hence, we

define three different feature sets to better understand the capabilities of the neural networks to predict the surface composition and what additional information they could give for the models of the release processes:

- F01 - all 11 elements' exospheric densities + the altitude of the single measurement;
- F07 - all 11 elements' exospheric densities + the altitude of the single measurement + sun incidence angle on the surface below + $\mathrm{TAA_M}$ + proton flux virtual data;
- F08 - the log10 of all 11 elements' exospheric densities + the altitude of the single measurement + sun incidence angle on the surface below + $\mathrm{TAA_M}$ + proton flux virtual data.

4. Results

The three different DNN architectures - MVR, MVRx2 and MVRe2m - were thoroughly tested for the nine distinct network/task combinations (Table 1) with a variety of settings for training in order to identify their capability to reconstruct accurately the surface composition.

4.1. Performance Metric

The performance of the algorithms is evaluated using a custom metric, which employs the Euclidean distance between the actual surface composition $\boldsymbol{y}_i$ and the predicted surface composition $\hat{\boldsymbol{y}}_i$ in vector space - the Euclidean Similarity 3 (ES3). It shows a better resemblance to a percentile accuracy when measured on the analyzed data distribution than other traditional metrics used in regression tasks, such as the R^2. The ES3 is defined as:

$$\mathrm{ES3} = 1 - \frac{\sqrt{\sum_i (\hat{y}_i - y_i)^2}}{\sqrt{\sum_i y_i^2}}. \tag{9}$$

To compare with the results reported in [25], at some places is noted also the Euclidean Similarity 1 (ES1) performance metric defined as:

$$\mathrm{ES1} = 1 - \frac{\sqrt{\sum_i (\hat{y}_i - y_i)^2}}{\sum_i y_i}. \tag{10}$$

4.2. Training and Test Campaigns

A total of three training and test campaigns have been undertaken, which amount to 534 separate tests for the inference of the algorithms on the test datasets. The training consists of 60 different training runs for a total training time of approximately 23 hours. The majority of the training runs (48 of them) are undertaken for the first of two test campaigns on the algorithms, where the learning curves of the algorithms are examined in conjunction with the bias/variance inspection with the growing sample size, and the same feature set F01. While the other 12 are undertaken with the aim to examine the feature sets F07 and F08. The third and final campaign uses the best training instances found in the previous two campaigns for each of the three tasks (elements, minerals, e2m) to survey the ranges of the network predictions when we vary the data generation model parameters in the test sets. No new trainings are performed for the last campaign.

4.2.1. Algorithms Campaign 1 - Bias/Variance, Learning Curves

For the first campaign a total of 144 tests were made with 3 test datasets - Map001, Map007 and Map013 (see Table 5) to inspect the performance of all 9 different algorithms. The fixed parameters in the campaign are the exosphere generation model parameters (model variation M01), the initial $\mathrm{TAA_M} = 0$ (True Anomaly Angle of Mercury's orbit around the Sun), the orbits of the test sets (MPO-P/A), and the feature set (F01). The objective was to show that the algorithms behave nominally under training and inference by examining their learning curves, bias and variance metrics, and find the optimal training time for inference to avoid overfitting [44].

The tests on the MVR DNN in both tasks *"elements"* and *"minerals"* display the expected behavior - the variance on ES3 metric shown between training set prediction and test set inference decreases with increasing the number of training examples, over which it learns. The learning curves in Figure 4 show that for the *MVR/elements* combination the optimal training time is around epoch #400, while for the *MVR/minerals* combination it is around epoch #150. As observed, the "minerals" task is harder to predict than the "elements" one, resulting in about 11 % ES3 difference in training between the two. The learning curve for the *MVR/minerals* algorithm (Figure 4.b) gives an indication that there is further potential to improve the training

prediction due to its fast rate of increase towards the end of training. However, the variance has to be reduced in order to achieve good inference, possibly through more training examples and/or further regularization.

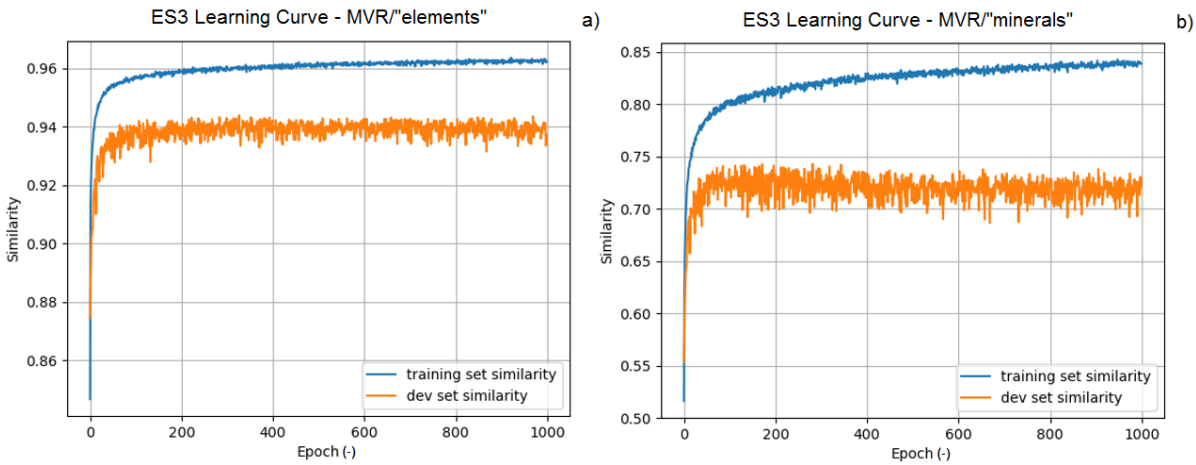

Figure 4. Learning curve for the MVR/"elements" training (a) and for the MVR/"minerals" training (b). The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (40 training sets) and the development hold-out validation dataset (1 training set) respectively.

This is also reflected in Tables 6 and 7, where the inference similarity results averaged on the three test datasets are compared for the two algorithm/task combinations. In these tables are reported also the ES3 similarities split for altitude volumes, day/night and presence of incoming H+ flux. In the following are summarized the observed implications of the Algorithms Campaign 1 results.

Implications of the tests on the MVR/"elements" algorithm/task combination (Table 6):

- Increasing the training epochs for a quarter (10/40) of the training set would reduce the bias, while not affecting the variance (increase both the training and test similarity), until reaching the optimal training time around epoch number 400 (Figure 4.a). The same observation could be made when we use half (20/40) and the total (40/40) of the training set.
- The variance is reduced by increasing the number of training examples. This is an indication of a correctly set up architecture for the applied task.

Test Set Studies - Algorithms Campaign 1 - Bias/Variance, Learning Curves												
DNN Algorithm	DNN Task	Training Set Used	Training Epochs	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVR	Elements	10/40	40	0.9613	0.9263	0.9452	0.9246	0.8978	0.9173	0.9364	0.9274	0.9257
MVR	Elements	10/40	100	0.9614	0.9267	0.9569	0.9227	0.8825	0.9197	0.9347	0.9298	0.9249
MVR	Elements	10/40	400	0.9678	0.9393	0.9629	0.9273	0.9134	0.9340	0.9452	0.9329	0.9430
MVR	Elements	10/40	1000	0.9670	0.9374	0.9623	0.9191	0.9159	0.9295	0.9462	0.9275	0.9433
MVR	Elements	20/40	40	0.9653	0.9344	0.9620	0.9305	0.8943	0.9294	0.9401	0.9363	0.9334
MVR	Elements	20/40	100	0.9686	0.9411	0.9701	0.9333	0.9027	0.9374	0.9453	0.9420	0.9407
MVR	Elements	20/40	400	0.9661	0.9345	0.9648	0.8956	0.9247	0.9220	0.9484	0.9149	0.9461
MVR	Elements	20/40	1000	0.9684	0.9397	0.9684	0.9043	0.9292	0.9318	0.9486	0.9264	0.9477
MVR	Elements	40/40	40	0.9712	0.9458	0.9723	0.9369	0.9122	0.9430	0.9488	0.9450	0.9462
MVR	Elements	40/40	100	0.9737	0.9506	0.9743	0.9386	0.9246	0.9471	0.9545	0.9480	0.9521
MVR	Elements	40/40	400	0.9750	0.9529	0.9744	0.9381	0.9332	0.9490	0.9572	0.9486	0.9555
MVR	Elements	40/40	1000	0.9737	0.9504	0.9724	0.9350	0.9304	0.9481	0.9529	0.9470	0.9524

Table 6. Algorithms campaign 1 for MVR/"elements". Shown are average test set similarities for a number of training runs - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+. Color of valued cells represents the ES3 similarity: green is higher, red is lower.

Test Set Studies - Algorithms Campaign 1 - Bias/Variance, Learning Curves												
DNN Algorithm	DNN Task	Training Set Used	Training Epochs	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVR	Minerals	10/40	40	0.8172	0.6601	0.7722	0.6536	0.4871	0.6438	0.6783	0.6750	0.6512
MVR	Minerals	10/40	100	0.8044	0.6338	0.7858	0.6183	0.4063	0.6112	0.6590	0.6406	0.6298
MVR	Minerals	10/40	400	0.8304	0.6800	0.7912	0.6083	0.5737	0.6363	0.7288	0.6440	0.7014
MVR	Minerals	10/40	1000	0.8174	0.6558	0.7756	0.5611	0.5585	0.6107	0.7061	0.6060	0.6853
MVR	Minerals	20/40	40	0.8302	0.6820	0.8138	0.6861	0.4668	0.6633	0.7028	0.6958	0.6737
MVR	Minerals	20/40	100	0.8334	0.6870	0.8136	0.6735	0.4979	0.6713	0.7046	0.7004	0.6790
MVR	Minerals	20/40	400	0.8461	0.7095	0.8113	0.6814	0.5750	0.6808	0.7417	0.7082	0.7104
MVR	Minerals	20/40	1000	0.8334	0.6859	0.7972	0.6065	0.5872	0.6422	0.7346	0.6514	0.7064
MVR	Minerals	40/40	40	0.8375	0.6966	0.8390	0.7154	0.4499	0.6912	0.7027	0.7328	0.6750
MVR	Minerals	40/40	100	0.8633	0.7417	0.8326	0.7085	0.6296	0.7156	0.7710	0.7352	0.7457
MVR	Minerals	40/40	400	0.8608	0.7371	0.8304	0.6971	0.6277	0.7098	0.7676	0.7260	0.7437
MVR	Minerals	40/40	1000	0.8615	0.7383	0.8203	0.6966	0.6487	0.7070	0.7733	0.7226	0.7476

Table 7. Algorithms campaign 1 for MVR/"minerals". Shown are average test set similarities for a number of training runs - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+.

- The test set prediction similarity is reduced with altitude. The sparsely populated exosphere at higher altitudes, and the fact that at higher altitudes with the current trajectory model the exospheric volume collects particles from multiple adjacent surface tiles, results in an increased error when reconstructing the surface from these kind of measurements.
- The test set prediction similarity is slightly worse on the dayside of the planet. This is due to more active processes, some of which expel not only source, but also ambient material.
- The test set prediction similarity is slightly worse with H+ flux present.
- By using the optimal training time and the full training set, the MVR network has succeeded

to learn a good approximation of the relationships between the features in F01 and to infer very accurately the elemental composition below, with the ES3 metric reaching 95.29 % on the test sets and 96.254 % in training.

Implications of the tests on the MVR/"minerals" algorithm/task combination (Table 7):

- With a smaller training set the overfitting is evident earlier and to a higher extent, the optimal number of training epochs is found around 150 (Figure 4.b).
- When comparing to the earlier work of Kazakov et al. (2020) that included only the MIV process in a simplified setting, the training (and testing) gives a slightly better ES1 score - 87.94 % training and 85.95 % test similarity for the former, and 90.11 % training and 86.33 % test similarity for the latter. The fact that the good similarity of the prediction on the mineral surface composition is translated to the more complex data generation model is a good indication for the future application of the method to a diverse selection of models.
- Inference at high altitudes is more complicated for the minerals task, evident by the higher loss of ES3 similarity - 80.21 % in training and 74.17 % in testing.
- A reduction of the similarity on the dayside and in the H+ flux, observed in the "elements" task, is apparent also for the "minerals" task.
- There is an observed decrease in similarity when compared to the "elements" task due to the higher number of relationships to estimate in the "minerals" task (also seen on Figure 4).

Implications of the MVRx2-1000 network for both tasks (Table A1):

- The learning curves give similar optimal training epoch number as the MVR algorithm - 400 for the "elements" and 150 for the "minerals" task.
- The two algorithms show a good reduction of the error when training on more examples - with the ES3 increasing from 92.02 % to 95.17 % on the "elements" task, and from 65.19 % to 71.55 % on the "minerals" task.
- There is no overall similarity improvement in the "elements" task and a slightly lower overall similarity for the "minerals" task than the MVR network. It should be noted that the two

parallel networks are trained each with approximately half of the total examples when they are split by altitude, which may partially explain the slight increase of the prediction error in both tasks with respect to the MVR network. This may well warrant a further investigation of this newly developed network to analyze its full potential to split the data distribution estimation with altitude.

Implications of the MVRx2-DN network for both tasks (Table A1):

- Again, the learning curves give similar optimal training epoch number as the MVR algorithm - 400 for the "elements" and 150 for the "minerals" task.
- These two algorithms also show nominal performance with increasing the examples, and the training time.
- The parallel network responsible for learning the relationships on the night side gives greater inference, but the daytime network gives the same bias and a bigger variance, signifying a potential for improvement on the inference part.
- There is the same slight reduction in overall similarity when compared to the MVR network justified by the half and half split of training examples to the two parallel networks.

Implications of the MVRx2-H+ network for both tasks (Table A1):

- Similar optimal training epoch number as the MVR algorithm - 400 for the "elements" and 150 for the "minerals" task.
- Both algorithms show a good reduction of the error when training on more examples, as expected.
- This time the training examples are split differently, H+ part of the network has about one third of the total examples, while the parallel network responsible for the reconstruction when there is no H+ flux processes the other approximately two thirds of the total examples. This results in the same slight reduction in overall similarity when compared to the MVR network.

Implications of the MVRe2m/"e2m" algorithm (Table A1):

- The learning curves give similar optimal training epoch number as the MVR/"elements" algorithm on the first neural network in the series and an optimal epoch around number 100 in the second neural network.
- The training time for the 2 networks is slightly more than for the other networks due to having to train both networks with the total number of examples.
- There is a slight improvement in the inference compared to the MVR/"minerals" algorithm that may warrant a further investigation and optimization of this more complex serial network. Another reason for this is, because it gives flexibility on both modeling the processes involved in the two steps, and in the network designs (modularity).

The test summary of the first algorithms campaign for the three subtypes of the MVRx2 network and the MVRe2m network is given in the Appendix A: Table A1.

4.2.2. Algorithms Campaign 2 - Feature Sets, TAA-M and Orbits

The second batch of tests, 174 in number, investigate the effects of different orbits, $\mathrm{TAA_M}$ at aphelion and the use of augmented feature sets for the algorithms. Here, the three base regoliths for testing are viewed for two Mercury True Anomaly Angles with Mercury at perihelion and aphelion respectively, and from three different orbits each - MPO-P/A, circular-500 and circular-1500. This results in a total of 18 different test sets. The parameters, which are kept invariant in this test campaign are the fixed training parameters, which resulted in the optimal ES3 similarities in the previous campaign, and the model parameters (base model M01). The tabular results of the Algorithms Campaign 2 are shown in Appendix A., Tables A2 and A3 and have the following implications:

Implications of the tests on the MVR/"elements" algorithm/task combination (Table A2):

- There is not a big difference in the similarity metrics between $\mathrm{TAA_M} = 0$ and $\mathrm{TAA_M} = 180$, although the aphelion measurements ($\mathrm{TAA_M} = 180$) show slightly higher errors.
- For the circular orbit of 500 km there is a slight drop in the ES3 accuracy metric between night and day measurements and "No H+" and "H+" measurements.

- As expected the test sets with a low altitude orbit of 500 km give more accurate predictions reaching ES3 of 97.29 % compared to 95.00 % for the MPO-P/A orbit and just 91.95 % for the high altitude orbit of 1500 km.
- Adding new features (in F07) doesn't affect the similarity greatly, implying also that the linear scale of the densities is not appropriate to advance further the algorithms. However, using the logarithm of the densities (in F08) improves the similarity of the inference, especially at high altitudes by as much as 2 % ES3. This would warrant a more thorough investigation on the engineering of all input features (e.g. the sun incidence angle as a cosine instead of in degrees, or even a $\log_{100}$ of the exospheric densities) as a further improvement on the method.

Implications of the tests on the MVR/"minerals" algorithm/task combination (Table A2):

- Directly inferring the minerals gives higher difference between the ES3 similarities of measurements on the night (76.76 %) and day (70.98 %) sides or in the H+ flux (72.60 %) and away from it (74.37 %).
- There is another pronounced difference between $\mathrm{TAA_M} = 0$ and $\mathrm{TAA_M} = 180$ for the day/night and H+/NoH+ measurements - a swapped similarity gap - the day side is somewhat easier to predict at aphelion than the night side, and the same goes for the open field-line measurements.
- The deterioration in the ES3 metric (with F07 set) at high altitudes indicates that the linear scales of the new features - the $\mathrm{TAA_M}$, the Sun incidence angle, and the H+ flux - are not synergistic with the linear scales of the exospheric density features (addressed in F08 set). This is an indication of the big importance of the correct scale for the input features. The F08 feature set greatly reduces the prediction error for this algorithm, especially at high altitudes, where it reaches ES3 of 71.75 % for the perihelion test datasets, up from 62.77 %.
- The algorithm trains faster for the F08 set - in approximately 100 epochs it reaches optimal training with respect to the variance of the network.

Implications of the MVRx2 and MVRe2m networks for all tasks (Table A3):

- MVRx2 all and MVRe2m all - show similar patterns of improvement for the F08 set as the former algorithms.
- However, none of the MVRx2 algorithms do not outperform the MVR network, and the MVRe2m networks show only very slight improvements.

4.2.3. Model Variations Campaign

The final 216 tests are undertaken in order to examine the range of the DNNs inference predictions when the models for the test data generation is differing from the one used for generation of the training dataset. A total of 25 different model variations are evaluated for the three combinations *MVR/"elements"*, *MVR/"minerals"* and *MVRe2m/"e2m"*. The fixed parameters in this campaign are the training parameters found from the Algorithms Campaign 1 and the feature set found from the Algorithm Campaign 2, as well as the orbit (fixed to MPO-P/A) and the $\mathrm{TAA_M} = 0$. The tabular results are shown in Appendix A., Tables A4, A5 and A6, where each test run for the different model variations is compared to the reference M01, and the implications are:

- The *MVR/"elements"* algorithm shows a narrow range of results in terms of ES3 in the variations of the present exosphere generation model (Table A4, especially when considering a single parameter variation applied for all species (M02-M05 and M08-M13). This implies that when the ratio between the effective temperatures of the different species or between their process efficiencies remains close to the base model, the algorithms are still able to accurately reconstruct the surface element composition (Table A4).
- There is a very slight reduction in accuracy (ES3, all algorithms) when the mean temperature of the MIV vapor is varied (both up and down, M02 and M03) - this shows that the DNN was able to find the correct relation between the ratio of the elemental densities.
- There is a slight reduction in accuracy (ES3) when the mean temperature of the SP released atoms is varied (M04 and M05) - especially noticeable in the H+ flux regions in all maps. This is more pronounced with the *MVR/"minerals"* and the *MVRe2m/"e2m"* algorithms. The reason for this is that the ratios between the different species' altitude distributions are altered with the release temperature.

- For the *MVR/"minerals"* combination the error becomes bigger, especially when varying the temperature of the processes (hence the surface velocity of the ejected atoms, M06 and M07) (Table A5).
- When reducing the SP yields (M10) in the same ratio for all species, the prediction accuracy improves, implying that the algorithms had learned to better estimate the ever-present MIV related vapor than the ejecta from the SP. This is valid for all the algorithms.
- When we use for the ambient population the mean of the regolith (M17), instead of more closely reflecting the source, the dayside/nightside asymmetry in the predictions is more pronounced.
- When simulating an increase in the area of the magnetic cusp footprints by CME (M18), both the dayside and the H+ part show reduced accuracy.
- The MVRe2m network seems that is not as robust to the model variations as the MVR network at its present configuration/setting. In fact, the ES3 values decrease by up to 15 % (M25) when parameters are varied. This may warrant a further investigation on this newly developed DNN.
- The accuracy is dropped again, when the SP has varying yields for each species (M21-M23) - especially, when they reflect their binding energies (although, this is because the training was performed with a uniform SP yield across the table). This may warrant an improvement of the base model to inspect further.
- When there is only the MIV acting on the test map (M20), there are mispredictions on the dayside. This is probably due to the presence in the F08 set of the sun incidence angle feature. This may warrant the engineering of more advanced feature sets in a future iteration of this model.

4.3. Reconstruction of the Surface Composition

After we have examined the tabular results of the test campaigns in terms of the accuracy metrics defined above, we proceed to show the visual representation of the Deep Neural Network's predictions. By running the "static-swipe" observation of the exosphere, the algorithms are able to generate a predictions map for each surface tile, or in other words to reconstruct the surface composition maps. As an example, Figure 5.b,e,h shows the true elemental relative composition

maps of the Al, Ca and Mg on the surface, and their predicted values from the best inference made with the MVR/"elements" algorithm (Figure 5.a,d,g) and the worst inference with highly different model parameters (M25) of the data generation model (Figure 5.c,f,i).

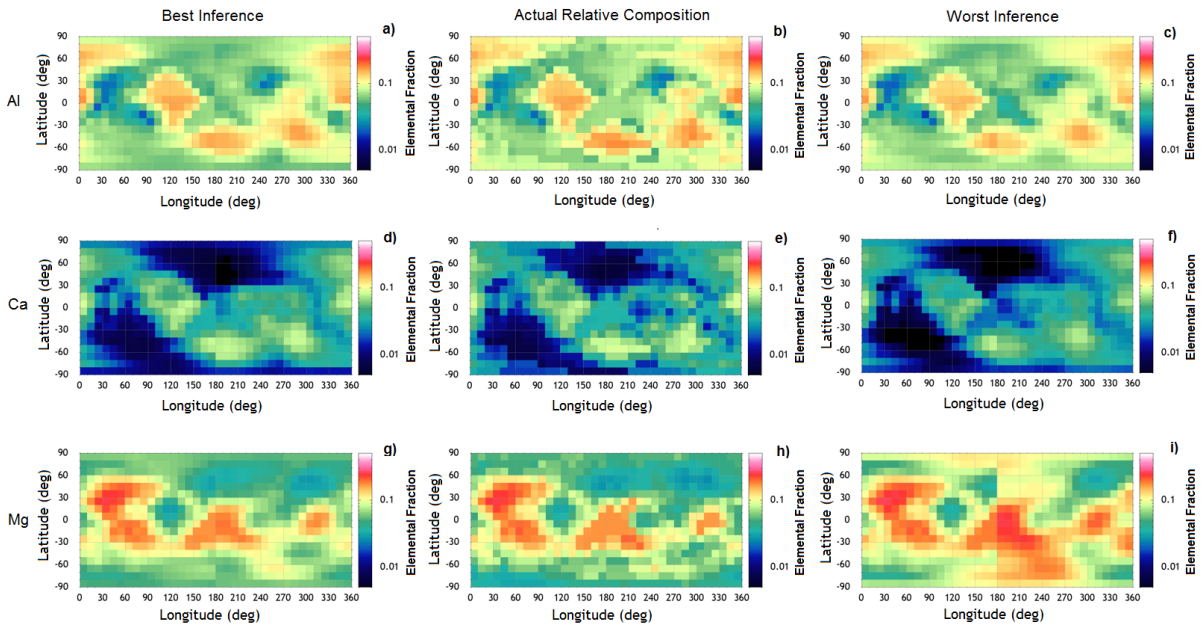

Figure 5. Examples of the maps (longitudes and latitudes) of the surface relative elemental compositions obtained by the DNN for Aluminium (first row: panels a, b, c), Calcium (second row: d, e, f), and Magnesium (third row: g, h, i). Middle panels (b, e, h): Actual ("ground truth") compositions. Left panels (a, d, g): Predicted by MVR/"elements" (M01) - best inference. Right panels (c, f, i): Predicted by MVR/"elements" (M25) - worst inference.

Figure 6.b,e,h shows the true mineral relative composition maps of the Anorthite, Enstatite and Albite minerals on the surface, and their predicted values from the best inference of the MVR/"minerals" algorithm (Figure 6.a,d,g) and the worst inference with highly different process parameters (M24) of the data generation model (Figure 6.c,f,i). Displayed on Figure 7 are the respective mean altitude of the measurements (b, e) and the respective test ES3 similarity maps for the elemental composition best (a) and worst (c) inference, and for the mineral composition best (d) and worst (f) inference.

We could observe that on both the elemental (Figure 5) and on the mineral (Figure 6) maps the predicted surface composition is very closely matching the actual one. Some of the implications mentioned earlier from the tabular analysis of the ES3 results in the test campaigns are also visible on the maps - the altitude and cusp footprint increase in errors are evident especially for the test datasets, which are using a more distant variation of the model parameters

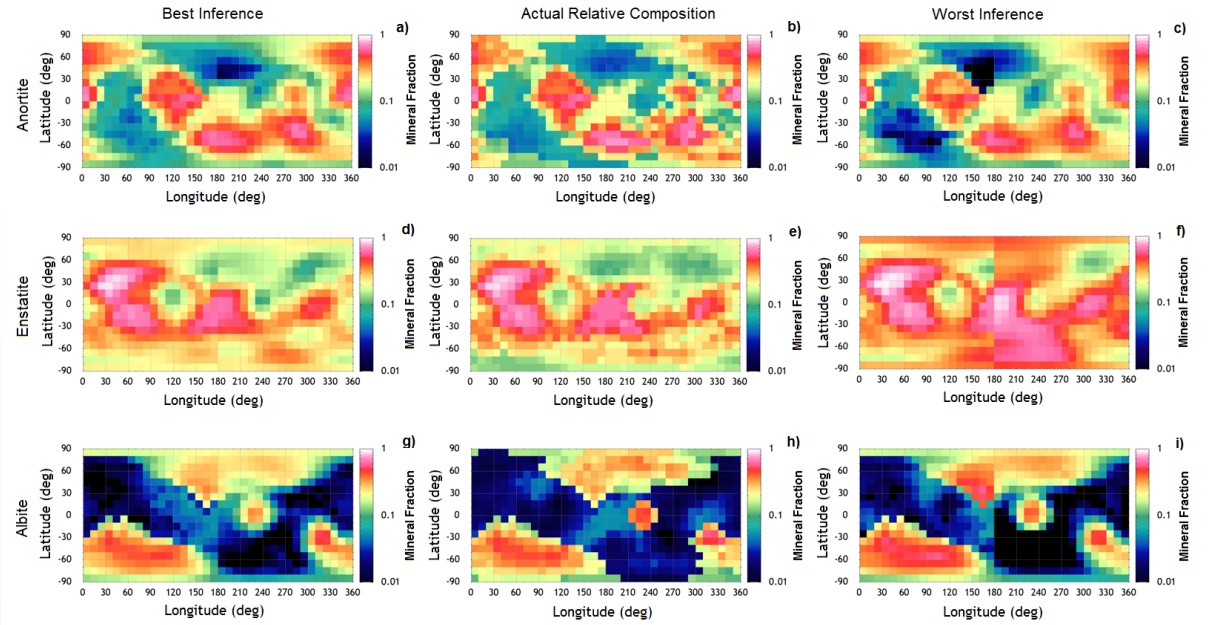

Figure 6. Examples of the maps (longitudes and latitudes) of the surface relative mineral compositions obtained by the DNN for Anorthite (first row: panels a, b, c), Enstatite (second row: d, e, f), and Albite (third row: g, h, i). Middle panels (b, e, h): Actual ("ground truth") compositions. Left panels (a, d, g): Predicted by MVR/"minerals" (M01) - best inference. Right panels (c, f, i): Predicted by MVR/"minerals" (M24) - worst inference.

compared to the M01. The cusp footprint higher error is explained by the presence of more surface release processes in that region.

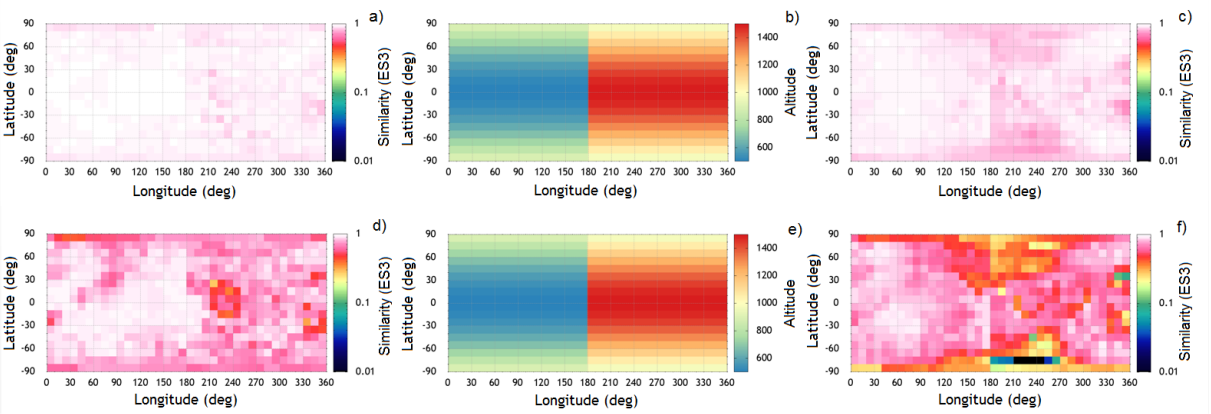

Figure 7. Middle panels: Mean orbital altitude of measurements in the "elements" (b) and "minerals" (e) tests. Left panels: ES3 similarity maps for the "elements" tests that give the best (a, M01) and worst (d, M25) inference. Right panels: ES3 similarity maps for the "minerals" tests that give the best (c, M01) and worst (f, M24) inference.

5. Conclusion

This study has shown a novel method for the reconstruction of the surface composition by utilizing measurements of the exospheric densities in orbit around Mercury as inputs to Deep Neural Networks within a multivariate regression inference task. The algorithms have been greatly improved over the earlier work of Kazakov et al. (2020), through the utilization of Bayesian hyper-parameter tuning and domain knowledge on the feature sets. We have also developed two DNN architectures in addition to the MVR DNN architecture, which resulted in a total of nine algorithms for the surface reconstruction. At the same time, elaborating the models to construct more plausible datasets has not resulted in decreased capability to accurately predict the mineralogy. The mineral composition prediction (task "minerals") is also supplemented by two new tasks - the prediction of the elemental composition ("elements"), and the sequential prediction of both elemental and mineral composition ("e2m"). The algorithms for all three resulting reconstruction tasks give good similarity on their inference compared with the actual surface composition. The similarity metric (ES3) for the "elements" task reaches 95.72 %, while for the "minerals" and "e2m" tasks give 77.10 % and 77.29 % accurate inference respectively. The reconstructed relative elemental and mineral composition maps confirm the adequacy of the algorithms, especially applied on virtual measurements of low altitudes. This has promising implications for an application of the method to constrain the models for the generation of the exosphere when real data becomes available, by comparing the predicted surface composition from several models to the composition measured by imagers such as BepiColombo/MPO/MIXS, BepiColombo/MPO/MGNS, BepiColombo/MPO/MERTIS and BepiColombo/MPO/SIMBIO-SYS [4]. However, our tentative method requires further development to put it in line with the real complexity of the processes involved in the interaction between the surface and the exosphere. Such developments would aim to eliminate the present limitations related to: (1) the simplified assumptions on the release mechanisms of each process in the exospheric generation model; (2) the exclusion of processes in the exosphere that could produce accelerations, loss or chemical variation of the particles; (3) the exclusion of magnetospheric models, which would reproduce relevant dynamics for highly variable in time processes, such as the ion sputtering.

As implied by the training and test campaigns undertaken with this work, there is a wide frontier for improvements of the DNN surface reconstruction algorithms. Thus, we have drawn

a roadmap for their future development in order to address the requirements for the application to real measurements in orbit around Mercury:

- Most essentially, we aim at integrating the algorithms within the SERENA analysis software by designing interfaces between the method and other more realistic models for the simulation of the Hermean environment. This way we would be able to train the DNNs with data generated from external simulations and compare the predictions for each of the models.
- We have noted that more training examples would improve the variance of the test set prediction, which is still high for the *"minerals"* and *"e2m"* tasks - about 9 % ES3, compared to just 1 % for the "elements" task. We could exploit this potential for improving the reconstruction of the mineralogy in a future iteration.
- The big improvements introduced with the feature set that uses the logarithm of the exospheric measurements (F08) is prompting us to consider further feature engineering also for some other input parameters, such as the altitude of the measurement or the sun direction.
- Another route for exploration of the capabilities of the algorithms warrants the further development of parallel DNN architectures - the MVRx4 and MVRx8 - combining the three subtypes of the MVRx2 to split the generation processes into different data subsets. This way we could characterize separately the data distributions, when different processes are acting on the surface to release neutral particles to the exosphere.
- One more path for extending the Deep Neural Network method would be to backtrack the released particles fluxes as a first step and then reconstruct the surface elemental and mineral compositions - a task dubbed *"f2e2m"*.

While still in a tentative state, the inclusion of this method in the data analysis module of the SERENA ground system is well underway with an advanced AI algorithm aimed to aid the scientific mission in the search for clues about Mercury's exosphere origins and dynamics. Further sophistication of the DNNs presented here have a big potential to give us new tools to investigate and describe not only the Mercury environment, but also the physical processes on other planetary bodies in our Solar System.

Acknowledgments

This work was supported by the Italian Space Agency (ASI) - SERENA contract 2018-8-HH.O "Scientific participation in the mission BepiColombo SERENA - Phase E1".

Appendix A. Training and Test Campaign Supplementary Tabular Results

Test Set Studies - Algorithms Campaign 1 - Bias/Variance, Learning Curves												
DNN Algorithm	DNN Task	Training Set Used	Training Epochs	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVRx2-1000	Elements	10/40	400	0.9589	0.9202	0.9682	0.8627	0.9009	0.8972	0.9459	0.8857	0.9407
MVRx2-1000	Elements	20/40	400	0.9624	0.9265	0.9720	0.8654	0.9147	0.9043	0.9513	0.8924	0.9468
MVRx2-1000	Elements	40/40	400	0.9744	0.9517	0.9737	0.9364	0.9319	0.9447	0.9596	0.9461	0.9551
MVRx2-DN	Elements	10/40	400	0.9659	0.9345	0.9607	0.9195	0.9076	0.9261	0.9439	0.9263	0.9394
MVRx2-DN	Elements	20/40	400	0.9690	0.9415	0.9693	0.9230	0.9157	0.9355	0.9483	0.9376	0.9438
MVRx2-DN	Elements	40/40	400	0.9723	0.9477	0.9702	0.9328	0.9266	0.9408	0.9490	0.9432	0.9504
MVRx2-H+	Elements	10/40	400	0.9670	0.9372	0.9603	0.9224	0.9150	0.9280	0.9475	0.9283	0.9426
MVRx2-H+	Elements	20/40	400	0.9683	0.9401	0.9687	0.9214	0.9132	0.9365	0.9442	0.9357	0.9428
MVRx2-H+	Elements	40/40	400	0.9720	0.9471	0.9734	0.9323	0.9198	0.9421	0.9527	0.9377	0.9527
MVRx2-1000	Minerals	10/40	400	0.8146	0.6519	0.8004	0.5594	0.5068	0.5912	0.7198	0.5981	0.6840
MVRx2-1000	Minerals	20/40	400	0.8201	0.6616	0.8163	0.5732	0.5024	0.6064	0.7233	0.6266	0.6824
MVRx2-1000	Minerals	40/40	400	0.8489	0.7155	0.8215	0.6836	0.5780	0.6706	0.7658	0.6983	0.7258
MVRx2-DN	Minerals	10/40	400	0.8226	0.6675	0.7862	0.6187	0.5266	0.6293	0.7104	0.6494	0.6784
MVRx2-DN	Minerals	20/40	400	0.8355	0.6888	0.8031	0.6526	0.5420	0.6506	0.7315	0.6825	0.6925
MVRx2-DN	Minerals	40/40	400	0.8376	0.6951	0.8102	0.6566	0.5493	0.6502	0.7452	0.6785	0.7049
MVRx2-H+	Minerals	10/40	400	0.8198	0.6590	0.7840	0.6039	0.5142	0.6313	0.6901	0.6325	0.6749
MVRx2-H+	Minerals	20/40	400	0.8378	0.6934	0.8069	0.6574	0.5478	0.6555	0.7357	0.6782	0.7024
MVRx2-H+	Minerals	40/40	400	0.8537	0.7249	0.8245	0.6942	0.5961	0.6970	0.7561	0.6893	0.7460
MVRe2m	E2M	10/40	500	0.8335	0.6843	0.7893	0.6268	0.5741	0.6542	0.7180	0.6551	0.7018
MVRe2m	E2M	20/40	500	0.8357	0.6881	0.7887	0.6024	0.6127	0.6529	0.7274	0.6520	0.7096
MVRe2m	E2M	40/40	500	0.8689	0.7517	0.8280	0.7183	0.6632	0.7288	0.7774	0.7405	0.7584

Table A1. Algorithms campaign 1 for MVRx2 and MVRe2m algorithms (all combinations). Shown are average test set similarities for a number of training runs - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+.

Test Set Studies - Algorithms Campaign 2 - Feature Sets, TAA_M and Orbits													
DNN Algorithm	DNN Task	Feature Set	TAA_M	Orbit	Test Set Similarity								
					Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
					ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVR	Elements	F01	0	P/A	0.9750	0.9529	0.9744	0.9381	0.9332	0.9490	0.9572	0.9486	0.9555
MVR	Elements	F01	180	P/A	0.9720	0.9472	0.9725	0.9330	0.9208	0.9687	0.9230	0.9607	0.9426
MVR	Elements	F01	any	P/A	0.9735	0.9500	0.9735	0.9356	0.9270	0.9589	0.9401	0.9547	0.9491
MVR	Elements	F01	any	500	0.9857	0.9729	0.9729	-	-	0.9685	0.9781	0.9618	0.9781
MVR	Elements	F01	any	1500	0.9573	0.9195	-	-	0.9195	0.9204	0.9185	0.9206	0.9200
MVR	Elements	F07	0	P/A	0.9711	0.9458	0.9742	0.9327	0.9134	0.9426	0.9493	0.9461	0.9456
MVR	Elements	F07	180	P/A	0.9610	0.9266	0.9707	0.9135	0.8692	0.9668	0.8817	0.9584	0.9160
MVR	Elements	F07	any	P/A	0.9661	0.9362	0.9725	0.9231	0.8913	0.9547	0.9155	0.9523	0.9308
MVR	Elements	F07	any	500	0.9842	0.9701	0.9701	-	-	0.9652	0.9756	0.9591	0.9755
MVR	Elements	F07	any	1500	0.9380	0.8832	-	-	0.8832	0.8951	0.8700	0.9011	0.8761
MVR	Elements	F08	0	P/A	0.9781	0.9589	0.9768	0.9387	0.9504	0.9579	0.9600	0.9517	0.9632
MVR	Elements	F08	180	P/A	0.9766	0.9555	0.9743	0.9386	0.9423	0.9706	0.9387	0.9624	0.9532
MVR	Elements	F08	any	P/A	0.9773	0.9572	0.9756	0.9387	0.9464	0.9642	0.9493	0.9571	0.9582
MVR	Elements	F08	any	500	0.9867	0.9748	0.9748	-	-	0.9736	0.9762	0.9689	0.9776
MVR	Elements	F08	any	1500	0.9647	0.9334	-	-	0.9334	0.9350	0.9317	0.9320	0.9344
MVR	Minerals	F01	0	P/A	0.8608	0.7371	0.8304	0.6971	0.6277	0.7098	0.7676	0.7260	0.7437
MVR	Minerals	F01	180	P/A	0.8544	0.7266	0.8223	0.6869	0.6132	0.8086	0.6350	0.7788	0.7092
MVR	Minerals	F01	any	P/A	0.8576	0.7319	0.8264	0.6920	0.6205	0.7592	0.7013	0.7524	0.7265
MVR	Minerals	F01	any	500	0.9080	0.8232	0.8232	-	-	0.7961	0.8535	0.7770	0.8442
MVR	Minerals	F01	any	1500	0.7745	0.5752	-	-	0.5752	0.5647	0.5870	0.5664	0.5803
MVR	Minerals	F07	0	P/A	0.8476	0.7105	0.8241	0.6790	0.5602	0.6913	0.7320	0.7162	0.7071
MVR	Minerals	F07	180	P/A	0.8197	0.6634	0.8120	0.6081	0.4809	0.7936	0.5179	0.7565	0.6324
MVR	Minerals	F07	any	P/A	0.8337	0.6869	0.8181	0.6436	0.5206	0.7424	0.6250	0.7364	0.6697
MVR	Minerals	F07	any	500	0.9018	0.8113	0.8113	-	-	0.7877	0.8377	0.7679	0.8323
MVR	Minerals	F07	any	1500	0.7088	0.4496	-	-	0.4496	0.4778	0.4182	0.4585	0.4381
MVR	Minerals	F08	0	P/A	0.8802	0.7710	0.8414	0.7120	0.7175	0.7590	0.7845	0.7480	0.7847
MVR	Minerals	F08	180	P/A	0.8714	0.7564	0.8361	0.7098	0.6753	0.8216	0.6835	0.7801	0.7485
MVR	Minerals	F08	any	P/A	0.8758	0.7637	0.8388	0.7109	0.6964	0.7903	0.7340	0.7641	0.7666
MVR	Minerals	F08	any	500	0.9110	0.8289	0.8289	-	-	0.8155	0.8440	0.7966	0.8430
MVR	Minerals	F08	any	1500	0.8074	0.6358	-	-	0.6358	0.6332	0.6388	0.6150	0.6457

Table A2. Algorithms campaign 2 for the MVR algorithm (both tasks). Shown are average test set similarities inferred with the best training run for each task - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+. Each row represents a set of tests with different feature set, $\mathrm{TAA_M}$ or orbit.

Test Set Studies - Algorithms Campaign 2 - Feature Sets, TAA_M and Orbits													
DNN Algorithm	DNN Task	Feature Set	TAA_M	Orbit	Test Set Similarity								
					Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
					ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVRx2-1000	Elements	F01	0	P/A	0.9744	0.9517	0.9737	0.9364	0.9319	0.9447	0.9596	0.9461	0.9551
MVRx2-1000	Elements	F01	180	P/A	0.9749	0.9525	0.9744	0.9374	0.9326	0.9711	0.9317	0.9627	0.9492
MVRx2-1000	Elements	F08	0	P/A	0.9771	0.9568	0.9775	0.9354	0.9450	0.9544	0.9594	0.9492	0.9613
MVRx2-1000	Elements	F08	180	P/A	0.9751	0.9526	0.9759	0.9329	0.9351	0.9721	0.9309	0.9621	0.9495
MVRx2-DN	Elements	F01	0	P/A	0.9723	0.9477	0.9702	0.9328	0.9266	0.9408	0.9490	0.9432	0.9504
MVRx2-DN	Elements	F01	180	P/A	0.9728	0.9485	0.9693	0.9354	0.9281	0.9651	0.9298	0.9595	0.9447
MVRx2-DN	Elements	F08	0	P/A	0.9760	0.9549	0.9759	0.9304	0.9457	0.9532	0.9567	0.9461	0.9601
MVRx2-DN	Elements	F08	180	P/A	0.9767	0.9558	0.9735	0.9393	0.9438	0.9692	0.9407	0.9598	0.9544
MVRx2-H+	Elements	F01	0	P/A	0.9720	0.9471	0.9734	0.9323	0.9198	0.9421	0.9527	0.9377	0.9527
MVRx2-H+	Elements	F01	180	P/A	0.9677	0.9394	0.9718	0.9350	0.8917	0.9685	0.9069	0.9574	0.9333
MVRx2-H+	Elements	F08	0	P/A	0.9765	0.9557	0.9752	0.9354	0.9448	0.9522	0.9597	0.9438	0.9628
MVRx2-H+	Elements	F08	180	P/A	0.9754	0.9531	0.9698	0.9359	0.9436	0.9666	0.9380	0.9576	0.9516
MVRx2-1000	Minerals	F01	0	P/A	0.8489	0.7155	0.8215	0.6836	0.5780	0.6706	0.7658	0.6983	0.7258
MVRx2-1000	Minerals	F01	180	P/A	0.8529	0.7240	0.8226	0.7057	0.5845	0.8094	0.6285	0.7748	0.7071
MVRx2-1000	Minerals	F08	0	P/A	0.8733	0.7584	0.8391	0.6960	0.6919	0.7367	0.7827	0.7304	0.7751
MVRx2-1000	Minerals	F08	180	P/A	0.8616	0.7397	0.8330	0.7028	0.6276	0.8177	0.6527	0.7764	0.7275
MVRx2-DN	Minerals	F01	0	P/A	0.8376	0.6951	0.8102	0.6566	0.5493	0.6502	0.7452	0.6785	0.7049
MVRx2-DN	Minerals	F01	180	P/A	0.8326	0.6871	0.8008	0.6674	0.5246	0.7848	0.5779	0.7570	0.6637
MVRx2-DN	Minerals	F08	0	P/A	0.8763	0.7644	0.8381	0.7080	0.7026	0.7435	0.7876	0.7405	0.7786
MVRx2-DN	Minerals	F08	180	P/A	0.8652	0.7457	0.8248	0.7021	0.6629	0.8111	0.6727	0.7776	0.7351
MVRx2-H+	Minerals	F01	0	P/A	0.8537	0.7249	0.8245	0.6942	0.5961	0.6970	0.7561	0.6893	0.7460
MVRx2-H+	Minerals	F01	180	P/A	0.8460	0.7121	0.8176	0.6948	0.5609	0.8050	0.6083	0.7518	0.6989
MVRx2-H+	Minerals	F08	0	P/A	0.8715	0.7556	0.8357	0.6976	0.6875	0.7357	0.7778	0.7159	0.7792
MVRx2-H+	Minerals	F08	180	P/A	0.8674	0.7494	0.8215	0.7142	0.6691	0.8088	0.6829	0.7618	0.7452
MVRe2m	E2M	F01	0	P/A	0.8632	0.7416	0.8362	0.7175	0.6144	0.7268	0.7582	0.7445	0.7399
MVRe2m	E2M	F01	180	P/A	0.8447	0.7077	0.8309	0.6840	0.5342	0.8162	0.5864	0.7800	0.6836
MVRe2m	E2M	F08	0	P/A	0.8809	0.7729	0.8453	0.7130	0.7170	0.7564	0.7914	0.7558	0.7831
MVRe2m	E2M	F08	180	P/A	0.8723	0.7580	0.8329	0.7170	0.6790	0.8207	0.6878	0.7841	0.7492

Table A3. Algorithms campaign 2 MVRx2 and MVRe2m algorithms (all combinations). Shown are average test set similarities inferred with the best training run for each task - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+. Each row represents a set of tests with different feature set, TAA_M or orbit.

Test Set Studies - Model Variations Campaign												
DNN Algorithm	DNN Task	Feature Set	Model Number	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVR	Elements	F08	M01	0.9781	0.9589	0.9768	0.9387	0.9504	0.9579	0.9600	0.9517	0.9632
MVR	Elements	F08	M02	0.9730	0.9494	0.9642	0.9347	0.9402	0.9496	0.9491	0.9473	0.9506
MVR	Elements	F08	M03	0.9768	0.9564	0.9752	0.9384	0.9443	0.9564	0.9564	0.9523	0.9589
MVR	Elements	F08	M04	0.9751	0.9528	0.9743	0.9230	0.9483	0.9465	0.9599	0.9350	0.9634
MVR	Elements	F08	M05	0.9694	0.9426	0.9729	0.9061	0.9308	0.9329	0.9535	0.9209	0.9555
MVR	Elements	F08	M06	0.9705	0.9444	0.9616	0.9239	0.9373	0.9394	0.9500	0.9327	0.9514
MVR	Elements	F08	M07	0.9690	0.9418	0.9715	0.9104	0.9257	0.9330	0.9516	0.9239	0.9525
MVR	Elements	F08	M08	0.9731	0.9552	0.9700	0.9369	0.9500	0.9558	0.9546	0.9506	0.9580
MVR	Elements	F08	M09	0.9763	0.9555	0.9715	0.9387	0.9467	0.9548	0.9563	0.9505	0.9585
MVR	Elements	F08	M10	0.9784	0.9595	0.9774	0.9397	0.9506	0.9572	0.9620	0.9519	0.9640
MVR	Elements	F08	M11	0.9782	0.9591	0.9773	0.9383	0.9509	0.9576	0.9607	0.9518	0.9635
MVR	Elements	F08	M12	0.9776	0.9580	0.9764	0.9380	0.9486	0.9563	0.9600	0.9517	0.9618
MVR	Elements	F08	M13	0.9774	0.9576	0.9748	0.9390	0.9487	0.9555	0.9600	0.9502	0.9620
MVR	Elements	F08	M14	0.9748	0.9527	0.9649	0.9348	0.9512	0.9492	0.9566	0.9434	0.9583
MVR	Elements	F08	M15	0.9757	0.9545	0.9693	0.9390	0.9462	0.9528	0.9563	0.9486	0.9579
MVR	Elements	F08	M16	0.9783	0.9591	0.9777	0.9366	0.9519	0.9587	0.9596	0.9514	0.9637
MVR	Elements	F08	M17	0.9759	0.9548	0.9721	0.9363	0.9457	0.9502	0.9599	0.9461	0.9600
MVR	Elements	F08	M18	0.9751	0.9532	0.9745	0.9317	0.9406	0.9501	0.9566	0.9467	0.9654
MVR	Elements	F08	M19	0.9779	0.9585	0.9762	0.9386	0.9501	0.9577	0.9593	0.9514	0.9627
MVR	Elements	F08	M20	0.9721	0.9476	0.9705	0.9327	0.9259	0.9333	0.9635	0.9356	0.9548
MVR	Elements	F08	M21	0.9778	0.9583	0.9766	0.9381	0.9491	0.9568	0.9600	0.9505	0.9630
MVR	Elements	F08	M22	0.9775	0.9577	0.9770	0.9338	0.9507	0.9556	0.9601	0.9483	0.9634
MVR	Elements	F08	M23	0.9664	0.9365	0.9710	0.8857	0.9320	0.9238	0.9507	0.9023	0.9569
MVR	Elements	F08	M24	0.9640	0.9324	0.9520	0.9077	0.9257	0.9181	0.9484	0.9085	0.9466
MVR	Elements	F08	M25	0.9582	0.9217	0.9469	0.8855	0.9168	0.9088	0.9355	0.8924	0.9387

Table A4. Model Variations Campaign MVR/"elements" combination. Shown are average test set similarities for each model variation - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+.

Test Set Studies - Model Variations Campaign												
DNN Algorithm	DNN Task	Feature Set	Model Number	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVR	Minerals	F08	M01	0.8802	0.7710	0.8414	0.7120	0.7175	0.7590	0.7845	0.7480	0.7847
MVR	Minerals	F08	M02	0.8587	0.7309	0.7966	0.6951	0.6615	0.7279	0.7341	0.7312	0.7306
MVR	Minerals	F08	M03	0.8676	0.7481	0.8335	0.7065	0.6531	0.7447	0.7519	0.7468	0.7489
MVR	Minerals	F08	M04	0.8647	0.7404	0.8313	0.6284	0.7070	0.7055	0.7794	0.6698	0.7825
MVR	Minerals	F08	M05	0.8346	0.6855	0.8282	0.5466	0.5962	0.6359	0.7410	0.5949	0.7396
MVR	Minerals	F08	M06	0.8446	0.7031	0.7856	0.6272	0.6469	0.6774	0.7317	0.6586	0.7296
MVR	Minerals	F08	M07	0.8269	0.6720	0.8211	0.5676	0.5377	0.6297	0.7192	0.6060	0.7113
MVR	Minerals	F08	M08	0.8698	0.7525	0.8126	0.7030	0.7059	0.7485	0.7569	0.7431	0.7581
MVR	Minerals	F08	M09	0.8742	0.7595	0.8288	0.7097	0.6983	0.7498	0.7703	0.7435	0.7690
MVR	Minerals	F08	M10	0.8810	0.7724	0.8456	0.7120	0.7156	0.7558	0.7909	0.7498	0.7859
MVR	Minerals	F08	M11	0.8807	0.7717	0.8434	0.7105	0.7183	0.7578	0.7873	0.7485	0.7856
MVR	Minerals	F08	M12	0.8785	0.7678	0.8326	0.7096	0.7224	0.7525	0.7850	0.7447	0.7816
MVR	Minerals	F08	M13	0.8745	0.7607	0.8302	0.7124	0.6978	0.7399	0.7839	0.7373	0.7747
MVR	Minerals	F08	M14	0.8674	0.7468	0.7916	0.6952	0.7267	0.7305	0.7650	0.7192	0.7633
MVR	Minerals	F08	M15	0.8693	0.7506	0.8173	0.7096	0.6849	0.7333	0.7700	0.7316	0.7619
MVR	Minerals	F08	M16	0.8794	0.7698	0.8455	0.7061	0.7123	0.7612	0.7793	0.7475	0.7830
MVR	Minerals	F08	M17	0.8680	0.7483	0.8116	0.7024	0.6928	0.7154	0.7851	0.7199	0.7653
MVR	Minerals	F08	M18	0.8639	0.7411	0.8326	0.6750	0.6607	0.7125	0.7730	0.7064	0.8058
MVR	Minerals	F08	M19	0.8793	0.7693	0.8405	0.7110	0.7136	0.7583	0.7816	0.7466	0.7828
MVR	Minerals	F08	M20	0.8481	0.7138	0.7832	0.7128	0.6037	0.6425	0.7935	0.6863	0.7302
MVR	Minerals	F08	M21	0.8720	0.7552	0.8359	0.6818	0.6995	0.7378	0.7747	0.7168	0.7781
MVR	Minerals	F08	M22	0.8765	0.7642	0.8413	0.6849	0.7202	0.7459	0.7846	0.7263	0.7868
MVR	Minerals	F08	M23	0.8560	0.7240	0.8283	0.6010	0.6789	0.6854	0.7664	0.6392	0.7740
MVR	Minerals	F08	M24	0.8144	0.6486	0.7252	0.5896	0.5851	0.5812	0.7240	0.5634	0.6994
MVR	Minerals	F08	M25	0.8165	0.6495	0.7259	0.5644	0.6123	0.6038	0.7005	0.5630	0.7011

Table A5. Model Variations Campaign MVR/"minerals" combination. Shown are average test set similarities for each model variation - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+.

Test Set Studies - Model Variations Campaign												
DNN Algorithm	DNN Task	Feature Set	Model Number	Test Set Similarity								
				Full Set		With Altitude (ES3)			Day/Night (ES3)		With H+ Flux (ES3)	
				ES1	ES3	500-850km	850-1275km	1275-1500km	Day	Night	H+	No H+
MVRe2m	E2M	F08	M01	0.8809	0.7729	0.8453	0.7130	0.7170	0.7564	0.7914	0.7558	0.7831
MVRe2m	E2M	F08	M02	0.8542	0.7220	0.7808	0.6913	0.6586	0.7235	0.7203	0.7236	0.7211
MVRe2m	E2M	F08	M03	0.8718	0.7547	0.8342	0.7062	0.6762	0.7453	0.7653	0.7488	0.7582
MVRe2m	E2M	F08	M04	0.8648	0.7401	0.8252	0.6338	0.7105	0.7055	0.7789	0.6724	0.7806
MVRe2m	E2M	F08	M05	0.8390	0.6935	0.8214	0.5664	0.6159	0.6419	0.7512	0.6078	0.7446
MVRe2m	E2M	F08	M06	0.8405	0.6956	0.7674	0.6343	0.6421	0.6762	0.7173	0.6548	0.7199
MVRe2m	E2M	F08	M07	0.8357	0.6879	0.8190	0.5897	0.5763	0.6401	0.7412	0.6218	0.7272
MVRe2m	E2M	F08	M08	0.8722	0.7562	0.8114	0.7013	0.7229	0.7508	0.7623	0.7427	0.7642
MVRe2m	E2M	F08	M09	0.8722	0.7550	0.8186	0.7074	0.7011	0.7431	0.7685	0.7407	0.7636
MVRe2m	E2M	F08	M10	0.8812	0.7723	0.8403	0.7152	0.7204	0.7539	0.7928	0.7516	0.7845
MVRe2m	E2M	F08	M11	0.8808	0.7711	0.8395	0.7093	0.7235	0.7553	0.7888	0.7495	0.7840
MVRe2m	E2M	F08	M12	0.8782	0.7667	0.8366	0.7056	0.7159	0.7494	0.7860	0.7487	0.7774
MVRe2m	E2M	F08	M13	0.8762	0.7631	0.8276	0.7100	0.7128	0.7425	0.7859	0.7371	0.7785
MVRe2m	E2M	F08	M14	0.8673	0.7464	0.7901	0.6953	0.7273	0.7253	0.7699	0.7163	0.7642
MVRe2m	E2M	F08	M15	0.8691	0.7494	0.8079	0.7081	0.6971	0.7323	0.7684	0.7299	0.7610
MVRe2m	E2M	F08	M16	0.8837	0.7769	0.8440	0.7120	0.7348	0.7699	0.7848	0.7550	0.7900
MVRe2m	E2M	F08	M17	0.8668	0.7455	0.8087	0.6975	0.6924	0.7102	0.7849	0.7167	0.7626
MVRe2m	E2M	F08	M18	0.8665	0.7462	0.8347	0.6754	0.6753	0.7228	0.7723	0.7167	0.8013
MVRe2m	E2M	F08	M19	0.8793	0.7684	0.8371	0.7074	0.7196	0.7549	0.7835	0.7453	0.7822
MVRe2m	E2M	F08	M20	0.8485	0.7132	0.8098	0.6712	0.6005	0.6416	0.7932	0.6687	0.7397
MVRe2m	E2M	F08	M21	0.8759	0.7620	0.8352	0.6989	0.7079	0.7468	0.7789	0.7347	0.7782
MVRe2m	E2M	F08	M22	0.8777	0.7657	0.8386	0.6896	0.7252	0.7475	0.7860	0.7345	0.7843
MVRe2m	E2M	F08	M23	0.8431	0.6998	0.8179	0.5612	0.6494	0.6526	0.7524	0.5960	0.7616
MVRe2m	E2M	F08	M24	0.8154	0.6498	0.7176	0.5976	0.5935	0.5746	0.7339	0.5631	0.7015
MVRe2m	E2M	F08	M25	0.8007	0.6210	0.7173	0.5128	0.5751	0.5819	0.6646	0.5252	0.6780

Table A6. Model Variations Campaign MVRe2m/"e2m" combination. Shown are average test set similarities for each model variation - ES1 and ES3 for the full dataset, and ES3 only for the dataset examples split by altitude, day/night, H+/No H+.

Appendix B. Range/Interval of the Predictions

To supplement the tabular results of the model variations test campaign, we display for each element and mineral the range for their predictions by the MVR/"elements" and the MVR/"minerals" algorithm respectively, when using datasets created with different model variations. In Figure B1 are shown the 11 elemental surface plots with their prediction ranges and best fits over each tile of the surface. The largest prediction interval is observed for the elements Fe and Mg - about ±5 to ±8 % of the total fraction. On Figure B2 the Ca predictions are compared for the best (M01) and two worst predictions (M24 and M25). The best fit in this case does not differ from the actual surface profile by more than 2 %, while the range of predictions when varying the model parameters can extend to ±4 % at some difficult to predict places on the surface.

In the same manner Figure B3 shows the prediction ranges for each tile of the surface of the mineral ratios (mineral surface profiles), and on Figure B4 are broken the predictions for the best and the two worst inferences of the mineral Enstatite. The interval for the mineral composition is wider than the one for the elemental composition - it can reach 20-25 % when the measurement is at apoherm. The best fit does not exceed ±6 % difference from the actual fraction of any given mineral.

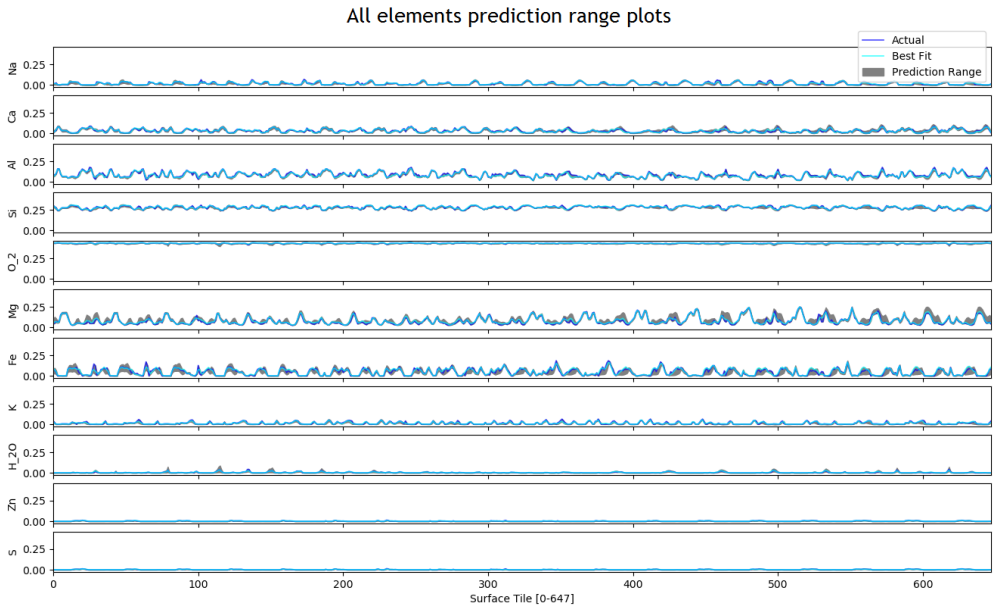

Figure B1. Prediction range plots for all the 11 elements on an example test subset.

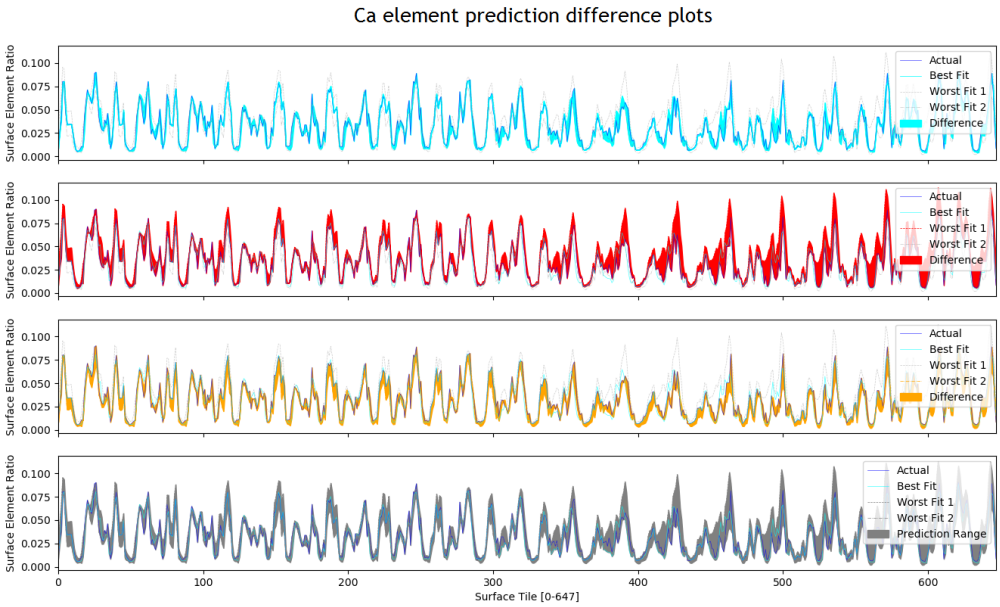

Figure B2. Difference between prediction and "ground truth" for the Ca surface elemental relative composition: for the best inference (M01, panel (a)), and the two worst inferences (M24, panel (b) and M25, panel (c)). Panel (d): Prediction range plot for Ca.

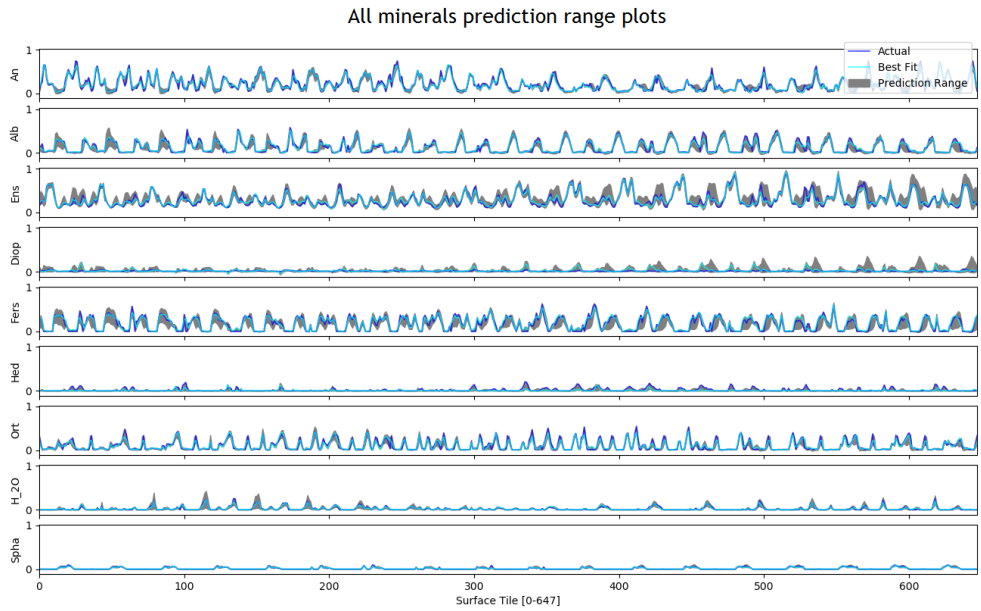

Figure B3. Prediction range plots for all the 9 minerals on an example test subset.

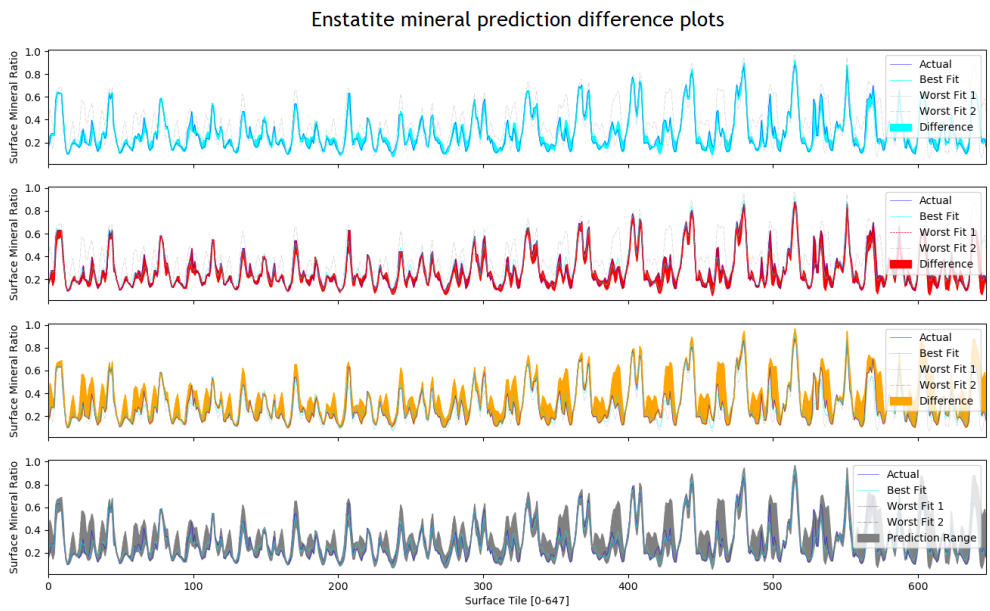

Figure B4. Difference between prediction and "ground truth" for the Enstatite surface mineral relative composition: for the best inference (M01, panel (a)), and the two worst inferences (M24, panel (b) and M25, panel (c)). Panel (d): Prediction range plot for Enstatite.

References

[1] Milillo, A., Wurz, P., Orsini, S., Delcourt, D., Kallio, E., Killen, R., Lammer, H., Massetti, S., Mura, A., Barabash, S., Cremonese, G., Daglis, I., De Angelis, E., Di Lellis, A., Livi, S., Mangano, V., Torkar, K. (2005). Surface-Exosphere-Magnetosphere System Of Mercury. *Space Science Reviews*, **117**, 397-443, doi: 10.1007/s11214-005-3593-z

[2] Domingue, D. L., Koehn, P. L., Killen, R. M., et al. (2007). Mercury's Atmosphere: A Surface-Bounded Exosphere. *Space Science Reviews*, **131**, (1–4): 161–186

[3] Killen, R. M., Cremonese, G., Lammer, H., et al. (2007). Processes that Promote and Deplete the Exosphere of Mercury. *Space Science Reviews*, **132**, (2–4): 433–509

[4] Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M., Ferri, P., Middleton, H. R., & Ziethe, R. (2010). BepiColombo - Comprehensive exploration of Mercury: Mission overview and science goals. *Planetary and Space Science*, **58**, 1: 2-20

[5] Milillo, A., Fujimoto, M., Kallio, E., Kameda, S., Leblanc, F., Narita, Y., Cremonese, G., Laakso, H., Laurenza, M., Massetti, S., McKenna-Lawlor, S., Mura, A., Nakamura, R., Omura, Y., Rothery, D., Seki, K., Storini, M., Wurz, P., Baumjohann, W., Sprague, A. (2010). The BepiColombo mission: An outstanding tool for investigating the Hermean environment. *Planetary and Space Science*, **58**, 40-60. doi: 10.1016/j.pss.2008.06.005

[6] Milillo, A., Fujimoto, M., Murakami, G. et al. (2020). Investigating Mercury's Environment with the Two-Spacecraft BepiColombo Mission. *Space Science Reviews*, **216**, article no. 93. https://doi.org/10.1007/s11214-020-00712-8

[7] Sarantos, M., Slavin, J., Benna, M., Boardsen, S., Killen, R., Schriver, D., Trávnícek, P. (2009). Sodium-ion pickup observed above the magnetopause during MESSENGER'S first Mercury flyby: Constraints on neutral exospheric models. *Geophysical Research Letters*, **36**. doi: 10.1029/2008GL036207.

[8] Cassidy, T. A., Merkel, A. W., Burger, M. H., Sarantos, M., Killen, R. M., McClintock, W. E., and Vervack, R. J. (2015). Mercury's seasonal sodium exosphere: MESSENGER orbital observations. *Icarus*, **248**, 547–559. https://doi.org/10.1016/j.icarus.2014.10.037.

[9] Plainaki, C., Mura, A., Milillo, A., Orsini, S., Livi, S., Mangano, V., Massetti, S., Rispoli, R., and De Angelis, E. (2017). Investigation of the possible effects of comet Encke's meteoroid stream on the Ca exosphere of Mercury. *J. Geophys. Res. Planets*, **122**, 1217-1226. doi: 10.1002/2017JE005304.

[10] Mura, A., Milillo, A., Orsini, S., Massetti, S. (2007). Numerical and analytical model of Mercury's exosphere: dependence on surface and external conditions. *Planetary and Space Science*, **55**, 1569–1583

[11] Mangano, V., et al. (2015). THEMIS Na exosphere observations of Mercury and their correlation within-situ magnetic field measurements by MESSENGER. *Planetary and Space Science*, **115**, 102-109.

[12] Wurz, P. and Lammer, H. (2003). Monte-Carlo simulation of Mercury's exosphere. *Icarus*, **164**, 1-13

[13] Rothery, D. A., Massironi, M., Alemanno, G. et al.(2020). Rationale for BepiColombo Studies of Mercury's Surface and Composition. *Space Sci Rev*, **216**, 66. doi: https://doi.org/10.1007/s11214-020-00694-7

[14] Killen, R. M., and M. H. Burger. (2019). Understanding Mercury's Exosphere: Models derived from

MESSENGER observations. In: *Mercury: The View After MESSENGER.* London: ISBN: 978-1107154452.

[15] Gamborino, D., Vorburger, A., and Wurz, P. (2019), Mercury's subsolar sodium exosphere: an ab initio calculation to interpret MASCS/UVVS observations from MESSENGER. *Ann. Geophys.*, **37**, 455–470

[16] Wurz, P., Whitby, J. A., Rohner, U., Martin-Fernandez, J. A., Lammer, H. and Kolb, C. (2010). Self-consistent modelling of Mercury's exosphere by sputtering, micrometeorite impact and photon-stimulated desorption. *Planetary and Space Science*, **58**, 1599-1616

[17] Mura, A., Wurz, P., Lichtenegger, H., Schleicher, H., Lammer, H., Delcourt, D., Milillo, A., Orsini, S., Massetti, S., Khodachenko, M. (2009). The sodium exosphere of Mercury: Comparison between observations during Mercury's transit and model results. *Icarus*, **200**, 1-11. doi: 10.1016/j.icarus.2008.11.014.

[18] Leblanc, F., Delcourt, D., and Johnson, R. E. (2003). Mercury's sodium exosphere: Magnetospheric ion recycling. *J. Geophys. Res.*, **108**, 5136. doi:10.1029/2003JE002151

[19] Russell, S., and Norvig, P., (2009). *Artificial Intelligence: A Modern Approach, 3rd edition.* Prentice Hall Press, USA

[20] LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, **521**, 436–444. https://doi.org/10.1038/nature14539

[21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* Cambridge, MA: MIT Press

[22] Minsky, M. and Papert, S.A. (2017). *Perceptrons: An introduction to computational geometry.* Cambridge, MA: MIT press

[23] Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in cognitive sciences*, **11**(10), 428–434

[24] Ciresan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep big simple neural nets for handwritten digit recognition. *Neural Computation*, **22**, 1–14.

[25] Kazakov, A. N., et al. (2020). Deep neural networks for analysis of Mercury's planetary exosphere. *J. Phys.: Conf. Ser.*, **1548**, 012014. doi: 10.1088/1742-6596/1548/1/012014

[26] Orsini S., Livi, S., Torkar, K., Barabash, S., Milillo, A., Wurz, P., Di Lellis, A.M., Kallio, E. (2010). SERENA: A suite of four instruments (ELENA, STROFIO, PICAM and MIPA) on board BepiColombo-MPO for particle detection in the Hermean environment. *Planetary and Space Science*, **58**, 1: 166-181

[27] Orsini, S., Livi, S., Lichtenegger, H., et al. (2020). SERENA: particle instrument suite for Sun-Mercury interaction insights on-board BepiColombo. *Space Sci Rev*, in press.

[28] Milillo, A. and Wurz, P. (2014). *SERENA Science Performance Report*

[29] Gulli, A., and Pal, S. (2017). *Deep learning with Keras.* Packt Publishing Ltd.

[30] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, **15**, 315-323

[31] Bishop, C. M. (1995). Regularization and complexity control in feed-forward networks. In *Proceedings International Conference on Artificial Neural Networks ICANN'95*, **1**, 141–148

[32] Rumelhart, D., Hinton, G., and Williams, R. (1986a). Learning representations by back-propagating errors. *Nature*, **323**, 533–536.

[33] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning internal representations by error

propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge.

[34] Kingma, D. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations.*

[35] Bergstra, J., Bardenet, R., Bengio, Y., & Balazs, K. (2011) Algorithms for Hyper-Parameter Optimization. *Proceedings of the 24th International Conference on Neural Information Processing Systems*, 2546-2554

[36] Head, T., et al. (2018). scikit-optimize/scikit-optimize: v0.5.2 (Version v0.5.2). Zenodo. http://doi.org/10.5281/zenodo.1207017

[37] Cintala, M. J. (1992). Impact-induced thermal effects in the lunar and Mercurian regoliths. *J. Geophys. Res.*, **97**, 947– 973

[38] Pokorný, P., Sarantos, M., and Janches, D. (2017). Reconciling the Dawn–Dusk Asymmetry in Mercury's Exosphere with the Micrometeoroid Impact Directionality. *The Astrophysical Journal Letters*, **842**, (2): L17

[39] Killen, R. (2016). Pathways for energization of Ca in Mercury's exosphere. *Icarus*, **268**, 32-36. doi: 10.1016/j.icarus.2015.12.035.

[40] Berezhnoy, A. (2018). Chemistry of impact events on Mercury. *Icarus*, **300**, 210-222. doi: 10.1016/j.icarus.2017.08.034.

[41] Schaible, M. J., et al. (2017). Solar wind sputtering rates of small bodies and ion mass spectrometry detection of secondary ions. *Journal of Geophysical Research: Planets*, **122**, 1968–1983. https://doi.org/10.1002/2017JE005359

[42] Killen, R. M., Potter, A. E., Reiff, P., Sarantos, M., Jackson, B. V., Hick, P., Giles, B. (2001). Evidence for space weather at Mercury. *J. Geophys. Res.*, **106**, 20509–20525.

[43] Mangano, V., Milillo, A., Mura, A., Orsini, S., De Angelis, E., Di Lellis, A. M. and Wurz, P. (2007). The contribution of impulsive meteoritic impact vapourization to the Hermean exosphere. *Planetary and Space Science*, **55**, 1541-56

[44] Bengio, Y. (2015). Early inference in energy-based models approximates back-propagation. Technical Report arXiv:1510.02777, Universite de Montreal.