

1 **Deep neural networks for surface composition
2 reconstruction from in-situ exospheric measurements at
3 Mercury**

4 **Adrian Kazakov¹, Anna Milillo¹, Alessandro Mura¹, Stavro Ivanovski², Valeria
5 Mangano¹, Alessandro Aronica¹, Elisabetta De Angelis¹, Pier Paolo Di
6 Bartolomeo¹, Luca Colasanti¹, Miguel Escalona-Morán³, Francesco
7 Lazzarotto⁴, Stefano Massetti¹, Martina Moroni¹, Raffaella Noschese¹,
8 Fabrizio Nuccilli¹, Stefano Orsini¹, Christina Plainaki⁵, Rosanna Rispoli¹,
9 Roberto Sordini¹, Mirko Stumpo¹, Nello Vertolli¹**

10 ¹INAF-IAPS, via Fosso del Cavaliere 100, 00133, Rome, Italy

11 ²INAF-Osservatorio Astronomico di Trieste, via Giambattista Tiepolo 11, 34143, Trieste, Italy

12 ³Augmented Intelligence Lab, Rua Lugo 2, 36470, Salceda de Caselas, Spain

13 ⁴INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122, Padova, Italy

14 ⁵ASI - Italian Space Agency, Via del Politecnico, 00133, Rome, Italy

15 **Key Points:**

- 16 • This study introduces a machine learning model that predicts Mercury's surface
17 composition using simulated exospheric data.
- 18 • The trained deep neural network is an estimator, a data-driven representation, of
19 the surface-exosphere interactions.
- 20 • This method has potential to enhance future studies of planetary exospheres, in-
21 cluding Mercury's, using data from space missions.

22 **Abstract**

23 Surface information derived from exospheric measurements at planetary bodies comple-
 24 ments surface mapping provided by dedicated imagers, offering critical insights into sur-
 25 face release processes, dynamics of various interactions within the planetary environment,
 26 erosion, space weathering, and planetary evolution. This study explores a method for
 27 deriving the elemental composition of Mercury's regolith from in-situ measurements of
 28 its neutral exosphere using deep neural networks (DNNs). We present a supervised feed-
 29 forward DNN architecture—a network of fully-connected neural layers, the so-called mul-
 30 tilayer perceptron (MLP). This network takes exospheric densities and proton precip-
 31 itation fluxes, derived from a simulated orbital run through Mercury's exosphere, as in-
 32 puts and predicts the chemical elements of the surface regolith below. It serves as an es-
 33 timator for the surface-exosphere interaction and the processes leading to exosphere for-
 34 mation, including micrometeoroid impact vaporization, ion sputtering, photon-stimulated
 35 desorption, and thermal desorption. Extensive training and testing campaigns demon-
 36 strate the MLP DNN's ability to accurately predict and reconstruct surface composition
 37 maps from simulated exospheric measurements. These results not only affirm the algo-
 38 rithm's robustness but also illuminate its extensive capabilities in handling complex data
 39 sets for the creation of estimators for modeled exospheric generation. Furthermore, the
 40 tests reveal substantial potential for further development, suggesting that this method
 41 could significantly enhance the analysis of complex surface-exosphere interactions and
 42 reduce uncertainties in planetary exospheres models. This work anticipates the analy-
 43 sis of data from the SERENA (Search for Exospheric Refilling and Emitted Natural Abun-
 44 dances) instrument package aboard the BepiColombo Mercury Planetary Orbiter, with
 45 its nominal phase starting in 2026.

46 **Plain Language Summary**

47 Mercury is subjected to a variety of environmental effects that influence the com-
 48 plex interactions between its outer planetary layers. This study focuses on Mercury's in-
 49 teraction with its surrounding space environment, particularly the connection between
 50 its surface and its thin exosphere (a layer of gases around the planet). We developed a
 51 computer model that uses data from the exosphere to predict the chemical elements present
 52 on Mercury's surface. This model, a deep neural network, is trained on simulated data
 53 that includes gas densities in the exosphere and particles from the solar wind. By learn-
 54 ing from this information, the model can estimate the processes that form Mercury's ex-
 55 sphere, such as vaporization from micrometeoroid impacts and the release of gas from
 56 particle bombardment. Our tests show that this method is able to predict surface com-
 57 positions, which could help scientists better understand the interactions between the planet's
 58 surface, its exosphere, and its environment. This research is especially relevant for fu-
 59 ture space missions, like the European-Japanese BepiColombo mission, which will be-
 60 gin collecting real data from Mercury in 2026. Our method shows further potential to
 61 enhance how scientists interpret that data and provide new insights into Mercury's dy-
 62 namics and evolution.

63 **1 Introduction**

64 Celestial bodies within our Solar System are continuously influenced by external
 65 forces such as solar wind, solar radiation, and micrometeoroids. These agents contribute
 66 to their reshaping by adding, removing, altering, or relocating material, affecting both
 67 their surfaces and atmospheres. Mercury's atmosphere, being exceptionally tenuous, is
 68 known as an exosphere - a planetary envelope where constituent particle collisions are
 69 so infrequent that their trajectories are essentially ballistic (Milillo et al., 2005; Domingue
 70 et al., 2007). This exosphere arises from a variety of environmental interactions with Mer-
 71 cury's surface. The external factors acting on the planet, such as dust particles, solar

72 wind protons, and heavy ions, as well as solar radiation and intense heat, have profound
 73 effects on the exosphere (R. M. Killen et al., 2007).

74 The active processes in the formation of Mercury's tenuous atmosphere are widely
 75 discussed in the literature (Mura et al., 2007; Wurz et al., 2010; R. M. Killen & Burger,
 76 2019; Grava et al., 2021; Milillo et al., 2023). Four predominant processes release atoms
 77 and molecules from the surface into the exosphere: micrometeoroid impact vaporization
 78 (MIV), sputtering after solar wind and heavy ion impacts (SP), thermal desorption (TD),
 79 and photon-stimulated desorption (PSD). MIV and SP are particularly indicative of the
 80 regolith composition below, as they involve higher energy transfers capable of dislodging
 81 neutral species from their minerals. Conversely, TD and PSD, being less energetic,
 82 tend to release atoms and molecules that are weakly bonded to minerals, such as volatile
 83 elements, most of which eventually fall back and are reabsorbed by the surface (R. M. Killen
 84 et al., 2007; Mura et al., 2009; Gamborino et al., 2019; Leblanc et al., 2023). Once in the
 85 exosphere, the released particles undergo further transformations due to interactions with
 86 radiation pressure, photons, and charged particles. Such interactions can modify the charge,
 87 chemical state, and movement of these exospheric constituents. However, in a first ap-
 88 proximation in the sparse exosphere, the atomic and molecular abundances resulting from
 89 these actors could be traced back to the planet, connecting the surface properties, like
 90 composition, mineralogy, and physical state to the different processes and the dynam-
 91 ics of matter around the planet (Milillo et al., 2020; Rothery et al., 2020). There has al-
 92 ready been direct evidence that this is the case for the Magnesium exosphere, which is
 93 directly related to the Magnesium-rich surface below, as shown by Merkel et al. (2018).

94 To gain a deeper understanding of Mercury's exosphere, scientists use sophisticated
 95 models to simulate the active processes and their effects on the planetary surface, thereby
 96 attempting to replicate the generation of the exosphere. This extensively applied method
 97 compares the results of simulations to those measured from space (e.g. Sarantos et al.
 98 (2009), Cassidy et al. (2015), Plainaki et al. (2017)) or from Earth (e.g. Wurz and Lam-
 99 mer (2003), Mura et al. (2009), Mangano et al. (2015)). However, the inherent complex-
 100 ity of these interactions, which includes electromagnetic, chemical, mechanical, thermal,
 101 and other effects, and the validity of the chosen parameters in their mathematical rep-
 102 resentation add significant challenges. Some of these effects have not been precisely eval-
 103 uated for each release process, leading to a broad range of simulated results with con-
 104 siderable uncertainty, depending on the assumptions made at the outset. To address these,
 105 a multifaceted approach is required, involving the refinement of models through improved
 106 measurements, continual reassessment of the model structure, and advanced statistical
 107 methods to better understand and quantify uncertainties.

108 In parallel, machine learning algorithms, particularly deep neural networks (DNNs),
 109 offer a novel approach to capture the relationships between the variables. These algo-
 110 rithms can resolve, to some extent, the data generation mechanisms (Russell & Norvig,
 111 2009; LeCun et al., 2015; Goodfellow et al., 2016), providing a tool to explore in depth
 112 the relationships between the components of Mercury's environment. This work will demon-
 113 strate how DNNs, especially multilayer perceptrons (MLPs), can be employed within the
 114 data analysis of Mercury's exosphere to reconstruct the elemental surface map under-
 115 neath. Suitable for nonlinear regression tasks, DNNs scale effectively with increasing train-
 116 ing data and input parameters (Minsky & Papert, 2017; Hinton, 2007; Ciresan et al., 2010),
 117 offering a promising direction for tackling the complexities inherent in modeling Mer-
 118 cury's exosphere.

119 This study builds upon the preliminary work of Kazakov et al. (2020) by extend-
 120 ing, refining, and further advancing the application of deep neural networks in predict-
 121 ing Mercury's surface composition from exospheric measurements. It involves predict-
 122 ing the elemental composition of the surface using data from more sophisticated and
 123 realistic models that simulate all major processes—MIV, SP, TD, and PSD—and consider
 124 the influence of solar radiation pressure and photolysis on the exosphere. The multilayer

125 perceptrons have been extensively optimized through a comprehensive exploration of their
 126 building blocks, resulting in the development of a robust predictive algorithm. The method-
 127 ological advancements showcased in this research pave the way for improved predictive
 128 accuracy and reliability in the DNNs employed in this application.

129 In **Section 2**, we introduce and detail the algorithm - the multilayer perceptron
 130 deep neural network - outlining its structure for the multivariate regression task of pre-
 131 dicting surface composition. This section methodically breaks down each component of
 132 the algorithm, providing a comprehensive guide for constructing effective neural network
 133 architectures. **Section 3** delves into the mechanisms behind exospheric data generation,
 134 encompassing the models of Mercury's surface, its environment, and the processes gen-
 135 erating the exosphere. It also elaborates on the creation of the datasets used in the al-
 136 gorithms, including feature selection and data augmentation for the DNN inputs. The
 137 findings from an extensive training and testing campaign are explored in **Section 4**. This
 138 section details the selection of the architectures' hyperparameters, offering insights into
 139 the optimal choices within the DNN hyperparameter space to develop an effective MLP
 140 DNN. The testing of the algorithms is presented, showcasing their performance on a va-
 141 riety of surface-exosphere pairs and culminating in the visual demonstration of recon-
 142 structed surface elemental composition maps. The paper concludes in **Section 5**, pre-
 143 senting a discussion and future perspectives. This final section outlines the method's po-
 144 tential and the wide scope for further research and development in this field.

145 2 Method

146 2.1 Prediction Task and General Characteristics of the Method

147 In this study, we develop a machine learning algorithm and apply it to supervised
 148 multivariate regression of exospheric data at Mercury using a multilayer perceptron deep
 149 neural network. The objective of this DNN is to infer the regolith source material, be-
 150 lieved to be a primary contributor to Mercury's exosphere. Specifically, the DNN pre-
 151 dicted elemental surface composition fractions from exospheric density measurements, gov-
 152 erned by the equation:

$$\sum_{i=1}^n \hat{y}_i = 1, \quad (1)$$

153 where \hat{y}_i is the fraction of an elemental species predicted by the neural network to
 154 be present in the surface area below the exospheric measurement, and n is the total num-
 155 ber of elements in the prediction task (Figure 1).

156 The methodological and algorithmic developments in this study include:

- 157 • Devising the algorithm's application for predicting surface elemental composition.
- 158 • Building DNNs capable of operating in a multi-process environment, integrating
 the four primary active processes (MIV, SP, TD, PSD) for neutral species release.
- 159 • Implementing a data production model with plausible assumptions for the exo-
 sphere generation processes.
- 160 • Training the algorithms with augmented datasets.
- 161 • Employing hyperparameter tuning to optimize the DNN design parameters.
- 162 • Investigating the formation of the physical processes parameter space.

163 Ultimately, the goal of the MLP DNN is to encapsulate the complex relationships
 164 between various surface processes and their impact on the generation of the exosphere,
 165 thereby formulating an estimator for these interactions.

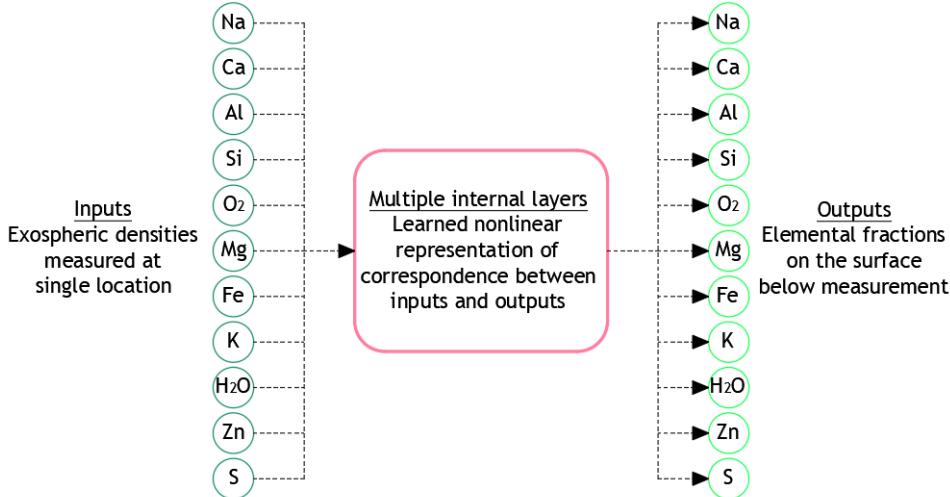


Figure 1. DNN prediction task schematics. The input to the neural network on the left are the exospheric densities at a single location in the exosphere. The output is the relative surface elemental composition as fractions summing up to 1 at a surface area just below the exospheric measurement. The hidden layer box consists of multiple layers and represents the complex, often nonlinear, relationships between the inputs and the outputs of the neural net.

168 2.2 Deep Neural Network Architecture

169 The MLP, a class of feedforward neural network, excels in multivariate regression
 170 by modeling complex nonlinear functions with its multi-layered, fully connected struc-
 171 ture and nonlinear activation functions (Minsky & Papert, 2017; D. Rumelhart et al.,
 172 1986; Kingma & Ba, 2014; LeCun et al., 2015). This architecture, combined with opti-
 173 mization techniques like backpropagation, allows MLPs to identify intricate patterns in
 174 high-dimensional data, making them ideal for robust predictive modeling.

175 Training involves preprocessing data for network suitability, building the model us-
 176 ing the Keras framework with TensorFlow (Abadi et al., 2015; Chollet et al., 2015), and
 177 iteratively tuning the network through backpropagation to minimize error (D. E. Rumel-
 178 hart et al., 1986). This process ensures effective and reproducible model performance.

179 The architecture of a multilayer perceptron enables complex data processing through
 180 a structured network of layers: an input layer, multiple hidden layers for nonlinear trans-
 181 formations, and an output layer for predictions. The network's effectiveness hinges on
 182 key components like the loss function, which guides accuracy improvements, and the reg-
 183 ularizer, which ensures generalizability. Efficiently chosen optimization algorithms and
 184 precise hyperparameter tuning further enhance the network's performance. Figure 2 il-
 185 lustrates this interplay, crucial for tasks like analyzing Mercury's exosphere, with sub-
 186 sequent sections detailing each component's role in predictive capabilities. The inner con-
 187 nectiveness of the MLP DNN neural units is shown on Figure 3.

188 *Input Layer*

189 The input layer of the MLP introduces data, in our case Mercury's exospheric den-
 190 sity measurements, into the network, with each neuron representing a distinct data fea-
 191 ture. For example, distinct elemental density measurements are represented by separate
 192 neurons. Before entry, data undergo normalization to ensure uniform influence on the

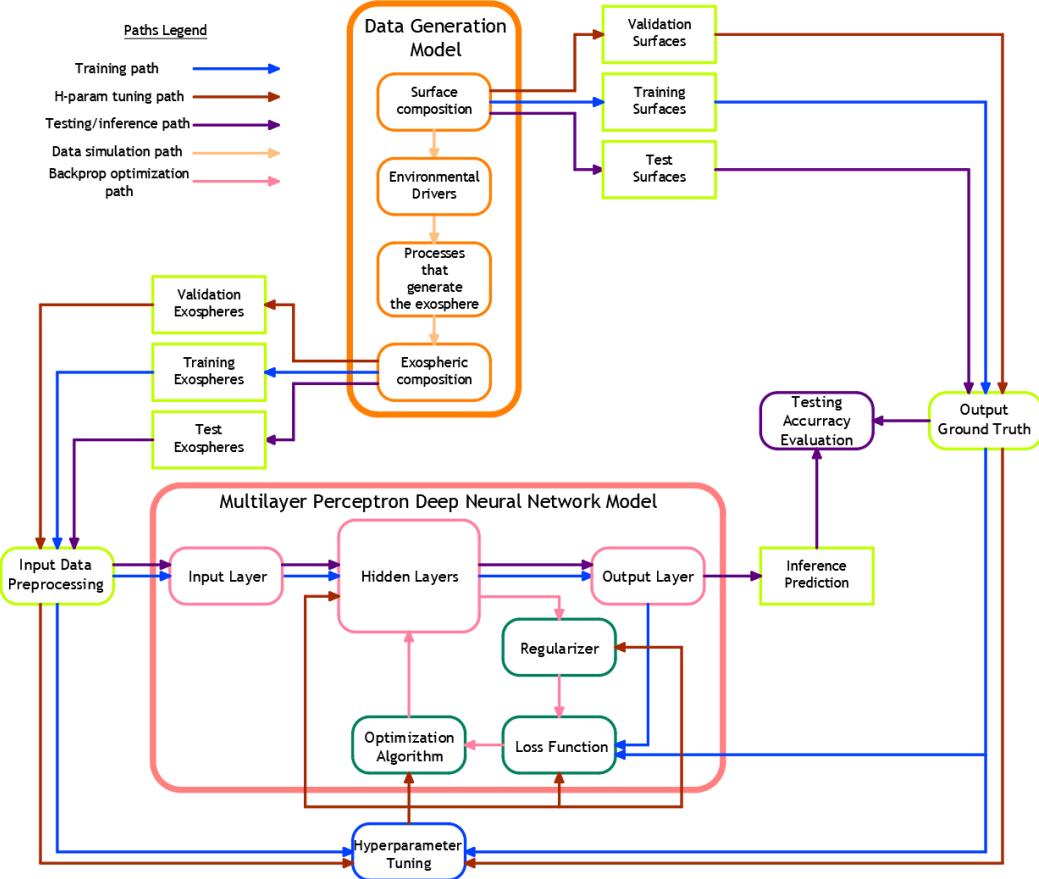


Figure 2. MLP DNN architecture overview. The data generation model produces both the inputs and the outputs for training, validating, and testing the algorithm. This data is passed through the MLP DNN model in the training, hyperparameter tuning and testing phases, respectively. The backpropagation optimization uses the loss function, regularizer and optimization algorithm to adjust the weights (internal parameters) of the neural network. In a separate process, the hyperparameter tuner adjusts/optimizes the MLP DNN by minimizing the errors on the validation dataset. After the final training, the previously unseen data from the testing sets is passed through the network and the accuracy of the predictions (performance of the network) is evaluated.

learning process, thereby preventing bias (Goodfellow et al., 2016). This involves standardizing each feature to zero mean and unit variance as per the equation:

$$\mathbf{x} = \frac{\mathbf{x}_{\text{orig}} - \boldsymbol{\mu}}{\sigma} \quad (2)$$

where \mathbf{x} is the standardized vector of input features, \mathbf{x}_{orig} is the original vector of input features, $\boldsymbol{\mu}$ is the vector of means of the feature values, and σ is the standard deviations vector. Such standardization enhances the efficiency and stability of the network's learning process.

199 *Hidden Layers*

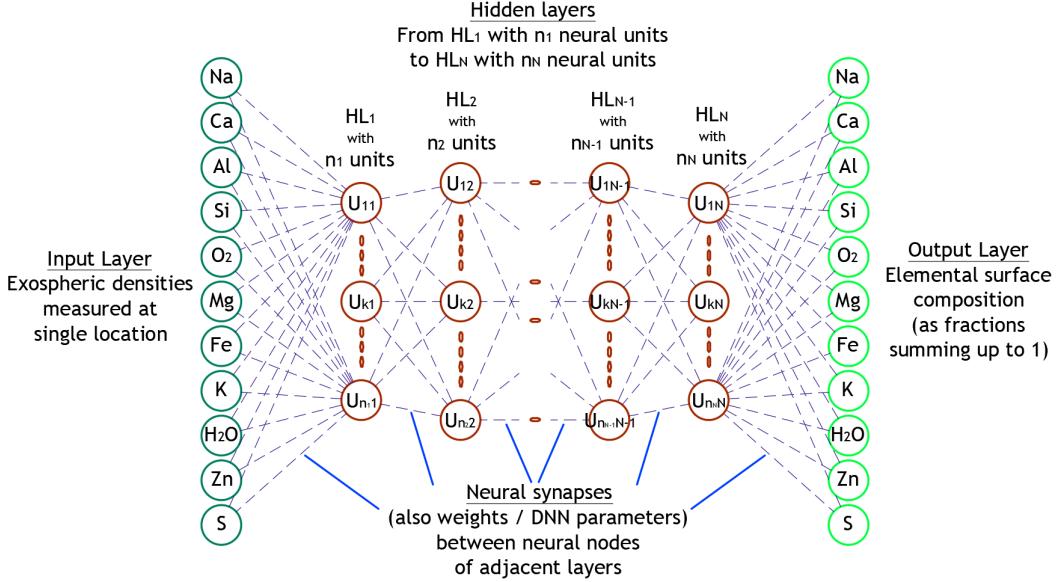


Figure 3. MLP DNN basic structure schematics. Exospheric densities form the input layer on the left. The output layer is formed from the relative surface elemental composition as fractions summing up to 1. There are N number of hidden layers with n_1 to n_N number of neural units (neurons). This structure represents the relationships between the inputs and the outputs of the MLP. The neural synapses, connections between the neural units, form the weight (DNN parameter) matrices W_1 to W_{N+1} .

The hidden layers form the core of the MLP architecture, where the actual processing and learning occur (Minsky & Papert, 2017; Hinton, 2007). Positioned between the input and output layers, they transform input data into a form usable for predictions. Each hidden layer is composed of a set of neural units - neurons - and each neuron in these layers is fully connected to all neurons in the preceding and succeeding layers, creating a dense network of synaptic connections. The structure of these layers is represented mathematically by combination matrices or weight matrices, which, along with the activation function applied at each neuron, helps form an estimation of the relationships among the processes acting between the input layer and the output layer.

In our study on Mercury's exosphere, multiple hidden layers with a substantial number of neurons allow the MLP to capture the nuances of Mercury's exospheric composition and the underlying processes that govern it. A key component of these hidden layers is the activation function, in this case the Rectified Linear Unit (ReLU), essential for introducing nonlinearity and aiding in effective gradient propagation to avoid vanishing gradients (Glorot et al., 2011). The ReLU function is defined as $a(\mathbf{z}) = \max(0, \mathbf{z})$, where \mathbf{z} is the input to the activation function.

The transformation within each hidden layer then follows the equation:

$$\mathbf{h} = a(\mathbf{z}) = a(\mathbf{W}^T \mathbf{x} + \mathbf{b}) = \max(0, \mathbf{W}^T \mathbf{x} + \mathbf{b}), \quad (3)$$

where \mathbf{W}^T represents the weight matrix, \mathbf{x} is the input vector to the hidden layer (input features or activations from a previous hidden layer), and \mathbf{b} is the bias vector of the affine transformation. This equation encapsulates the affine transformation followed

220 by the application of the ReLU activation function, enabling the network to learn and
221 represent complex nonlinear relationships.

222 Finally, the output from the hidden layers is passed on to the output layer, where
223 the final prediction is made. The architecture and depth of the hidden layers are crit-
224 ical and typically determined through empirical methods and hyperparameter tuning.
225 This ensures the network has the requisite complexity for effective learning while avoid-
226 ing overfitting to the empirical distribution present in the training data.

227 *Output Layer*

228 The output layer is the final layer in an MLP, playing the role of determining the
229 format and nature of its predictions. In the context of our study, this layer is tailored
230 to predict the elemental composition of Mercury's surface, with each neuron correspond-
231 ing to one of the elements being analyzed. For example, if predicting the fractions of 11
232 different elements, the output layer would consist of 11 neurons.

233 The activation function used in the output layer is crucial and depends on the na-
234 ture of the prediction task. In our case, where the output is a set of continuous values
235 that sum to 1 (representing fractions), the softmax function is used (Joachims, 2002).
236 The softmax function converts the raw output of the network into a probability distri-
237 bution, ensuring that the predicted fractions are non-negative and sum up to one, align-
238 ing perfectly with the physical reality of our task.

239 The formula for the softmax function is as follows:

$$\hat{y}_i = \text{softmax}(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}, \quad (4)$$

240 where \hat{y}_i is the predicted fraction for the i -th element, and \mathbf{z} represents the raw out-
241 put values from the final hidden layer. This configuration allows the network to deliver
242 accurate, meaningful predictions of Mercury's surface composition, synthesizing the rep-
243 resentation insights gained from all previous layers.

244 *Loss Function*

245 The loss function plays a pivotal role in guiding the optimization process, quan-
246 tifying the discrepancy between the network's predictions and the actual target values
247 to gauge model accuracy. In our most successful MLP tests on predicting Mercury's sur-
248 face elemental composition, the Kullback-Leibler (KL) divergence (Cover & Thomas, 2006)
249 has proven particularly effective. It measures how one probability distribution, repre-
250 senting the predicted elemental composition (the output from the MLP), diverges from
251 the actual distribution (the true elemental composition). The formula for KL divergence
252 is:

$$KL(P||Q) = J(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y}) = \sum_i P(i) \log \frac{P(i)}{Q(i)}, \quad (5)$$

253 where P represents the true distribution of the fraction of element i in the data,
254 and Q is the predicted distribution from the MLP. Other loss functions like Mean Ab-
255 solute Error (MAE) and Mean Squared Error (MSE) were also considered. However, the
256 KL divergence was preferred for our regression task because it aligns better with the prob-
257 abilistic requirements, focusing on relative proportions rather than absolute quantities
258 of elements.

259 *Regularizer*

260 Regularization is an essential technique in neural network training, designed to enhance
 261 model generalization by adding constraints or penalties to the loss function. In
 262 our study, we use L2 regularization (weight decay) on the weights of each hidden layer(Bishop,
 263 2006). This technique constrains the magnitude of the weights, preventing them from
 264 becoming excessively large and helping to avoid overfitting the model to the specific dataset
 265 used for training.

266 The L2 regularization is mathematically represented as:

$$\hat{J}(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y}) = J(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y}) + \lambda \sum_{i=1}^m |\theta_i|^2, \quad (6)$$

267 where $J(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y})$ is the original loss function, λ is the regularization coefficient,
 268 and θ denotes the vector of all weight parameters (unfolded from the \mathbf{W} matrices). The
 269 right choice of λ is critical. If λ is too large, it can lead to underfitting, where the model
 270 is overly simplified and fails to capture the underlying trends in the data. Conversely,
 271 a very small λ might not effectively prevent overfitting.

272 In a multivariate regression task such as ours, where the model needs to understand
 273 complex relationships between various features in the surface-exosphere interaction at
 274 Mercury, L2 regularization helps in maintaining a balance between MLP model complex-
 275 ity and its ability to generalize. The addition of this regularization term (penalty) to the
 276 loss function thus ensures that the model not only fits the training data well but also
 277 maintains the flexibility to perform accurately on new, unseen data.

278 *Optimization (Learning) Algorithm*

279 The training of our multilayer perceptron for predicting Mercury's surface compo-
 280 sition employs the Adam optimization algorithm, a refinement of stochastic gradient de-
 281 scent known for its effectiveness with large-scale data and complex models (Kingma &
 282 Ba, 2014). The fundamental mechanism of Adam involves dynamically and adaptively
 283 updating the weights of the combination matrices for each hidden layer to minimize the
 284 total error as indicated by the loss function. This is achieved through backpropagation
 285 optimization (D. Rumelhart et al., 1986; D. E. Rumelhart et al., 1986), where the weights
 286 are adjusted following their gradients with respect to the loss function:

$$\boldsymbol{\theta} := \boldsymbol{\theta} - \alpha \frac{1}{m} \nabla_{\boldsymbol{\theta}} \text{KL}(P|Q) = \boldsymbol{\theta} - \alpha \frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_{j=1}^m \sum_i P_j(i) \log \frac{P_j(i)}{Q_j(i)}, \quad (7)$$

287 In this equation, α represents the learning rate and $\nabla_{\boldsymbol{\theta}} \text{KL}(P|Q)$ is the gradient of
 288 the KL divergence with respect to the model parameters $\boldsymbol{\theta}$. The stochastic nature of the
 289 gradient descent implies that learning iterations are not performed on the entire dataset
 290 but rather on a random subset known as a mini-batch. Here, m denotes the number of
 291 examples in the mini-batch.

292 *Hyperparameter Tuning*

293 Key hyperparameters of our MLP DNN are:

- 294 • **Learning Rate:** This parameter governs the size of the steps taken during the
 295 backpropagation optimization algorithm along the weight gradients of the loss func-
 296 tion. A well-balanced learning rate is critical—it must be large enough to navi-
 297 gate plateaus in the loss function's parameter space, yet sufficiently small to con-
 298 verge to (or remain near) the minimum of the error.

- **Mini-Batch Size:** This refers to the size of the random subset of examples used in each training iteration, impacting both the speed and stability of the learning process.
- **Number of Hidden Layers and Neurons:** These parameters determine the depth and width of the neural network, influencing its ability to model complex relationships in the data.
- **L2 Regularization Coefficient:** This defines the degree of penalty imposed on large weight values, helping to prevent overfitting by controlling model complexity.

To fine-tune these hyperparameters, we employed a Bayesian optimization strategy using the Gaussian Process (GP) approach, as outlined in Bergstra et al. (2011). The tuning process was facilitated by the scikit-optimize library (Head et al., 2018), which utilizes a prior probability distribution function to identify the hyperparameter configuration that minimizes the total loss on a hold-out validation dataset. This systematic adjustment of hyperparameters not only enhances learning capabilities and overall performance, but also optimizes the balance between model complexity and efficiency.

315 Performance Metrics

To evaluate the performance of our machine learning model, we utilize both customized and standard metrics to ensure precise and insightful quantitative assessments. Our primary metric, the Euclidean similarity 4 (ES4), integrates elements of Euclidean distance and cosine similarity, providing a nuanced measure of prediction accuracy by considering both magnitude and directionality in multidimensional space:

$$\text{ES4} = \left(1 - \frac{\sqrt{\sum_i (\hat{\mathbf{y}}_i - \mathbf{y}_i)^2}}{\sqrt{\sum_i \mathbf{y}_i^2}} \right) \times \left(\frac{\hat{\mathbf{y}}_i \cdot \mathbf{y}_i}{\|\hat{\mathbf{y}}_i\| \|\mathbf{y}_i\|} \right), \quad (8)$$

where $\hat{\mathbf{y}}_i$ and \mathbf{y}_i represent the predicted and actual surface compositions, respectively.

Moreover, we incorporate the R-squared (R^2) metric into our evaluation framework. The R^2 metric, commonly used in regression analysis, quantifies the proportion of the variance in the dependent variable that is predictable from the independent variable(s). In the context of multivariate regression, R^2 is defined as:

$$R^2 = 1 - \frac{\sum_i (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_i (\mathbf{y}_i - \bar{\mathbf{y}})^2}, \quad (9)$$

where $\bar{\mathbf{y}}$ is the mean of the actual values. This metric is particularly useful for assessing the model's ability to capture the variance in the data, offering insights into how well the model's predictions approximate the actual data distribution compared to a naive model that only predicts the mean.

Additionally, we evaluate the model using absolute and relative residuals, which provide further granularity in understanding the model's performance. These residuals help identify the absolute and relative differences between predicted and actual values, offering a direct measure of prediction error.

By combining these metrics, we achieve a multidimensional evaluation of our DNN's performance, encompassing both the accuracy of individual predictions and the model's overall ability to capture the complexity of the data. This comprehensive assessment not only ensures validation of the model's outputs but also sheds light on areas for potential improvement, thereby contributing to the refinement of the model's predictive capabilities.

3 Data Generation Model

For the development of our MLP DNN, we employ a data generation model consisting of three components: a surface composition model, an exosphere generation model, and a model for the distributions of the drivers of surface release processes. The surface model in our study has been considered as a set of nearly unconstrained composition maps for each species. These maps are what the DNN aims to reconstruct. The exosphere model captures all the main physical processes that link the surface, external environment conditions, and the exosphere. The drivers model comprises maps of the surface where external drivers, such as ions, photons, and micrometeoroids, are active. These maps are derived from simplified functions in this prototype network. In the following sections, we describe the models used for these three components in greater detail.

Our focus extends beyond the inherent complexities of Mercury’s exosphere to include a thorough description of the physical parameter space used to represent the processes behind exospheric formation. The parameter space is an important element in our simulation approach, allowing us to explore and represent a variety of planetary environmental conditions. Our analysis encompasses a wide range of physical processes relevant to the surface-exosphere interaction, considering a highly multidimensional parameter space that includes both variable and fixed processes and drivers. It can be further subdivided into two subspaces: the parameter subspace constructed from physical interactions explicitly or implicitly considered in our model, and a second parameter space describing aspects omitted due to lack of computational resources or understanding.

The three components of the data generation model comprehensively describe the physical processes shaping the surface-exosphere interaction, allowing us to define a specific region in the parameter space. This region governs the data distribution that our multilayer perceptron deep neural network aims to estimate. The objective of this work is to demonstrate that our trained MLP DNN can closely approach the most representative region in the physical parameter space and serve as a robust estimator of the data generation mechanism. Understanding this concept is crucial for grasping the nature of the surface-environment-exosphere relationships that our deep neural network is trained to represent, laying the foundation for further exploration of this parameter space.

3.1 Surface and Regolith

Even if the Mercury surface composition has been partially identified by MESSENGER data (E. Vander Kaaden et al., 2017), for the purposes of this study—specifically, to validate our method—we decided to consider surface compositions that are in a broader range around the composition ratios expected for the main mineralogical components of Mercury. We selected a specific set of minerals (Table 1) believed to be present on Mercury’s surface (Wurz et al., 2010). These minerals are assumed to exist in varying proportions, contributing to the overall mineral composition of the regolith. These proportions delineate zones characterized by dominant primary minerals and their secondary counterparts, enforcing the presence of some of the minerals on the surface. Additionally, constraints on the minimal fractions of specific minerals and the presence of water ice further refine this parameter subspace.

Among the included minerals, hedenbergite (primarily deposited from meteorites), sphalerite (resulting from volcanic activity), and water ice are considered as rare minerals. Their presence is strongly reduced in the random surface generation, compared to the other six minerals. Furthermore, in the randomized creation of the surface, they are not allowed to be distributed everywhere on the surface. The overall minerals used in our datasets, including the split to their constituent elemental species (atoms or molecules), are reported in Table 1. This implies an assumption that the surface, on average, encompasses a complete pool of atoms and molecules derived from these minerals, which are then subjected to external environmental forces. This approach, as an approximation,

Mineralogy and Relationships (Baseline Model)					
Mineral Name	Chemical Formula	Decomposed to	Rarity	Mean Fraction	Range
Anorthite	$\text{CaAl}_2\text{Si}_2\text{O}_8$	$\text{Ca}, 2\text{Al}, 2\text{Si}, 4\text{O}_2$	-	0.134	0.049 - 0.319
Albite	$\text{NaAlSi}_3\text{O}_8$	$\text{Na}, \text{Al}, 3\text{Si}, 4\text{O}_2$	-	0.140	0.051 - 0.341
Orthoclase	KAlSi_3O_8	$\text{K}, \text{Al}, 3\text{Si}, 4\text{O}_2$	-	0.134	0.050 - 0.313
Enstatite	$\text{Mg}_2\text{Si}_2\text{O}_6$	$2\text{Mg}, 2\text{Si}, 3\text{O}_2$	-	0.137	0.053 - 0.312
Diopside	$\text{MgCaSi}_2\text{O}_6$	$\text{Mg}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	-	0.141	0.053 - 0.336
Ferrosilite	$\text{Fe}_2\text{Si}_2\text{O}_6$	$2\text{Fe}, 2\text{Si}, 3\text{O}_2$	-	0.137	0.052 - 0.329
Hedenbergite	$\text{FeCaSi}_2\text{O}_6$	$\text{Fe}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	Rare	0.065	0.010 - 0.213
Sphalerite	ZnS	Zn, S	Rare	0.069	0.012 - 0.243
Water Ice	H_2O	H_2O	Rare	0.044	0.012 - 0.136

Table 1. Mineral composition considered in the baseline surface model. The minerals are decomposed via the classical additive method to "elemental" species. The decomposition captures some of the relationships between mineralogy and elemental composition, while others are omitted (e.g. decomposition of water ice, or decomposition to heavier molecules). The mean mineral fraction reported in this table is for all the datasets generated in this study - 204,768 surface tiles.

392 considers the full fraction of volatile species (such as Na, K, H_2O , S, and O_2) as being
 393 readily available for release into the exosphere as they are loosely bound to the regolith
 394 grains. The list of elements resulting from the mineral break down is presented in Ta-
 395 ble 2. The elemental composition resulting from this process represents the 'actual' or
 396 'ground truth' data that we compare with the predictions obtained by our algorithms.

Elements and Relationships (Baseline Model)					
Element Name	Designation	From Mineral	Rarity	Mean Fraction	Range
Aluminium	Al	Anorthite, Albite, Orthoclase	-	0.072	0.043 - 0.113
Calcium	Ca	Anorthite, Diopside, Hedenbergite	-	0.047	0.025 - 0.072
Iron	Fe	Ferrosilite, Hedenbergite	-	0.048	0.020 - 0.098
Sodium	Na	Albite	-	0.019	0.007 - 0.042
Oxygen	O_2	All, except Sphalerite, Water Ice	Dominant	0.423	0.391 - 0.435
Sulfur	S	Sphalerite	Rare	0.012	0.002 - 0.047
Water Vapor	H_2O	Water Ice	Rare	0.008	0.002 - 0.026
Zinc	Zn	Sphalerite	Rare	0.012	0.002 - 0.047
Silicium	Si	All, except Sphalerite, Water Ice	Dominant	0.282	0.258 - 0.301
Potassium	K	Orthoclase	-	0.018	0.007 - 0.041
Magnesium	Mg	Enstatite, Diopside	-	0.059	0.030 - 0.109

Table 2. Elemental composition considered in the baseline surface model. The elements are broken down from minerals in the classical additive method. The ranges of variation for each element are in the last column.

397 For our modeling, we need to consider not only the composition map but also some
 398 characteristics of the planetary surface and regolith (the loose, heterogeneous material
 399 covering solid rock), which are part of our simulation model. These include the influence
 400 of surface composition, texture, and physical, chemical, and thermal properties, all in
 401 the context of forming the modeling parameter space and defining a region within that
 402 space.

403 Firstly, the surface in our model is represented as a grid comprised of 36×18 sur-
 404 face tiles in a modified Mercator projection. Each tile measures $10^\circ \times 10^\circ$, which, at the
 405 equator, translates to approximately 425 km \times 425 km. This averaging of composition in-
 406 evitably reduces the complexity of the parameter space, as it results in the loss of finer
 407 details in the spatial relations of the spread of the different species - elemental and min-
 408 eral - on the planetary surface. However, it is important to note that the potential res-
 409 olution of surface composition maps reconstructed from exospheric measurements by an
 410 orbiting spacecraft cannot be significantly higher.

411 Conversely, while our model omits certain surface qualities such as grain sizes, slope
 412 angles, and roughness, we do incorporate a simplified representation of porosity and the
 413 presence of microshadows in the ion-sputtering process acting on the surface.

414 3.2 Environmental Drivers

415 In our model definition, we incorporate the environmental conditions and various
 416 factors that contribute to changes in the sources or processes for the release of material
 417 from the planetary surface into the exosphere. These drivers encompass solar radiation,
 418 dust particles, and charged particles that enable surface material to escape into the ex-
 419 osphere. Mercury's proximity to the Sun significantly influences its interaction with the
 420 surrounding environment. Its highly eccentric orbit, varying between 0.31 and 0.46 as-
 421 tronomical units (AU), causes external conditions such as thermal radiation, photon flux,
 422 and solar wind intensity to fluctuate according to its distance from the Sun. In our model,
 423 we specifically focus on conditions at perihelion. This close distance markedly impacts
 424 the intensity of the solar influence, thereby affecting the range of effects and processes
 425 contributing to the generation of the exosphere.

426 One such influence is the equivalent photon flux, which is the photon flux at Earth's
 427 orbit adjusted for Mercury's closer position to the Sun by a factor of $1/r^2$, where r is
 428 the distance to the Sun in AU. A photon flux at Earth of $3.0 \times 10^{15} \text{ cm}^{-2}\text{s}^{-1}$ is con-
 429 sidered (Mura et al., 2009) resulting in an equivalent photon flux of $3.1 \times 10^{16} \text{ cm}^{-2}\text{s}^{-1}$.

430 Another environmental aspect is the activity level of the Sun, which we have as-
 431 sumed to be at a moderate level, devoid of extreme events such as coronal mass ejections
 432 or solar flares. This assumption sets the conditions for a solar wind velocity of 450 km/s
 433 and a solar wind density of 60 cm^{-3} at Mercury's perihelion (Wilson III et al., 2018).

434 Furthermore, the dust environment around Mercury is considered for particles smaller
 435 than $100 \mu\text{m}$ in diameter, with a mean flux of $1.0 \times 10^{-16} \text{ g/cm}^2\text{s}$ and mean velocity of
 436 20 km/s in Mercury's vicinity, in agreement with the modal impact velocity reported by
 437 Cintala (1992). This is compared to a planet velocity at perihelion of 59 km/s. However,
 438 our model does not differentiate between the origins of these dust particles—whether they
 439 come from the Main Belt Asteroids, Jupiter Family Comets, Oort Cloud Comets, or Hal-
 440 ley Type Comets—nor does it consider the full ranges and exact distributions of parti-
 441 cle sizes and velocities as in Pokorný et al. (2018). Additionally, no large meteorite im-
 442 pacts or increases of fluxes due to particularly dense cometary streams, such as fromc
 443 omet Encke (Plainaki et al., 2017), are considered. Grain size distribution influence of
 444 the dust particles is also not represented in our physical parameter space.

445 The environmental conditions on the planet itself present a diverse range of param-
 446 eters due to varying exposure to sunlight and shadow, as well as differences in particle
 447 fluxes on the planet's leading and trailing sides due to its high orbital velocity. Our model
 448 incorporates detailed maps that illustrate solar incidence angles and planetary velocity
 449 incidence angles at Mercury's perihelion (Figure 4). It's crucial to recognize Mercury's
 450 unique orbit-spin resonance, which alternates the sides facing the Sun at the same true
 451 anomaly angle in successive orbits, a fact that we have taken advantage of later in our
 452 study.

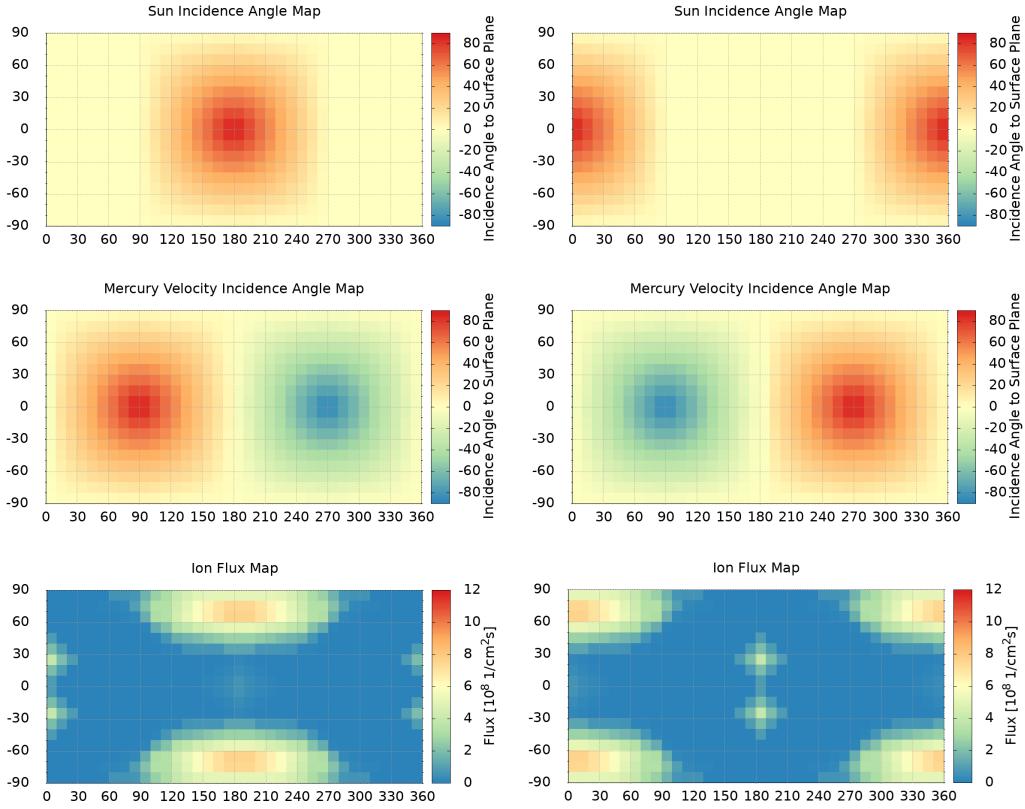


Figure 4. Maps of incidence angles due to planet orientation at two consecutive perihelia. Perihelion 1 on the left and perihelion 2 on the right. The top panels represent the incidence maps of solar photons on the surface. The middle panels represent the angles between the planet's velocity and the surface normal. The bottom panels show the magnetic field cusp footprints on the surface. A shift of 180 degrees in longitude between the two perihelia is notable due to the spin-orbit resonance of Mercury.

Our study does not encompass the full spectrum of influences that the interplanetary magnetic field and Mercury's own magnetic field might have on the planet's surface and exosphere. The sole aspect of Mercury's magnetic field that our model takes into account is the presence of open magnetic field lines. These lines serve as conduits for charged particles, channeling them through magnetic field cusps directly onto the planet's surface at specific locations known as cusp footprints (Massetti et al., 2003). The impact of this process is significant, as it concentrates ion bombardment in particular areas, altering the surface composition and influencing the generation of the exosphere. We have delineated the shape and relative impact areas of these cusp footprints in Figure 4, highlighting the regions on Mercury's surface that are most affected by the ion funneling. These zones are another variable that is dynamically changing and in many cases in very short timeframes. In our study, we have considered a fairly wide zones of ion precipitation. A more dynamic dependence on the magnetic reconnection rate driven by the Interplanetary Magnetic Field (IMF) strength and orientation is neglected.

A summary of the environmental conditions considered in our study and their contribution further to the release processes and to the represented parameter space are given in Table 3.

Environmental Conditions					
Group	Condition	Units	Value	Sources Affected	Processes Affected
Star Activity	Solar wind velocity	km/s	450	Proton flux	SP
	Solar wind density	1/cm ³	60	Proton flux	SP
Comets and Asteroids	Photon flux at Earth	1/cm ² s	3.0 x 10 ⁻¹⁵	Equivalent photon flux	PSD
	Dust particle size	μm	0-100	Micrometeorite flux	MIV
Magnetic fields	Mean flux of dust particles	g/cm ² s	1.0 x 10 ⁻¹⁶	Micrometeorite flux	MIV
	Mean velocity of dust particles	km/s	20	Micrometeorite flux	MIV
Cusp footprints size	Cusp footprints location	Map		Ion precipitation zone	SP
	Cusp footprints ion flux distribution	Map		Ion precipitation zone	SP
Planet	Cusp footprints area coefficient	-	0.4	Ion flux	SP
	Planet velocity	km/s	59	Micrometeorite flux	MIV
Planet	Planet orientation	deg	0 (local solar time offset)	All sources influence zones	All
	Distance from Sun	AU	0.31	All	All

Table 3. Environmental parameters and conditions and their effects on the populations of particles or other processes that release particles from the surface.

470

3.3 Exosphere Generation Model

471

The exospheric model is used for generating our own simulated datasets for training and testing the DNN. Our method is designed to be adaptable, not limited to a single model, but capable of reconstructing various exosphere generation processes and predicting surface compositions. For this purpose, we decided to use the Mura et al. (2007) model, as it is a comprehensive exosphere generation model able to describe all the main surface release processes for different species on Mercury as a function of external environmental drivers, with the flexibility to tune the simulated exosphere with many relevant parameters and driver inputs. It is important to note that the exosphere is non-collisional, meaning each released species does not interact with others. Therefore, we can consider the exosphere as the sum of different exospheres for each species.

481
482
483
484
485
486
487
488
489
490

An outline of the surface release processes used in our DNN is provided below, with more details available in Mura et al. (2007). These active processes include the four main ones, namely the micrometeorite impact vaporization (MIV), sputtering from protons and heavy ions (SP), thermal desorption (TD), and photon-stimulated desorption (PSD). Their respective sources, or drivers, are micrometeoroid fluxes, precipitating ions through the open field lines of Mercury's magnetic field, temperature effects on the surface, and solar photons that impact the dayside surface. However, there are still significant gaps in understanding these processes, making the problem not fully constrained in terms of what is observed versus what the model can reproduce. Despite these uncertainties, we have chosen a realistic range for their values.

491

3.3.1 Temperature map and Thermal desorption (TD)

492
493
494

TD becomes notably efficient at temperatures above 400 K (Mura et al., 2007). We consider this process for the release of Na, K, H₂O, and S which are loosely bound to Mercury's surface.

495
496
497
498

We assume the subsolar point temperature on Mercury reaches 700 K at perihelion, while the night side registers a much lower temperature of 110 K. The temperature distribution across the surface adheres to a quarter-power law, ranging from a minimum of 110 K to a maximum of 700 K at subsolar point:

$$T_s(\phi, \theta) = T_{\min} + (T_{\max} - T_{\min})(\cos\phi\cos\theta)^{1/4}, \quad (10)$$

499

where ϕ represents the latitude and θ the longitude, as outlined in Mura et al. (2007).

TD is considered only as a direct thermal ejection of species from the surface i.e. sticking coefficient equal to 1. The flux of atoms or molecules resulting from TD is given by Mura et al. (2007):

$$\Phi_n = \nu N c e^{-\frac{U_d}{k_B T}}, \quad (11)$$

where ν denotes the vibrational frequency of the species, N the surface density of the regolith, c the fractional presence of the species within the regolith, U_d the species' binding energy, k_B the Boltzmann constant, and T the temperature at which desorption occurs.

Exospheric Model Parameters													
Parameter	Process	Units	Al	Ca	Mg	Na	K	Fe	Si	Zn	S	O ₂	H ₂ O
Dayside temperature	TD	K						700					
Nightside temperature	TD	K						110					
Surface density	TD	1/cm ²						7.5 × 10 ¹⁴					
Vibrational frequency	TD	1/s						1.0 × 10 ¹³					
Binding energy	TD, SP	eV	3.36	2.1	1.54	2	0.93	4.34	4.7	1.35	2.88	2	0.5
Mean photon flux	PSD	1/cm ² s						3.1 × 10 ¹⁶					
beta coefficient	PSD	-	-	-	-	1	1	-	-	-	1	-	1
Temperature above regolith	PSD	K	-	-	-	200	200	-	-	-	200	-	0
PSD cross section	PSD	1/m ²	-	-	-	1 × 10 ⁻²⁵	1 × 10 ⁻²⁵	-	-	-	1 × 10 ⁻²⁵	-	1 × 10 ⁻²²
Mean flux of dust particles	MIV	g/cm ² s						1.0 × 10 ⁻¹⁶					
Mean velocity of dust particles	MIV	km/s						20					
Vapor phase production rate	MIV	-						5					
Vapor temperature	MIV	K						4000					
Mean ion flux	SP	1/cm ² s						1.08 × 10 ⁹					
Yield efficiency	SP	-						0.1					
Impact energy	SP	eV						1000					
Porosity coefficient	SP	-						0.35					
Microshadows coefficient	SP	-						0.4					
Photoionization lifetime	Exo	s	600	2500	25000	6000	4000	8000	5000	20000	8000	20000	50
Radiation acceleration	Exo	cm/s ²	5	5	5	15	25	5	5	5	5	5	5

Table 4. Parameters used in the processes of the exospheric model.

3.3.2 Photon flux and Photon-stimulated desorption (PSD)

PSD is initiated by the interaction of incoming photons with the surface, each photon possessing the capability to eject atoms or molecules from a population of loosely bound volatile species. The efficiency of this process is contingent upon the cross section for photon impact (Wurz & Lammer, 2003; R. M. Killen et al., 2001; Wurz et al., 2010). At perihelion, the incident photon flux is quantified as $3.1 \times 10^{16} \text{ cm}^{-2}\text{s}^{-1}$.

The model quantifies the neutral particle flux resulting from PSD as:

$$\Phi_n = N_c \int \Phi_\gamma(E) \sigma(E) dE, \quad (12)$$

where $\Phi_\gamma(E)$ denotes the energy-dependent differential photon flux, $\sigma(E)$ the relative differential cross-section for desorption, N the surface density of the regolith, and c the fraction of the specific neutral species being considered.

The photon flux as a function of incidence angles is described by the following relation:

$$\Phi_n(\phi, \lambda)^* = \Phi_n \cos(\phi) \cos(\lambda), \quad (13)$$

with ϕ representing the longitude in local solar time and λ the latitude, thereby factoring in the geometric reduction of flux due to the angle of solar incidence.

521 The PSD process' energy distribution is modeled using a formula adapted from Johnson
 522 et al. (2002):

$$f(E) = \beta(1 + \beta) \frac{EU^\beta}{(E + U)^{2+\beta}}, \quad (14)$$

523 in which β is set to 1 for our study to represent an energy cut-off, and U denotes
 524 the threshold energy.

525 3.3.3 Micrometeorite fluxes and impact vaporization

526 MIV is a highly energetic surface-release process capable of releasing the entire sur-
 527 face material in a given volume after a micrometeoroid impact (not only the volatile species).
 528 In the Mura et al. (2007) model, the MIV exosphere is simulated starting from a map
 529 of surface release particles at a given release energy, represented as a Maxwellian distri-
 530 bution at 4000 K (Wurz & Lammer, 2003).

531 The vaporized material includes not only single elements but also molecules such
 532 as CaO, NaOH, NaO, and others, resulting from the complex chemistry within the impact-
 533 produced cloud and a fraction of condensed material that re-impacts the surface (R. Killen,
 534 2016; Berezhnoy, 2018; Moroni et al., 2023). However, for our initial DNN analysis it-
 535 eration, we assume these species have very short photolysis lifetimes, quickly breaking
 536 down into their constituent elements without further energization. In other words, only
 537 single elements are released from the surface.

538 The distribution of this flux onto Mercury's surface is influenced by the velocity
 539 of Mercury and its projection onto the surface area where the flux is calculated, as il-
 540 lustrated in Figure 4. We employ a simple relationship between the angle of incidence
 541 and the modification of the mean flux onto the surface, defined as:

$$\Delta\phi_{MIV} = \frac{V_{mm} \cos \beta_{Surf}}{V_M}, \quad (15)$$

542 where V_{mm} is the mean dust velocity, V_M is the velocity of Mercury and β_{Surf} is
 543 the angle between Mercury's velocity vector and the surface normal vector. Consequently,
 544 the incoming flux of dust particles varies between approximately 0.7×10^{-16} and $1.4 \times$
 545 $10^{-16} \text{ gcm}^{-2}\text{s}^{-1}$ on the trailing and leading sides, respectively. At the chosen modal ve-
 546 locity of the incoming flux, a constant vapor phase production rate of about 5 is taken
 547 in accordance with Cintala (1992):

$$\frac{V_x}{V_P} = c + dv + ev^2, \quad (16)$$

548 where V_x and V_P are the volumes of the released vapor and the impactors, respec-
 549 tively, v is the velocity of the impactors, and c , d , and e are constants. This simplifica-
 550 tion of the parameter subspace for this complex vaporization process is deemed sufficient
 551 for our study.

552 This approach yields outflows of surface matter ranging from 3.5×10^{-16} to $7 \times$
 553 $10^{-16} \text{ gcm}^{-2}\text{s}^{-1}$. These values are conservatively estimated to be marginally smaller than
 554 those suggested by Cintala (1992) and two orders of magnitude smaller than those pro-
 555 posed by Pokorný et al. (2017), fitting within the parameter space of interest to mod-
 556 elers without overly emphasizing this omnipresent process. Additionally, this assump-
 557 tion is considered to present a more challenging scenario for the algorithm due to the in-
 558 herent representability of the surface composition by the part of the exosphere gener-
 559 ated due to the MIV.

560 **3.3.4 Ion precipitation and Ion sputtering (SP)**

561 The SP is initiated by a flux of bombarding ions, predominantly comprising solar
 562 wind protons, which efficiently ejects atoms/molecules from the surface regolith (Wurz
 563 & Lammer, 2003; Mura et al., 2005; Wurz et al., 2010; R. M. Killen et al., 2007). The
 564 ion flux's impact is localized mainly in areas where the open magnetic field lines inter-
 565 sect the surface.

566 In our model, the flux impacting Mercury's surface is assumed proportional to the
 567 solar wind's unperturbed upstream flux of protons (Massetti et al., 2003), represented
 568 as:

$$\phi = C \rho_{sw} v_{sw}, \quad (17)$$

569 where C denotes the ratio between the cusp area at the magnetic footprint and its
 570 corresponding area in the undisturbed solar wind, set at 0.4 for our study. Here, ρ_{sw} is
 571 the solar wind density (60 cm^{-3}), and v_{sw} is the solar wind velocity (450 km s^{-1}). The
 572 calculated flux impacting the surface is $1.08 \times 10^9 \text{ cm}^{-2} \text{s}^{-1}$.

573 To derive the flux for individual species, we employ the equation from Mura et al.
 574 (2007):

$$\frac{d\Phi_n}{dE_e} = Y c \int_{E_{\min}}^{E_{\max}} \frac{d\Phi_I}{dE_i} f_S(E_e, E_i) dE_i, \quad (18)$$

575 where Y is the yield of the process, c the surface relative abundance of the species,
 576 Φ_I the ion flux, Φ_n the neutral flux emitted from the surface, E_i the impact energy, E_e
 577 the energy of the ejected particles, and f_S an empirical model for the energy distribu-
 578 tion of ejected particles, defined as:

$$f_S(E_e, T_m) = c_n \frac{E_e}{(E_e + E_b)^3} \times \left[1 - \left(\frac{E_e + E_b}{T_m} \right)^{1/2} \right], \quad (19)$$

579 with T_m as the maximum transmitted energy, c_n the normalization constant, and
 580 E_b the surface binding energy of the ejected species. T_m is:

$$T_m = E_i \frac{4m_1 m_2}{(m_1 + m_2)^2}, \quad (20)$$

581 where E_i is the impact energy, taken as constant 1000 eV.

582 For this investigation, we assume the same yield efficiency of $0.1 \times 0.35 \times 0.4$ for
 583 all species. This is a rough approximation since each element has a different binding en-
 584 ergy, resulting in different yields. Nevertheless, by considering this low yield, we aim to
 585 account for the regolith's porosity and the microshadows within it (represented by the
 586 coefficients 0.35 and 0.4, respectively). This overall low sputtering effect is a deliberate
 587 choice to complicate the prediction of surface composition by DNN algorithms in high-
 588 latitude regions receiving solar wind precipitation. The angular distribution around the
 589 normal direction of the surface is taken as $\cos^2(\alpha_n)$.

590 **3.3.5 Dynamics of the Exosphere**

591 The dynamics of the exosphere, as simulated in the Mura et al. (2007) model, en-
 592 compass the movement and behavior of particles after they have been released into the

593 exosphere, including their interactions, trajectories, and eventual fate. Factors such as
 594 gravitational influences, electromagnetic forces, and collisions are examined to under-
 595 stand how they shape the structure and composition of the exosphere.

596 Once in the exosphere, each elemental species follows ballistic trajectories under
 597 the influence of Mercury's gravity. These particles are also subject to conditions that de-
 598 fine their mean lifetime in the exosphere before impacting the surface or undergoing pho-
 599 toionization. When an exospheric particle is ionized, it is no longer simulated, making
 600 the ionization process a net loss to the exosphere.

601 Another relevant effect included in the model is solar radiation pressure, which tends
 602 to push neutral elements away from the direction of incoming sunlight, effectively propelling
 603 them toward the night side of the planet. This movement is not uniform across all species;
 604 it varies depending on the physical properties of the particles, such as their mass and
 605 effective cross-section, which influence how much momentum they absorb or reflect from
 606 solar photons. The parameters for each of the elemental species used in our exospheric
 607 model are listed in Table 4.

608 3.4 Generation of the Datasets

609 We conduct separate simulation runs for each of the four primary surface release
 610 processes, reproduced for each distinct species. Following these individual simulations,
 611 we aggregate the outcomes to compose the overall exosphere. This process essentially
 612 involves summing the resulting individual exospheres generated for each species, with-
 613 out considering interactions between the various processes. For instance, we do not ac-
 614 count for potential competition among processes for a finite pool of particles at the sur-
 615 face. Similarly, the exospheres for different species are treated as non-interacting enti-
 616 ties.

617 The number of particles launched in the Monte Carlo simulation significantly af-
 618 fects the accuracy of the results. A higher number of particles better captures the sta-
 619 tistical behavior of the populations. However, the need for more computational resources
 620 and time increases with the number of particles. For our purposes, considering the coarse
 621 resolution and the high number of simulations required, we launch 50,000 particles for
 622 each simulation run.

623 The exospheric grid is a virtual representation of space around the planet, divided
 624 into discrete cells, serving as the framework for tracking particle positions and movements.
 625 For our study, we have extended the surface 2D grid in 100 km altitude steps around the
 626 planet to a final altitude of 5000 km.

627 This section further describes the comprehensive process of dataset generation, de-
 628 tailing how we simulate measurements and observations that mimic real-world exospheric
 629 data.

630 3.4.1 Measurements and Observations Creation

631 As explained above, the creation of our datasets starts from a randomly generated
 632 surface elemental composition based on realistic mineralogies, establishing the 'ground
 633 truth' for each dataset. We then simulate environmental effects to craft a static repre-
 634 sentation of the exosphere at a specific moment, taking into account the necessary phys-
 635 ical and chemical processes.

636 Next, we strategically define various positions within the exosphere to place a vir-
 637 tual sensor, aiming to comprehensively provide "measurements" of its constituents. This
 638 conceptual sensor operates without considering the complexities and potential inaccur-
 639 acies introduced by real-world sensor characteristics such as detector noise and sensi-

640 tivity limitations. To create a comprehensive and static snapshot of the exosphere, mea-
641 surements across the simulated exosphere are conducted simultaneously.

642 The positions in the exosphere are selected to maximize data diversity and rele-
643 vance, considering factors like altitude, latitude, and environmental conditions. To bal-
644 ance the required spatial variability and facilitate the reconstruction of complete surface
645 maps from DNN predictions, we align our virtual sensor positioning in the exosphere di-
646 rectly above each surface grid tile's center. These measurements collectively form what
647 we refer to as an observation or a data subset. Each observation/subset comprises 648
648 measurement data points per altitude level. Every data point within an observation is
649 a vector that encapsulates the measured densities of all neutral species present in the
650 exosphere as per our simulation's setup. To each data point there is also the correspond-
651 ing 'ground truth' vector formed from the elemental fractions of the surface tile below.

652 In terms of altitude coverage, we define two types of data subsets. The first is a
653 constant altitude observation, with different subsets at altitudes ranging from 200 km
654 to 1500 km. For the second type, the augmented subset, we applied data augmentation
655 techniques to enhance our datasets, especially for training. This involved incorporating
656 measurements from a range of altitudes, including 200, 500, 800, 1100, 1400, 1700, and
657 2000 km for a total of 4,536 data points per subset. This approach not only increases
658 the dataset volume but also introduces a variety of exospheric conditions, allowing the
659 algorithm to develop a nuanced understanding of how measurements at different alti-
660 tudes correlate with surface compositions.

661 *3.4.2 Training, Validation and Test Datasets*

662 We produce three distinct types of datasets: training, hold-out validation, and test-
663 ing datasets. This subdivision is a fundamental practice in machine learning, ensuring
664 that the algorithm is trained on a diverse set of data, validated for accuracy, and finally
665 tested for generalization to unseen data. Different subsets of generated exospheres are
666 used for training, validating, and testing the DNN. For the training datasets, multiple
667 observations/subsets representing different exospheres are aggregated, enhancing the dataset's
668 size, complexity, and variability. Conversely, the hold-out validation and testing datasets
669 are each derived from individual observations/subsets to maintain representability of par-
670 ticular planetary surface-exosphere pairs.

671 Given the mission-specific orbital characteristics of the BepiColombo/Mercury Plan-
672 etary Orbiter and the enhanced measurement capabilities of the STROFIO mass spec-
673 tometer within the SERENA instrument suite at lower altitudes, we define a baseline
674 constant altitude subset at 500 km. This altitude, corresponding to the periherm (clos-
675 est approach) of the MPO, is chosen for its potential to yield accurate measurements with
676 an improved signal-to-noise ratio. In our training campaign we examine both the base-
677 line dataset type with examples only at altitudes of 500 km (648 examples per subset)
678 and the augmented dataset type (4,536 examples per subset), the latter of which pro-
679 vides the MLP DNN with the capability to make predictions at diverse altitudes.

680 The training phase demands a complex strategy that captures a larger part of the
681 data distribution to train the algorithm to approximate the relationships between the
682 employed physical processes accurately. To address this, we aggregate multiple data sub-
683 sets or observations to form the training dataset, with the number of included subsets
684 ranging from 10 baseline observations (yielding a total of 6,480 data points) to 300 aug-
685 mented observations (resulting in a staggering 1,360,800 data points). The multiple train-
686 ing sets used in our study vary in size and complexity and present different empirical dis-
687 tributions to the DNN. Training the DNN on each set produces different estimators of
688 the processes, each tailored to the particular empirical distribution of its training set.
689 The construction of a representative training dataset plays a crucial role in ensuring the
690 resulting estimator of physical processes closely approximates the actual data distribu-

691 tion. Additionally, increasing the training set size reduces the probability of the algo-
 692 rithm becoming biased towards a non-representative smaller dataset distribution.

693 For the validation and testing phases of our DNN algorithms, we adopt a differ-
 694 ent strategy. Hold-out validation involves using a single validation dataset to monitor
 695 the trained network during each training epoch. This process helps identify when the
 696 algorithm begins to overfit the training data, thereby reducing its generalization capa-
 697 bilities. The validation set is also crucial during hyperparameter tuning to evaluate the
 698 algorithm's ability to generalize effectively. At the same time, the test datasets consist
 699 of examples (data points) that are never seen during training, ensuring that the MLP
 700 DNN algorithm's performance evaluation is not biased by improvements in accuracy due
 701 to evaluation on previously seen examples.

702 For validation and test dataset purposes, employing single data subsets allows for
 703 straightforward prediction and reconstruction of surface maps corresponding to individ-
 704 ual surface-exosphere simulations. Our hold-out validation dataset is selected at the MPO's
 705 periherm altitude of 500 km (baseline subset) to maintain consistency in evaluation con-
 706 ditions.

707 The test datasets are designed to thoroughly assess the algorithm's predictive ca-
 708 pabilities under varied conditions. To ensure a thorough assessment of the algorithm's
 709 performance, we generate a total of 135 test datasets derived from 15 distinct exosphere
 710 simulations, each at a constant altitude ranging from 200 km to 1500 km. This diver-
 711 sified testing ground allows us to explore the algorithms' responsiveness to altitude varia-
 712 tions. Such an approach facilitates a comprehensive analysis and evaluation across a
 713 broader spectrum of exospheric altitudes and conditions, enabling us to more accurately
 714 gauge the algorithm's efficacy and robustness in predicting the surface elemental com-
 715 position. A more detailed description of the datasets and their underlying surfaces is given
 716 in Appendix A.

717 *3.4.3 Feature Selection and Engineering*

718 The final step is to supplement the datasets with additional metadata, or features,
 719 from these observations, such as geometrical information, to serve as auxiliary inputs for
 720 the algorithms. The effectiveness of deep neural network algorithms in modeling com-
 721 plex relationships within data is significantly influenced by the selection and engineer-
 722 ing of input features. Carefully chosen or engineered, these features, describing each data
 723 point within all three types of datasets, can enhance the algorithm's ability to discern
 724 patterns and relationships, thereby improving its overall performance.

725 In our study, we have considered and incorporated a range of additional features
 726 to enrich our datasets:

- 727 • **Altitude of Measurement:** This feature is critical for capturing altitude-specific
 728 dynamics, enabling the algorithm to identify how the distribution of neutral species
 729 changes with altitude relative to their source points on the surface. Both the ac-
 730 tual altitude and a logarithm of the altitude were tested as features. The latter
 731 engineered feature is intended to highlight non-linear altitude effects on the mea-
 732 sured parameters, providing another layer of depth to the altitude-related anal-
 733 ysis.
- 734 • **Logarithm of Exospheric Density:** By applying the base 10 logarithm to the
 735 exospheric density of each species, we introduce a constraint that aids the algo-
 736 rithm in exploring nonlinear relationships, acknowledging the exponential decrease
 737 in density with altitude.
- 738 • **Subsolar Angle:** Represented as either the value of the angle ϕ directly, or as
 739 $\cos(\phi - 180)$, this feature helps differentiate between exospheric populations on

740 the dayside, nightside, and the transitional terminator regions, enhancing the model's
 741 spatial awareness.

- 742 • **Latitude Dependency:** Using $\sin(\gamma)$, this feature allows the algorithm to ac-
 743 count for latitude-specific phenomena, such as ion sputtering, which vary across
 744 different latitudinal zones.
- 745 • **Proton Flux Virtual Data:** Integrating virtual measurements akin to those from
 746 the MIPA ion detector of the SERENA suite, this feature hints at sputtering ef-
 747 fects induced by proton precipitation on the surface, offering a proxy for under-
 748 standing underlying ion induced physical processes.

749 Each data point in our datasets is defined by combinations of these features, form-
 750 ing distinct feature sets that illuminate to the algorithm various aspects of the exosphere's
 751 behavior. The compilation of these feature sets is crucial for unraveling the capabilities
 752 of neural networks in predicting surface compositions and contributing insights into the
 753 mechanisms governing particle release into the exosphere.

754 4 Results

755 In this section, we show the findings of our investigation, which are divided into
 756 two distinct phases to provide a comprehensive understanding of our study's outcomes.
 757 The first, training phase focuses on the configuration and optimization of the deep neu-
 758 ral network. This entails a systematic exploration of the hyperparameter space and other
 759 method characteristics to identify the optimal settings that enhance the DNN's ability
 760 to model the data accurately. The second, testing phase evaluates the performance of
 761 the DNN, now finely tuned with the optimal hyperparameter configuration, in interpret-
 762 ing and making predictions on data derived from unseen during learning examples.

763 4.1 Training Phase and DNN Finalization

764 An extensive training campaign was undertaken to explore both the empirical dis-
 765 tribution represented in the training datasets, and the hyperparameter space of the neu-
 766 ral network architecture. This effort aimed to develop an accurate estimator that demon-
 767 strates optimal generalization capabilities by closely approaching the true data gener-
 768 ating distribution. In this section we outline the most important findings of this inves-
 769 tigation phase to ultimately refine and finalize the components of the multilayer percep-
 770 tron deep neural network. The complete training campaign is detailed in Appendix B.

771 *Eliminating Skewed Predictions*

772 Initial analysis showed that accuracy metrics for predicting surface elemental com-
 773 position were skewed by the high prevalence of oxygen (O_2) and silicon (Si). To address
 774 this, we excluded these elements from the prediction vector, adjusting the model to fo-
 775 cus on the normalized proportions of the remaining nine elements. This adjustment im-
 776 proved the model's relevance and performance by aligning with our study's objectives
 777 more effectively.

778 *Training Set Size and Data Augmentation*

779 We analyzed MLP DNN performance in relation to the expansion of the training
 780 dataset size with training sets ranging from 10 to 200 unaugmented baseline data sub-
 781 sets, observing that larger datasets improved generalization accuracy. Additionally, in
 782 order to better approximate the true data-generating distribution, we augmented our train-
 783 ing datasets with examples varying in altitude, enhancing representability. This strate-
 784 gic choice expanded our dataset to 300 augmented subsets, totaling 1,360,800 examples,
 785 significantly improving the model's robustness and predictive accuracy across altitude-
 786 specific inputs.

787 *Learning Curve Examination for Optimal Training Duration*

788 Our examination of the MLP DNN's learning curves aimed to identify the optimal
 789 training duration to avoid overfitting, in line with the guidance provided by Bengio (2015).
 790 Analysis indicated that predictive performance on the validation dataset peaked at 40
 791 epochs, as shown in Figure 5. While longer training durations, up to 200 epochs, con-
 792 tinued to align the model to the training dataset, the best balance between training and
 793 inference accuracy was achieved at 40 epochs, suggesting this as the optimal training du-
 794 ration.

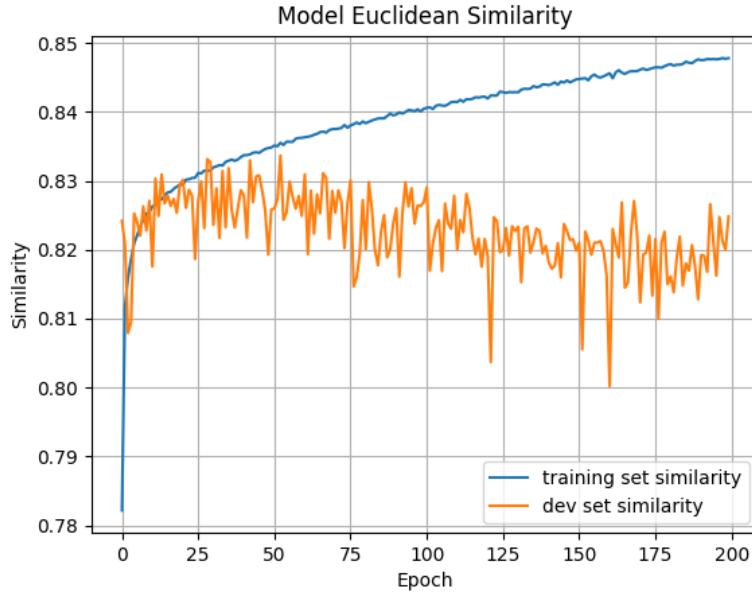


Figure 5. Learning curve for the MLP DNN training. The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (300 subsets, 1,360,800 data points) and the hold-out validation dataset (1 subset, 648 data points) respectively.

795 *Selected Feature Set*

796 After rigorous testing and evaluation, the feature set that emerged as superior, of-
 797 fering the most consistent and highest accuracy, comprised of the following features:

- 798 • Logarithmic transformations of elemental species exospheric densities.
- 799 • Logarithm of the altitude at which measurements were taken.
- 800 • Sun incidence angle.
- 801 • Presence of H+ ions arriving through open field lines.
- 802 • Local time.
- 803 • Latitude.

804 *Hyperparameter Optimization and DNN Structural Components Finalization*

805 An extensive hyperparameter optimization effort resulted in the selection of the fi-
 806 nal MLP DNN architecture, consisting of a four-layer structure with 600, 500, 350, and
 807 250 neurons in each layer respectively (Figure 6). The regularization coefficient was op-
 808 timized to a higher value of 1.0×10^{-5} to improve generalization, while the learning rate

809 was finely tuned to 0.5×10^{-4} . Training was conducted in mini-batches of 512 examples,
 810 identified as near-optimal through our optimization process.

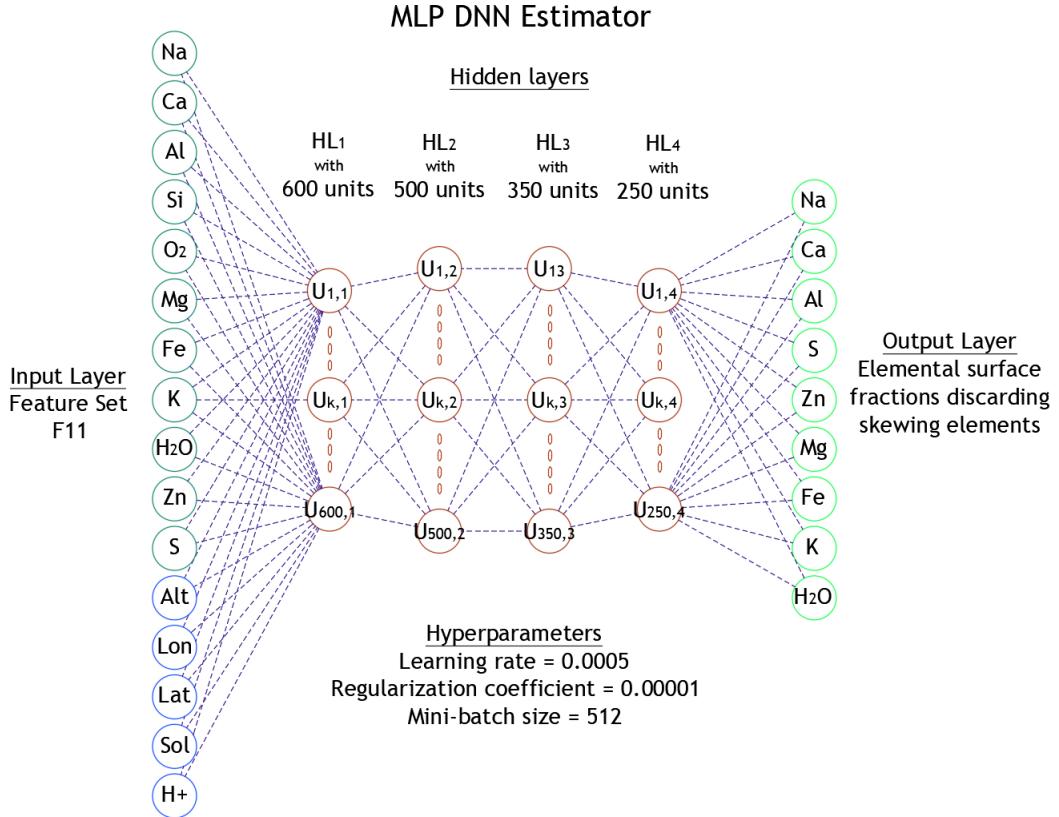


Figure 6. Finalized multilayer perceptron deep neural network. The input layer consists of the features collected in feature set F11, the output layer is adjusted to remove the skewing predictions towards better estimating the O₂ and Si species. There are four hidden layers with 600, 500, 350, and 250 neurons respectively.

811 The combined efforts of hyperparameter exploration, architectural fine-tuning, and
 812 dataset augmentation have significantly propelled our model's performance and its abil-
 813 ity to generalize from the empirical distribution to the true underlying data-generating
 814 distribution—by as much as a total of 10% in ES4 and up to 30% in R-squared. This
 815 optimized architecture, along with structural parameters and algorithmic characteris-
 816 tics refined during our comprehensive training campaign, ensures that the MLP DNN
 817 is a robust model for our sophisticated predictive tasks.

818 4.2 Testing Phase

819 The ultimate evaluation of our multilayer perceptron deep neural network algorithm's
 820 performance hinges on its ability to accurately predict surface compositions and recon-
 821 struct elemental surface maps using datasets it has not previously encountered. Our re-
 822 search incorporates two distinct test campaigns designed to assess the MLP network's
 823 predictive prowess. These campaigns were structured to apply the final network, fine-
 824 tuned with an extensive training set comprising 300 augmented subsets, across test datasets
 825 derived from a variety of altitudes not previously seen during training.

The scope of these test campaigns is broad, focusing not only on aggregate performance metrics across the entire dataset but also on detailed analyses for individual elemental species. This includes a thorough examination of residuals to identify any systematic errors or biases in predictions, with the final goal of reconstructing the surface composition maps for each species. This process entails a visual comparison between the original, or "ground truth", maps and the ones predicted by our algorithm.

4.2.1 Preliminary Test Campaign

In our preliminary test campaign, we embarked on a performance evaluation using single-simulation test datasets derived from 15 unique surface compositions, each leading to distinct exospheres. This approach encompassed data from both the dayside and nightside, allowing for a robust examination of the MLP DNN algorithm's predictive accuracy and its capability in reconstructing surface elemental maps under varying conditions.

The campaign tested the algorithm's performance across a spectrum of altitudes ranging from 200 km to 1500 km. This setup provided a rich dataset for analysis, comprising 15 sets of predictions for each of the 9 altitude levels, culminating in a total of 135 complete prediction sets. These predictions detailed the fractional composition of nine elements across the surface grid tiles, facilitating the reconstruction of elemental maps for the 15 different surfaces from measurements at each altitude level.

We utilized our suite of performance metrics, including the average ES4, R-squared, absolute, and relative residuals, to evaluate the predictions and reconstructions systematically. These metrics were plotted against the measurement altitudes to analyze the model's performance comprehensively, depicting them for the overall predicted output, individual elemental species, and separate analyses for dayside and nightside predictions (Figure 7).

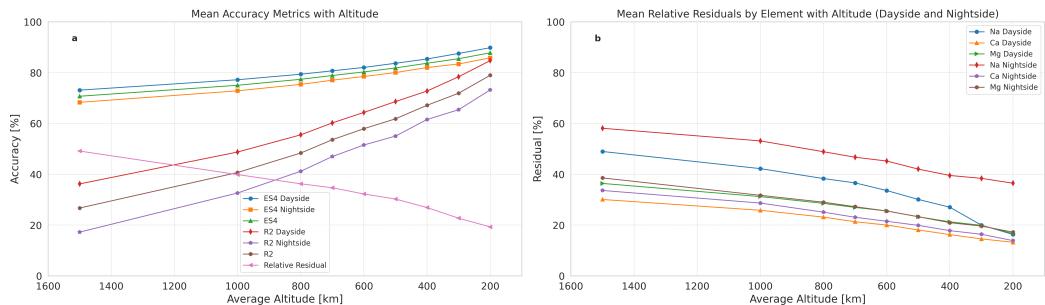


Figure 7. Preliminary test campaign. Panel **a** shows the mean accuracy (dayside, nightside, and overall) and relative residuals metrics of the MLP DNN predictions on the 15 test surfaces. In panel **b** are displayed the dayside and nightside mean relative residuals for the elements Na, Ca, and Mg.

The results of this preliminary testing suggested that predictions were generally more precise for the dayside, likely a consequence of particle movements influenced by solar radiation pressure. This was especially true for volatile species, such as sodium, on which we observed a notable discrepancy in predictive accuracy with respect to refractory ones, like magnesium and calcium.

856 **4.2.2 Main Test Campaign**

857 We note that Mercury has a unique rotational and orbital dynamics, particularly
 858 its 3:2 orbit-spin resonance. This results in the nightside hemisphere becoming the day-
 859 side in every following orbit. Thus, by using a combination of data from two consecu-
 860 tive years, we can fill the gap in the predictions of the volatile species, as leveraged in
 861 the main test campaign. This approach utilized double-simulation compound predictions,
 862 focusing on the same 15 surface compositions from the preliminary campaign but ob-
 863 served at two consecutive perihelia. During these two periods, different halves of Mer-
 864 cury's surface were illuminated by the Sun, allowing for comprehensive daylight obser-
 865 vation of the entire planet over the two simulations. For this campaign, predictions specif-
 866 ically targeted sunlit surface tiles, enabling an in-depth analysis of surface compositions
 867 that were previously on the nightside in the initial test phase. Measurements for this cam-
 868 paign were again taken at a range of altitudes from 200 to 1500 km.

869 A significant outcome of the combined odd-even orbit campaign was the improved
 870 accuracy in predicting volatile species' fractions, aligning more closely with the refrac-
 871 tory species' predictions observed in the preliminary campaign. This enhancement in pre-
 872 dictive accuracy for volatiles under daylight conditions underscores the importance of
 873 solar illumination in accurately assessing surface compositions.

874 The main campaign demonstrated a notable increase in overall prediction and map
 875 reconstruction accuracy, with the average ES4 metric reaching approximately 89.70% and
 876 the average R-squared metric reaching 83.41% at the lowest altitude of 200 km (Figure
 877 8). This accuracy diminished at higher altitudes, attributed to the exosphere's dynamic
 878 nature and the increased complexity in tracing back exospheric particles to their orig-
 879 inating surface tiles. There was, however, a marked improvement compared to the pre-
 880 liminary findings, which highlights the efficacy of considering Mercury's solar exposure
 881 in enhancing predictive models' accuracy. By focusing solely on the dayside observations
 882 across two perihelia, the campaign effectively capitalized on optimized conditions for sur-
 883 face composition reconstruction. The box statistical plots shown on Figure 8 suggest also
 884 reduction in the range of prediction accuracies and residuals with decrease in altitude.

885 Going deeper in the detailed statistics of the predictions by the MLP DNN for each
 886 elemental species, we can observe its tendencies in the box plots of the absolute and rel-
 887 ative predictions for Aluminium, Calcium, and Sodium (Figure 9). Box plots for the re-
 888 maining elements are provided in Appendix C. Almost all elements have good predic-
 889 tion statistics at the lowest altitude, with the median relative residual of Aluminium par-
 890 ticularly impressive at only 8.69% at 200 km. The algorithm has more difficulties with
 891 Sodium, with its median relative residual at 17.86% at 200 km, which is nevertheless a
 892 good result. The robustness in the prediction of the refractive elements is present through-
 893 out the altitudes, even up to 1500 km, where the median relative residuals of Aluminium
 894 is 21.18%, while that for Calcium is 30.76%. While the range of the prediction errors for
 895 the different types of surface elements is impressive throughout the altitudes.

896 Additionally, our visual comparisons from the map reconstructions (Figures 10 to
 897 12) highlighted the algorithm's strengths and weaknesses in predicting different elemen-
 898 tal distributions.

899 The discrepancy in predictive accuracy between refractory species, such as Aluminium
 900 (Figure 10) and Calcium (Figure 11), and volatile ones, like Sodium (Figure 12), is still
 901 present in the predictions of the particular test set 2 shown on the figures. However, there
 902 is a noticeable improvement by the daylight only predictions of the main testing campaign,
 903 compared to the one from the preliminary campaign, which included the night side pre-
 904 diction. This is shown on the map reconstructions of Sodium (test set 2), where the bottom-
 905 most panels in Figure 12 show that increased errors on the night side (longitudes 0-90
 906 and 270-360) and an average relative residual of 23.73% from the preliminary test, di-

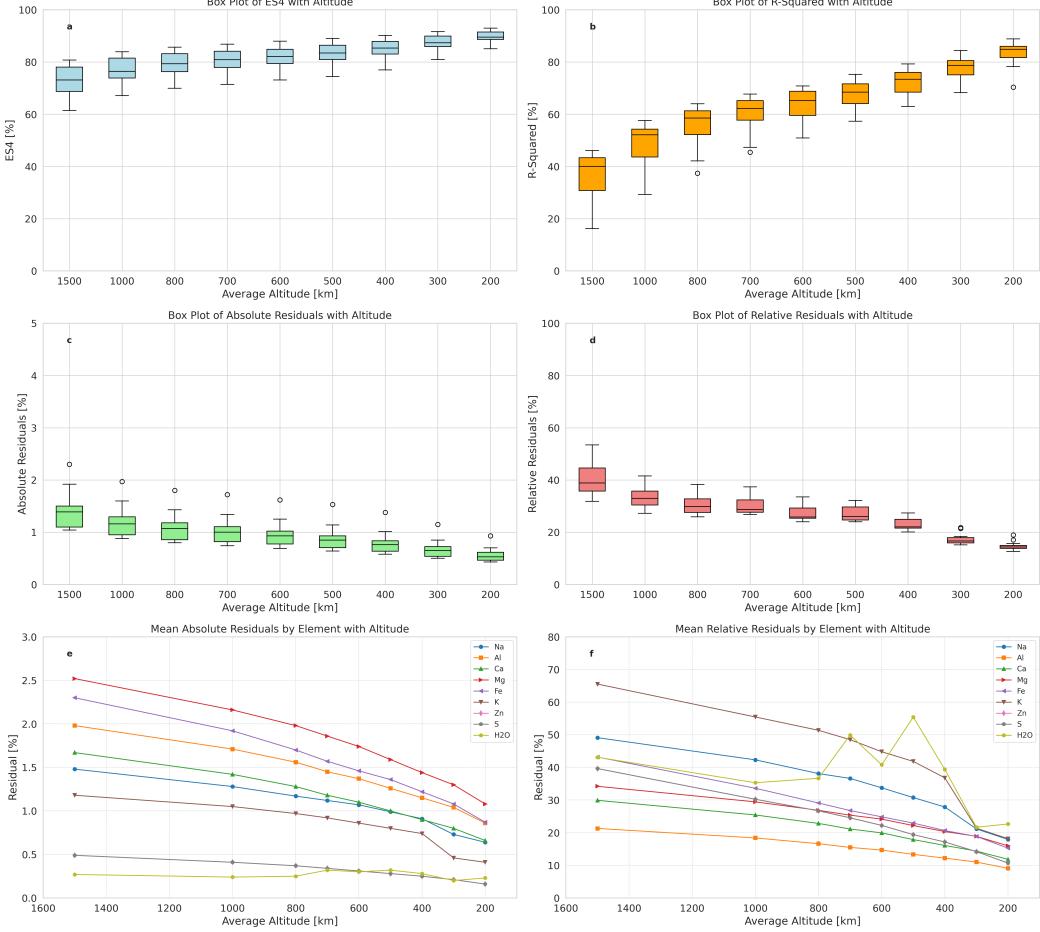


Figure 8. Box plots of the accuracy metrics (panels **a** and **b**) and the mean residuals (panels **c** and **d**) of the MLP DNN predictions on the 15 test surfaces of the main test campaign. Panels **e** and **f** give respectively the absolute and relative residuals for all predicted elements averaged over the 15 predicted surfaces.

rectly compared to the average of the main campaign at 500 km (third row from top) where the average relative residual is reduced to 20.62%.

From a closer examination of the reconstructed maps, we can clearly see that certain large scale patterns in the surface composition of all species are recognized from as far as 1500 km, even if details are missed by the MLP at that large distance. The algorithm starts to resolve with a relatively good accuracy at altitudes of 500-800 km. This is especially true for the refractory species (Ca, Al, etc.), and going closer to the planet (down to 200 km) produces the most accurate predictions and reconstructions, even for volatiles (Na). The full set of reconstructed maps for this particular test set number 2 in our main test campaign is provided in Appendix C.

5 Discussion and Future Work

This research presents a novel approach for deducing Mercury's surface composition using advanced deep neural networks to analyze exospheric density measurements. The algorithm is adept at predicting relative elemental compositions and reconstruct-

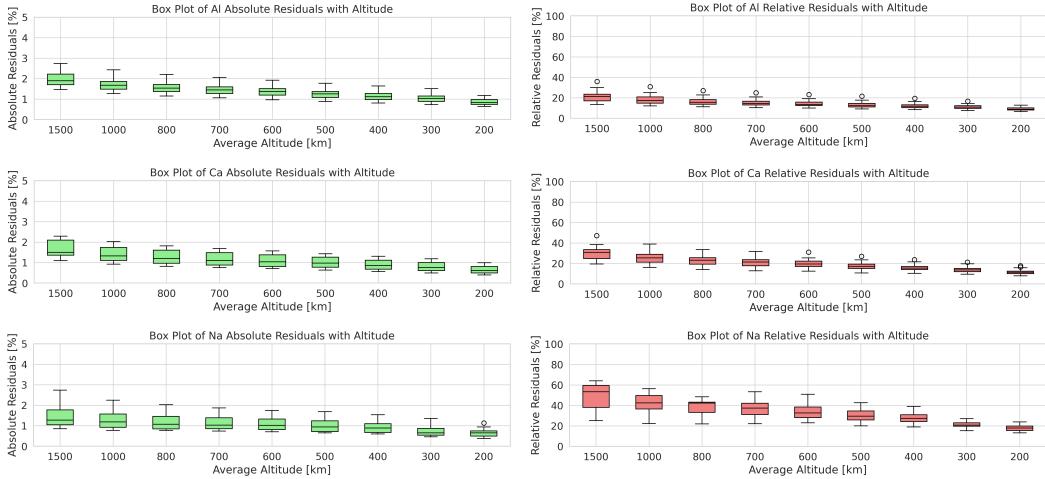


Figure 9. Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the elements Aluminium, Calcium, and Sodium.

ing surface maps, crucial for understanding how neutral atoms transit from the planet's regolith into its exosphere. Key characteristics of the method include the definition and utilization of the simulated model's physical parameter space, the development of a sophisticated multilayer perceptron DNN architecture, the application of Bayesian hyperparameter tuning for optimal configuration, and the integration of domain-specific knowledge into the feature selection process.

The principal accomplishment of this study is the creation of a DNN that accurately estimates surface-exosphere interactions on Mercury. This algorithm effectively represents a specific region of the physical parameter space, mapping complex relationships between exospheric and surface compositions. By yielding an estimator for exosphere generation processes, the study bridges theoretical modeling with practical data-driven predictions. This has the potential to enhance our understanding of Mercury's exosphere and offers new avenues for research in planetary science and machine learning, suggesting potential for significant future discoveries.

A preliminary test phase demonstrated the model's superior predictive performance on Mercury's dayside but highlighted less precise predictions for elements like sodium and potassium on the nightside. These results show that some elemental species have a stronger link with the surface, retained even on the nightside (refractory elements), while others are more strongly connected to the surface on the dayside (volatile elements) due to exospheric effects like radiation pressure, making nightside predictions less straightforward. Meanwhile, other species, such as oxygen and silicon, being widely present on the surface, have less significance in the analysis. This led to a focused test campaign using only simulated daylight observations from two consecutive Mercury years, leveraging the planet's spin-orbit resonance. This approach significantly improved predictions for volatile species and enabled a comprehensive reconstruction of Mercury's surface.

The developed algorithms for surface composition reconstruction achieved high fidelity in their predictions, with mean accuracy metrics reaching 89.70% (ES4) and 83.41% (R-squared) across 15 test sets at 200 km altitude. Additionally, both mean absolute and relative residuals of elemental predictions showed a robust trend of decreasing with altitude, reaching as low as 0.56% and 14.70% respectively. The effectiveness of these algorithms was particularly notable in reconstructing elemental composition maps from low-altitude measurements (200km to 700km), especially for sufficiently represented el-

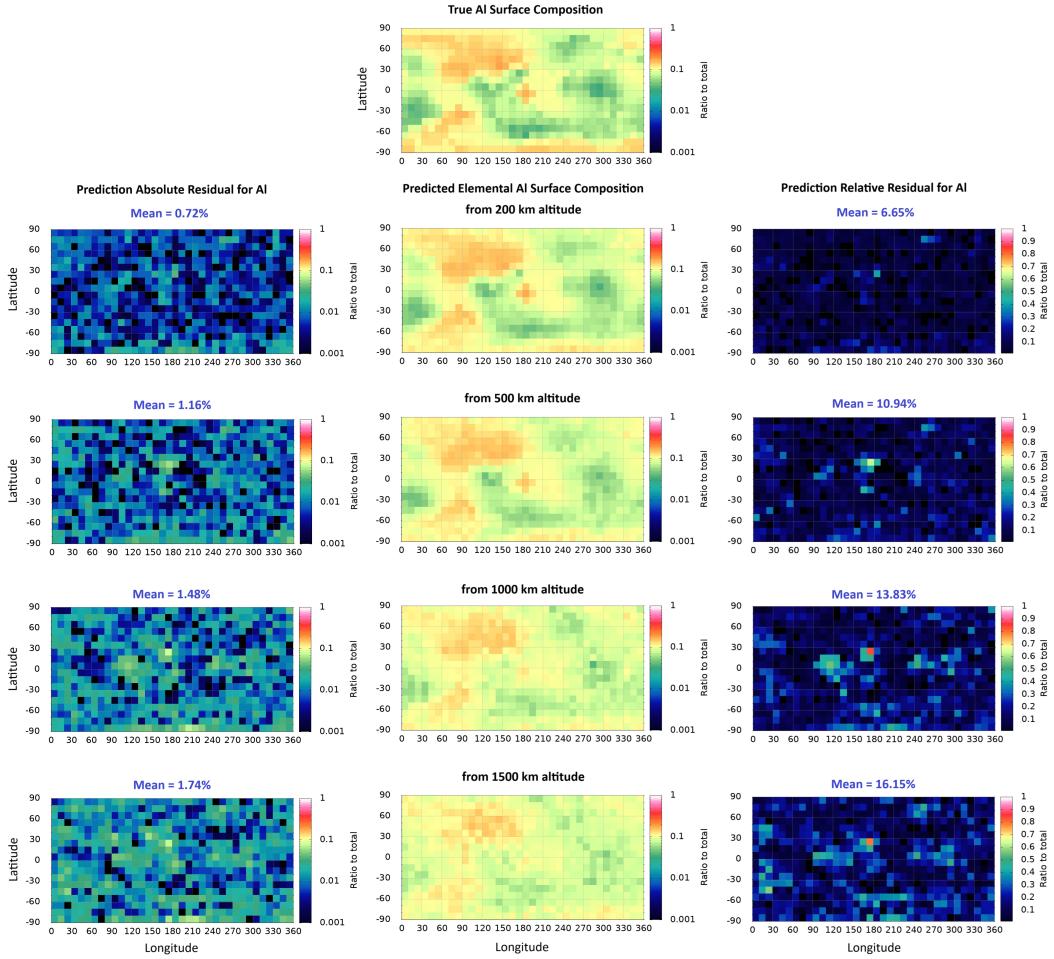


Figure 10. Main test campaign - MLP DNN map reconstructions of the same sample Aluminium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

elements, such as aluminium, calcium, magnesium, and sodium. At the same time, minor elements, such as water, exhibited jumps and anomalies in their predictions, attributed to their low overall fraction in our model and their randomized placement on the surface (not constrained to the poles only). Meanwhile, the MLP was able to capture the coupling of sulfur and zinc in its predictions, even if the input exospheric densities are widely different for the two elements.

We must note that the presented method is far from a finalized tool for the analysis of surface-exosphere interactions. The extensive training and testing campaigns conducted in this study highlight significant potential for refining the algorithms, while the data generation models can be further improved with even more detailed representations of the physical processes.

Future work includes expanding and refining the physical processes parameter space used in the data generation model. This endeavor will align the algorithm more closely

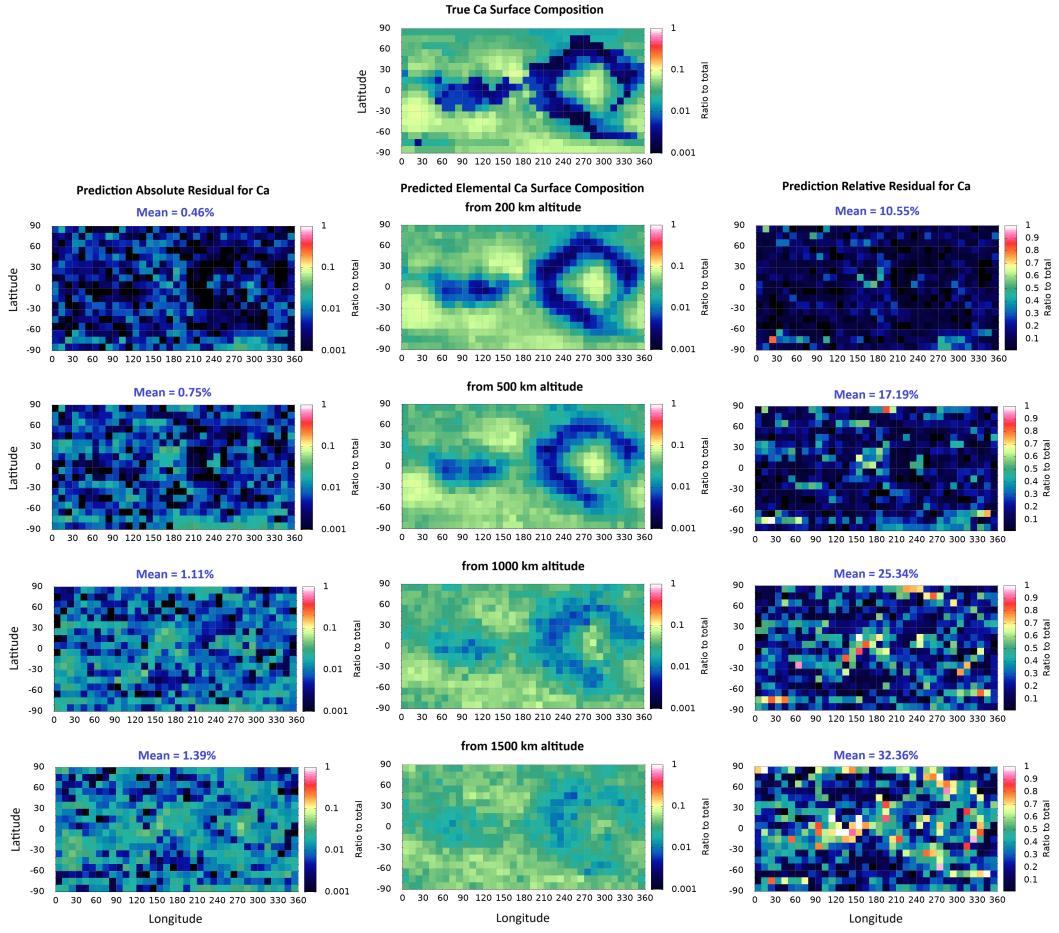


Figure 11. Main test campaign - MLP DNN map reconstructions of the same sample Calcium surface composition (test set number 2). Inputs are coming from the dayside of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

with the complex realities governing interactions between a planet's surface and its exosphere. A key part of this exploration includes modeling previously compressed dimensions within the parameter space, such as the complex chemistry within the MIV-produced cloud, more complex impacting dust populations, and more involved regolith effects on the release processes (e.g. diffusion), all aiming to more accurately reflect the intricate dynamics of real-world processes. Unraveling these dimensions, previously simplified in our model, is expected to provide deeper insights into planetary science mechanisms and enhance the estimator's ability to capture the full spectrum of surface-exosphere interactions. Moreover, we aim to explore the performance of the DNNs on data distributions derived from more constrained surface models that closely follow the observed mineralogy, elemental composition, and distribution of elements, especially water, on the surface.

Another critical area of future research is testing DNNs trained with data from one region of the physical processes parameter space against data distributions from differ-

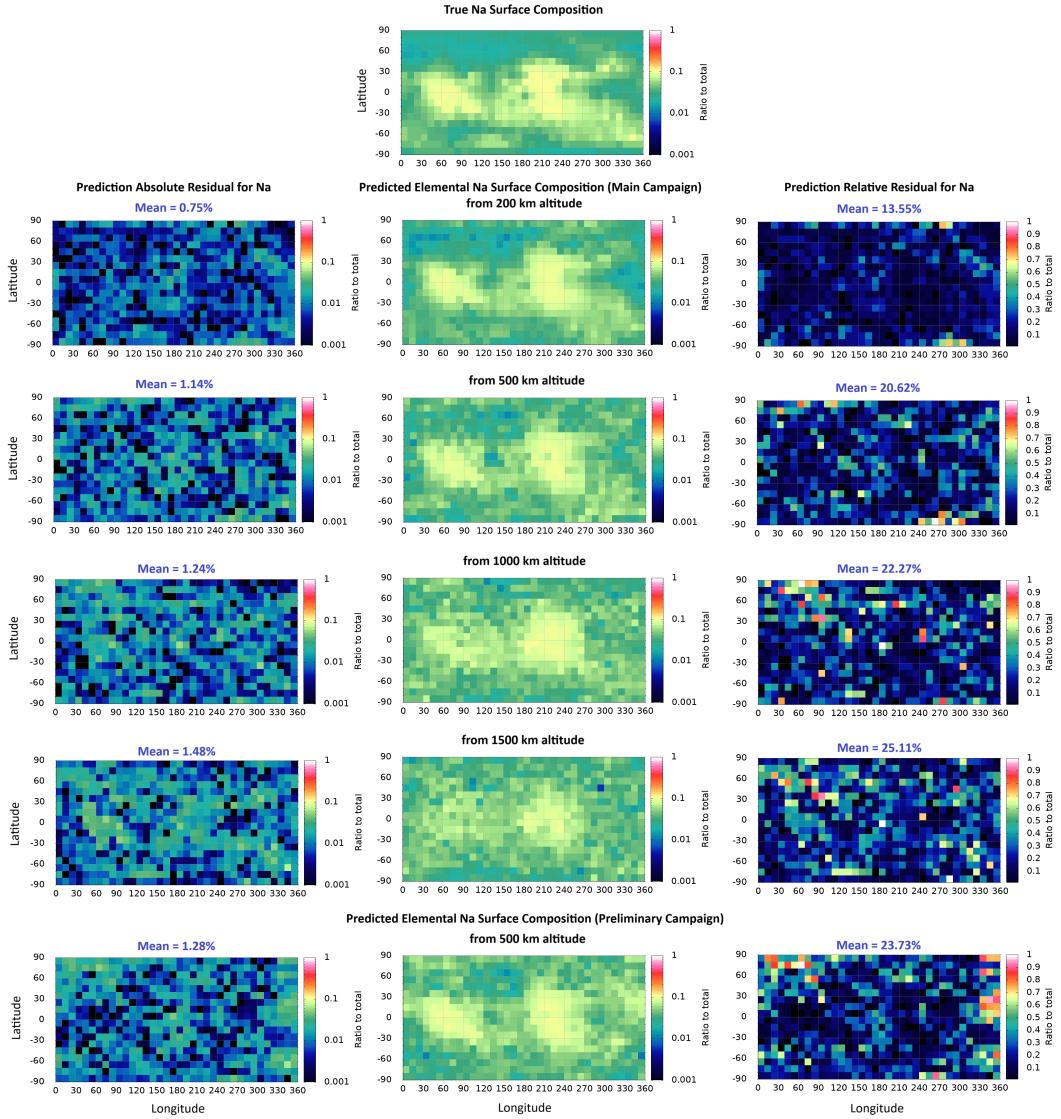


Figure 12. Main test campaign - MLP DNN map reconstructions of the same sample Sodium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals. Eclipsed areas in the preliminary campaign (bottom-most panels, 0-90 and 270-360 Lon) are markedly more accurately reconstructed in these dayside only combined maps of the main campaign.

ent regions. This exploration is essential to assess the DNNs' performance when applied to various models of exospheric production and their respective data distributions, potentially constructing another layer for analyzing the surface-exosphere interactions with this innovative tool.

One potential application of this approach involves utilizing multiple DNNs trained on different regions of parameter space. Input data could be passed through all of the

986 pre-built DNNs, and their predictions juxtaposed in a subsequent layer of the algorithm
 987 to estimate the input data's underlying generation mechanism. This would constrain the
 988 range of physical parameters to those of the DNNs with the highest accuracy on the in-
 989 put data.

990 Another potentially groundbreaking utilization may explore the mapping between
 991 the physical processes space, defined by analytical equations, and the purely data-driven
 992 DNN representation, built from the internal weight matrices of the neural nets. This map-
 993 ping could serve the dual purpose of understanding how a description of reality constructed
 994 strictly from data relates to the description by physical equations, and exploring poten-
 995 tial synergies between the two in describing the real world.

996 Additionally, expanding and elaborating on the hyperparameter space is identified
 997 as another area for development. This will involve constructing a hyperparameter space
 998 that considers aspects such as network layer connectivity, optimization of loss functions,
 999 and the functions used within the hidden and output units. Exploring alternative DNN
 1000 architectures also holds promise for enhancing the models' application, accuracy, and re-
 1001 liability. Further research into feature engineering by applying more domain-specific knowl-
 1002 edge to optimize input parameters can provide a better representation of empirical data
 1003 distributions. Observations of discrepancies between dayside and nightside predictions
 1004 may warrant an examination of split DNNs trained on data from only one side of the
 1005 planet (illuminated or shadowed). These developments aim to push the boundaries of
 1006 what these algorithms can achieve in surface composition analysis.

1007 The contrast between using simulated data and incorporating real observational
 1008 data into our algorithm development merits further exploration. While simulations pro-
 1009 vide a controlled environment for testing various scenarios, they do not capture the full
 1010 complexity and unpredictability of actual exospheric data. This could lead to significant
 1011 deviations in the parameter space from those assumed in our simulations, as real pro-
 1012 cesses and their interdependent variables may change over time and are not fully rep-
 1013 resented in simulations.

1014 Shifting our focus from simulated to real physical processes is a bold and poten-
 1015 tially transformative step. Developing an estimator capable of effectively processing and
 1016 analyzing real-world data from Mercury's exosphere would significantly advance our un-
 1017 derstanding of planetary surfaces and their interactions with their environments. This
 1018 progress would not only deepen our theoretical knowledge but also offer practical insights
 1019 into the formation, dynamics, and evolution of planetary exospheres.

1020 This study is performed in anticipation of the upcoming BepiColombo ESA/JAXA
 1021 mission. The mission will deploy two spacecraft—the Mercury Planetary Orbiter (MPO)
 1022 and the Mercury Magnetospheric Orbiter (MMO)—equipped with a suite of instruments
 1023 aimed at understanding Mercury's surface, exosphere, and magnetosphere (Benkhoff et
 1024 al., 2010; Milillo et al., 2010, 2020). Particularly, our study targets future utilization of
 1025 measurements from the SERENA (Search for Exospheric Refilling and Emitted Natu-
 1026 ral Abundances) suite on the MPO, which includes sensors like STROFIO (STart from
 1027 a ROTating Field mass spectrometer), MIPA (Miniature Ion Precipitation Analyser),
 1028 PICAM (Planetary Ion CAMera), and ELENA (Emitted Low Energy Neutral Atoms)
 1029 to analyze exospheric gas composition, plasma fluxes, and ion precipitation (Orsini et
 1030 al., 2010, 2021; Milillo & Wurz, 2014). Notable instruments aboard BepiColombo, which
 1031 may provide images of the surface, include MIXS, MGNS, MERTIS, and SIMBIO-SYS
 1032 (Benkhoff et al., 2010).

1033 The application of these methods to the observational data from BepiColombo's
 1034 suite of instruments offers a promising path to refine models of exosphere generation. By
 1035 comparing predicted surface compositions with actual measurements, we can more ac-
 1036 curately constrain our models, enhancing our understanding of planetary processes.

1037 In conclusion, this research establishes a solid foundation for advancing our understanding of planetary surface-exosphere interactions, particularly around Mercury. By
 1038 utilizing exospheric measurements as inputs to deep neural networks, we've taken a significant step forward, enhancing the capabilities of estimators and broadening our understanding of planetary science. The application of this method to the anticipated data
 1040 collected by the BepiColombo mission will represent a notable advance in space exploration.
 1041 With sophisticated AI algorithms, BepiColombo's potential to uncover insights into Mercury's exosphere dynamics will be greatly enhanced. Moreover, the ongoing development and refinement of deep neural networks in this study promise to revolutionize
 1043 our approach to studying planetary bodies within our Solar System, providing new tools for understanding the complex processes that govern the environments of celestial objects.
 1045

1049 Appendix A Detailed Datasets

1050 A1 Mean Mineral and Elemental Fractions in the Datasets

Surfaces	Albite	Anorthite	Diopside	Enstatite	Ferrosilite	Hedenbergite	Orthoclase	Sphalerite	Water Ice
Training x10	0.102	0.170	0.146	0.156	0.116	0.079	0.110	0.089	0.032
Training x20	0.109	0.136	0.162	0.149	0.126	0.072	0.131	0.077	0.038
Training x40	0.136	0.124	0.148	0.142	0.133	0.065	0.139	0.065	0.046
Training x60	0.133	0.124	0.153	0.137	0.152	0.062	0.137	0.056	0.046
Training x80	0.134	0.121	0.147	0.134	0.154	0.066	0.138	0.060	0.046
Training x100	0.131	0.127	0.144	0.134	0.151	0.070	0.139	0.058	0.048
Training x150	0.133	0.133	0.143	0.139	0.137	0.066	0.135	0.067	0.047
Training x200	0.139	0.134	0.143	0.136	0.133	0.065	0.133	0.072	0.045
Training x300	0.140	0.134	0.141	0.137	0.137	0.065	0.134	0.069	0.044
v01	0.042	0.310	0.088	0.119	0.103	0.061	0.101	0.062	0.114
t01	0.095	0.116	0.090	0.120	0.397	0.017	0.107	0.020	0.038
t02	0.384	0.124	0.131	0.098	0.053	0.020	0.138	0.032	0.021
t03	0.099	0.150	0.202	0.124	0.085	0.114	0.124	0.053	0.050
t04	0.148	0.182	0.173	0.132	0.088	0.069	0.139	0.034	0.034
t05	0.125	0.054	0.063	0.027	0.325	0.020	0.087	0.241	0.058
t06	0.097	0.113	0.095	0.153	0.095	0.140	0.116	0.149	0.043
t07	0.241	0.075	0.191	0.144	0.200	0.011	0.089	0.033	0.017
t08	0.074	0.146	0.064	0.073	0.219	0.153	0.178	0.051	0.043
t09	0.151	0.146	0.127	0.063	0.086	0.019	0.338	0.025	0.045
t10	0.285	0.176	0.196	0.182	0.055	0.022	0.047	0.020	0.017
t11	0.261	0.222	0.104	0.242	0.047	0.012	0.085	0.007	0.021
t12	0.172	0.237	0.187	0.208	0.086	0.010	0.074	0.008	0.018
t13	0.295	0.259	0.140	0.125	0.076	0.018	0.065	0.009	0.013
t14	0.178	0.240	0.131	0.157	0.040	0.008	0.166	0.069	0.011
t15	0.186	0.265	0.145	0.109	0.105	0.010	0.102	0.068	0.010

Table A1. Mean surface mineral fractions in the datasets used for training, validation, and testing.

1051 A2 Consolidated Datasets

Surfaces	Na	Al	Si	O _z	Ca	Mg	Fe	K	Zn	S	H ₂ O
Training x10	0.014	0.074	0.275	0.421	0.055	0.066	0.045	0.015	0.015	0.015	0.005
Training x20	0.015	0.069	0.279	0.421	0.052	0.066	0.047	0.018	0.014	0.014	0.007
Training x40	0.018	0.069	0.282	0.422	0.047	0.062	0.048	0.019	0.011	0.011	0.010
Training x60	0.017	0.068	0.284	0.424	0.047	0.061	0.052	0.018	0.009	0.009	0.009
Training x80	0.018	0.068	0.284	0.424	0.046	0.059	0.053	0.018	0.010	0.010	0.009
Training x100	0.017	0.069	0.283	0.424	0.047	0.059	0.053	0.019	0.010	0.010	0.009
Training x150	0.018	0.071	0.282	0.423	0.048	0.060	0.049	0.018	0.011	0.011	0.009
Training x200	0.018	0.072	0.282	0.422	0.048	0.060	0.047	0.018	0.012	0.012	0.008
Training x300	0.019	0.072	0.282	0.423	0.047	0.059	0.048	0.018	0.012	0.012	0.008
v01	0.006	0.109	0.257	0.421	0.067	0.047	0.039	0.014	0.010	0.010	0.021
t01	0.012	0.058	0.285	0.430	0.031	0.046	0.113	0.014	0.003	0.003	0.006
t02	0.048	0.095	0.302	0.435	0.034	0.042	0.016	0.017	0.004	0.004	0.003
t03	0.013	0.071	0.280	0.426	0.066	0.064	0.041	0.016	0.008	0.008	0.007
t04	0.019	0.085	0.284	0.431	0.056	0.059	0.034	0.018	0.005	0.005	0.005
t05	0.020	0.053	0.263	0.386	0.023	0.019	0.106	0.016	0.052	0.052	0.012
t06	0.014	0.062	0.269	0.405	0.054	0.058	0.051	0.016	0.032	0.032	0.008
t07	0.031	0.062	0.295	0.431	0.038	0.064	0.056	0.011	0.005	0.005	0.003
t08	0.010	0.069	0.283	0.426	0.050	0.030	0.087	0.023	0.008	0.008	0.007
t09	0.019	0.098	0.298	0.435	0.037	0.033	0.025	0.043	0.003	0.003	0.006
t10	0.035	0.084	0.289	0.434	0.051	0.076	0.017	0.006	0.003	0.003	0.002
t11	0.030	0.097	0.286	0.436	0.043	0.078	0.014	0.011	0.001	0.001	0.003
t12	0.020	0.090	0.280	0.435	0.056	0.081	0.024	0.009	0.001	0.001	0.003
t13	0.034	0.106	0.284	0.438	0.052	0.051	0.022	0.008	0.001	0.001	0.002
t14	0.023	0.105	0.281	0.431	0.049	0.059	0.012	0.021	0.009	0.009	0.002
t15	0.024	0.105	0.277	0.431	0.055	0.048	0.029	0.013	0.009	0.009	0.001

Table A2. Mean surface elemental fractions in the datasets used for training, validation, and testing.

Test and Validation Datasets						Total # of Datasets from Surf-Exo pair
Surf-Exo Pair Name	Resulting Dataset Type	# of Examples Per Dataset	Mercury TAA	Altitude Range [km]		
v01	Validation	648	0	500		1
t01	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t02	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t03	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t04	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t05	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t06	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t07	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t08	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t09	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t10	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t11	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t12	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t13	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t14	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18
t15	Test	648	0 and 360	200, 300, 400, 500, 600, 700, 800, 1000, 1500		18

Table A3. Test and validation datasets prepared for the generalization evaluation of the MPL DNN algorithm. Each test surface-exosphere pair gives rise to one dataset per altitude level and per Mercury TAA. The "0 and 360" TAA signifies that two simulations are performed on this surface coming from two consecutive perihelia.

Training Sets				
# Subsets	Data Augmentation	Examples per Subset	Total Examples	Altitude Range [km]
10	No	648	6,480	500
20	No	648	12,960	500
40	No	648	25,920	500
60	No	648	38,880	500
80	No	648	51,840	500
100	No	648	64,800	500
150	No	648	97,200	500
200	No	648	129,600	500
300	No	648	194,400	500
200	Yes	4,536	907,200	200-2000
300	Yes	4,536	1,360,800	200-2000

Table A4. Main training datasets used to train the MLP DNN algorithm. Each training dataset defines a different empirical data distribution which its respective MLP DNN is trained to approximate.

1052 **Appendix B Detailed Training Campaign**

1053 *Starting Point*

1054 Our journey in the big training campaign endeavour commenced with a baseline train-
 1055 ing dataset, which, being the simplest, was least representative of the true data gener-
 1056 ation distribution. This dataset comprised 10 subsets with a total of 6,480 data points
 1057 from the F00 feature set (only exospheric density measurements). The initial architec-
 1058 ture of the MLP DNN was modeled closely after the structure proposed in the prelim-
 1059 inary study of Kazakov et al. (2020), featuring a four-layered network with 400, 200, 200,
 1060 and 100 neurons, respectively. Notably, our study expanded the input layer to accom-
 1061 modate a greater number of elements - 11 total input elements.

1062 The output layer of the network employs softmax units, designed to predict the sur-
 1063 face composition of the same 11 elements provided as inputs. Initially, the minibatch size
 1064 was set to 1,024 examples. The regularization L-2 coefficient and the learning rate were
 1065 chosen as 1.0×10^{-6} and 0.5×10^{-4} , respectively, to balance the trade-off between learn-
 1066 ing efficiency and the risk of overfitting.

1067 *Eliminating Skewed Predictions*

1068 The initial analysis of predicting fractionated surface elemental composition revealed
 1069 that the accuracy metrics were significantly skewed by the prevalence of certain abun-
 1070 dant elements, notably oxygen (O_2) and silicon (Si), which are omnipresent in most of
 1071 the minerals in our model. This skewness, stemming from the algorithm's propensity to
 1072 more accurately predict these two elements, was addressed by excluding them from the
 1073 prediction vector and adjusting it to ensure a normalized sum of 1. Consequently, the
 1074 refined model focuses on predicting the normalized proportions of the remaining nine el-
 1075 ements, with a subsequent denormalization process applied for the map reconstruction
 1076 purposes. This strategic exclusion of the most abundant elements led to a marked en-
 1077 hancement of approximately 4% in the predictive R^2 accuracy for the other nine elements.
 1078 It is important to note, however, that the input vector maintained its original compo-
 1079 sition of 11 elements.

1080 This decision to modify the output layer by removing two elements was driven by
 1081 a clear rationale: the omnipresent elements, though significant, held less interest for the
 1082 objectives of our study compared to the other elements. This approach underscores our
 1083 commitment to optimizing the model's performance where it matters most, despite rec-
 1084 ognizing that alternative configurations of the output layer might exist.

1085 *Training Set Size and Learning Curve Examination*

1086 Exploring the behavior of the initial MLP DNN involved examining its perfor-
 1087 mance in relation to the expansion of the training dataset size and the extension of train-
 1088 ing duration. The aim was to demonstrate the algorithm's nominal operation during both
 1089 training and inference phases by analyzing its learning curves. This included assessing
 1090 training and generalization accuracies across a training dataset and the hold-out vali-
 1091 dation dataset, respectively. Additionally, identifying the optimal training duration for
 1092 inference was crucial to mitigate the risk of overfitting, in line with the guidance pro-
 1093 vided by Bengio (2015).

1094 Our investigation spanned training sets ranging from 10 to 200 unaugmented data
 1095 subsets. We observed a clear positive relationship between increasing the dataset size
 1096 and the enhancement of generalization accuracy.

1097 In parallel, the algorithm's behavior was monitored in terms of its optimization pro-
 1098 cess over multiple iterations (epochs) across the entire training dataset, employing stochas-
 1099 tic gradient descent to converge to the minimum of the loss function. Analysis of the learn-

ing curves revealed a maximum in predictive performance on the validation dataset after 40 epochs. This was in contrast to the outcomes observed at 200 epochs of SGD, despite the training dataset's distribution increasingly aligning with each additional training iteration.

The learning curve depicted in Figure B1 also hints at the potential for further enhancements in training predictions, given the rapid ascent observed towards the training's culmination. However, to ensure robust inference capabilities, it's imperative to diminish the variance. This could potentially be achieved by incorporating a greater number of training examples and/or intensifying the regularization measures.

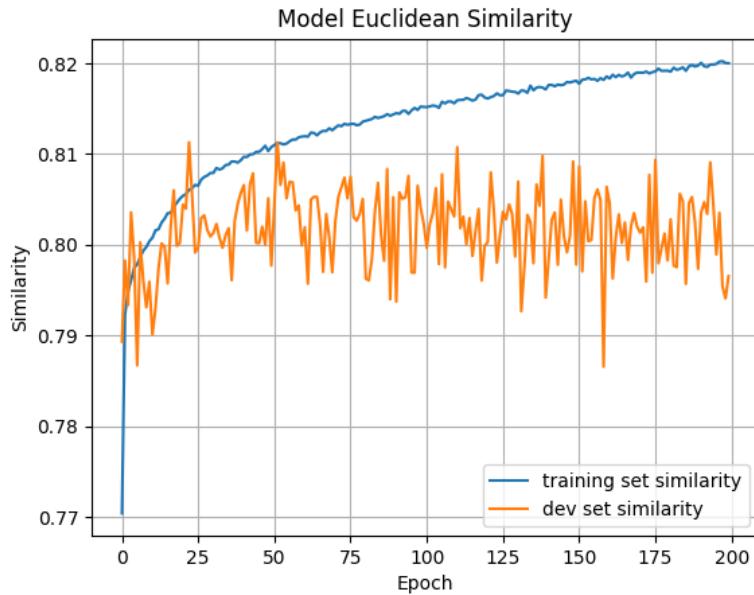


Figure B1. Learning curve for the MLP DNN training. The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (200 subsets, 129,600 data points) and the development hold-out validation dataset (1 subset, 648 data points) respectively.

1109 Feature Sets Examination

The evaluation of feature sets played a pivotal role in optimizing the performance of the MLP DNN, particularly through the training of the network with various input features across the unaugmented 200-subset training dataset. The assortment of the examined feature sets is detailed in Table B1. The iterative process of enhancing input features resulted in substantial improvements in prediction accuracy for several modified feature sets with ES4 going up by about 2%, while R^2 increasing with as much as 5% from F00 to F11. This underscores the critical importance of a well-curated and comprehensive feature set in the development of neural networks capable of tackling intricate tasks, such as predicting the surface elemental compositions.

After rigorous testing and evaluation, the feature set that emerged as superior, offering the most consistent and highest accuracy, was F11. This feature set encompasses: (1) logarithmic transformations of elemental species exospheric densities, which provide a normalized scale for comparing densities of various elements; (2) the logarithm of the altitude at which measurements were taken, introducing a scale that accommodates the

Feature Sets Explored						
Feature Set Name	Exospheric Densities	Altitude	Longitude	Latitude	Local Solar Time	Ion Precipitation
F00	linear	-	-	-	-	-
F01	linear	linear	-	-	-	-
F02	logarithmic	-	-	-	-	-
F03	logarithmic	logarithmic	-	-	-	-
F04	logarithmic	-	cos	sin	-	-
F05	logarithmic	linear	cos	sin	-	-
F06	logarithmic	logarithmic	cos	sin	-	-
F08	logarithmic	linear	-	-	linear	linear
F09	logarithmic	-	cos	sin	linear	linear
F10	logarithmic	linear	cos	sin	linear	linear
F11	logarithmic	logarithmic	cos	sin	linear	linear

Table B1. Feature sets used in training and testing of the MLP DNN algorithm.

wide range of altitudes without skewing the data; (3) Sun incidence angle, accounting for the variation in solar energy impacting the elemental composition; (4) the presence of H+ ions arriving through open field lines, a feature indicating solar wind interaction with the planetary surface; (5) Cosine of solar time longitude, offering a representation of the position in solar time longitude; and (6) Sine of latitude, providing a function to capture latitudinal variations.

The selection of F11 as the final feature set was predicated on its ability to yield the most reliable and accurate predictions, thereby encapsulating the intricate dynamics and characteristics vital for elemental composition analysis. This feature set's efficacy highlights the nuanced approach required in feature selection to enhance neural network performance for specific predictive tasks. All the future training and testing were performed with this feature set as inputs to the DNN.

Hyperparameter Optimization and DNN Structural Components Finalization

In our quest to fine-tune the multilayer perceptron for optimal performance, a significant focus was placed on hyperparameter optimization. This process was critically informed by the parameters outlined in Section 2.2, employing a Bayesian search strategy to navigate the hyperparameter space efficiently. Our methodology involved running the optimization process five times, with each iteration spanning 50 cycles and starting from a point incrementally closer to the previously identified minimum, for a total of 250 cycles. This approach was instrumental in inching towards the optimal hyperparameter settings, with subsequent iterations yielding diminishing returns, indicative of approaching a plateau near the optimal values in the hyperparameter space.

During this campaign, the selection of loss functions emerged as a pivotal consideration, with our experiments revealing substantial variations in model performance across different functions. The discerning application of loss functions, particularly the adoption of the KL-divergence for evaluating probability-like outputs, marked a jump in performance.

The culmination of our hyperparameter optimization efforts led to the finalization of the MLP DNN architecture, characterized by a four-layered structure with 600, 500, 350, and 250 neural units respectively (Figure 6). An adjustment was made to the regularization coefficient, setting it to the found higher value of 1.0×10^{-5} to enhance model generalization. Concurrently, the learning rate was optimized to 0.5×10^{-4} , balancing the trade-off between learning speed and stability. Training was conducted in mini-batches of 512 examples, a size determined through our optimization exercises to be close to optimal. This meticulously optimized structure and parameter set represent the culmina-

1159 tion of our comprehensive campaign to refine the MLP DNN, ensuring it stands as a ro-
 1160 bust model for our advanced predictive task.

1161 *Augmented Data Study*

1162 In the concluding phase of our training campaign, we embarked on a strategic ini-
 1163 tiative to enhance the representability of the empirical distribution, thereby aligning it
 1164 more closely with the true data-generating distribution—a target that remains inherently
 1165 elusive due to limited direct access. This endeavor was pursued through the deliberate
 1166 augmentation of our training datasets, an approach that involved the integration of ad-
 1167 ditional examples derived from the same exospheric observations that constituted our
 1168 initial datasets. However, these new inclusions were distinct in their representation of
 1169 varying altitudes, thereby enriching the diversity and depth of our training data.

1170 The initial expansion of our dataset to encompass 200 augmented subsets had al-
 1171 ready demonstrated significant promise in enhancing the model’s performance. Motivated
 1172 by these preliminary successes, we ambitively expanded our dataset even further to in-
 1173 clude a total of 300 augmented subsets, culminating in an impressive 1,360,800 exam-
 1174 ples. This substantial augmentation effort was driven by the rationale that incorporat-
 1175 ing measurements from varying altitudes would not only bolster the dataset’s compre-
 1176 hensiveness but also empower our model to predict with greater accuracy across a di-
 1177 verse range of altitude-specific inputs.

1178 The outcome of this labor were unmistakably positive, with the augmented datasets
 1179 markedly improving the robustness and accuracy of our MLP DNN, increasing further
 1180 the validation set’s ES4 to 84.0% (+1.5%) and its R^2 to 63.5% (+3.5%). The strategic
 1181 inclusion of altitude-varied examples was particularly impactful, enabling the algorithm
 1182 to achieve enhanced predictive precision for inputs across different altitudes.

1183 *Implications and Results of the Training*

1184 As our meticulous exploration of the hyperparameter space culminated in identi-
 1185 fying a region that, while not conclusively the ultimate minimum, demonstrates unpar-
 1186 alleled accuracy in inferences on the hold-out validation dataset, we arrived at several
 1187 pivotal implications and results from our training campaign. This journey through hy-
 1188 perparameter optimization has yielded a collection of finely tuned multilayer perceptron
 1189 deep neural networks, each reflecting a nuanced understanding of the underlying data-
 1190 generating processes.

1191 Firstly, one of the outcomes of this campaign is the demonstration of the algorithm’s
 1192 efficiency, achieving optimal training within 40 complete epochs. This not only highlights
 1193 the effectiveness of our chosen architecture but also underscores the potential for accu-
 1194 racy improvements with the expansion of the training dataset. Such findings affirm the
 1195 architectural decisions made in designing our MLP DNN for the task at hand.

1196 Secondly, our investigation revealed the critical role of specific features in guiding
 1197 the algorithm toward more precise predictions of exospheric measurements and surface
 1198 composition. The identification of these key features underscores the importance of thought-
 1199 ful feature selection in enhancing model performance.

1200 Thirdly, the exploration led to the refinement of the MLP’s internal structure, sig-
 1201 nificantly bolstered by experiments with various loss functions and output layers, along-
 1202 side the application of Bayesian hyperparameter optimization. While acknowledging that
 1203 the realm of possible architectural enhancements remains vast, the current configuration
 1204 stands as a testament to the robustness and efficacy of our model.

1205 Lastly, the strategic augmentation of our dataset with additional exospheric ob-
 1206 servations has unequivocally improved the algorithm’s predictive capabilities. This ex-

1207 pansion not only enriches the model’s training environment but also enhances its abil-
1208 ity to generalize across a broader spectrum of the empirical distribution, thereby mov-
1209 ing closer to the elusive true data-generating distribution.

1210 The combined efforts of hyperparameter exploration, architectural fine-tuning, and
1211 dataset augmentation have significantly propelled our model’s performance. Through
1212 this comprehensive training campaign, we have not only achieved a high degree of ac-
1213 curacy in our predictions but also laid a solid foundation for future research to build upon,
1214 promising even greater advancements in our understanding and representation of com-
1215 plex data-generating processes. Through this concerted effort, we have significantly ad-
1216 vanced the model’s capacity to generalize from the empirical distribution to the true un-
1217 derlying data-generating distribution.

1218
1219

Appendix C Box Plots and Map Reconstructions of the Main Test Campaign Predictions by Species

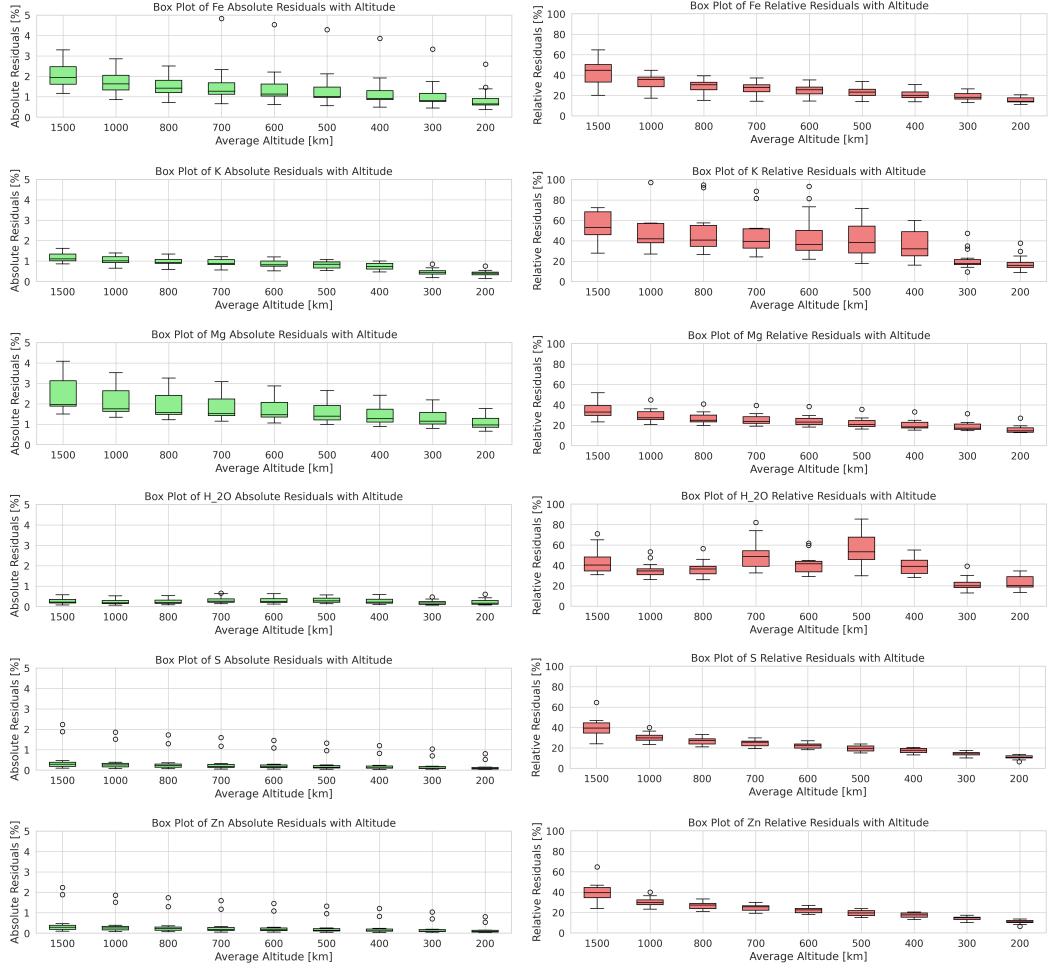


Figure C1. Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Iron, Potassium, Magnesium, Water, Sulfur, and Zinc.

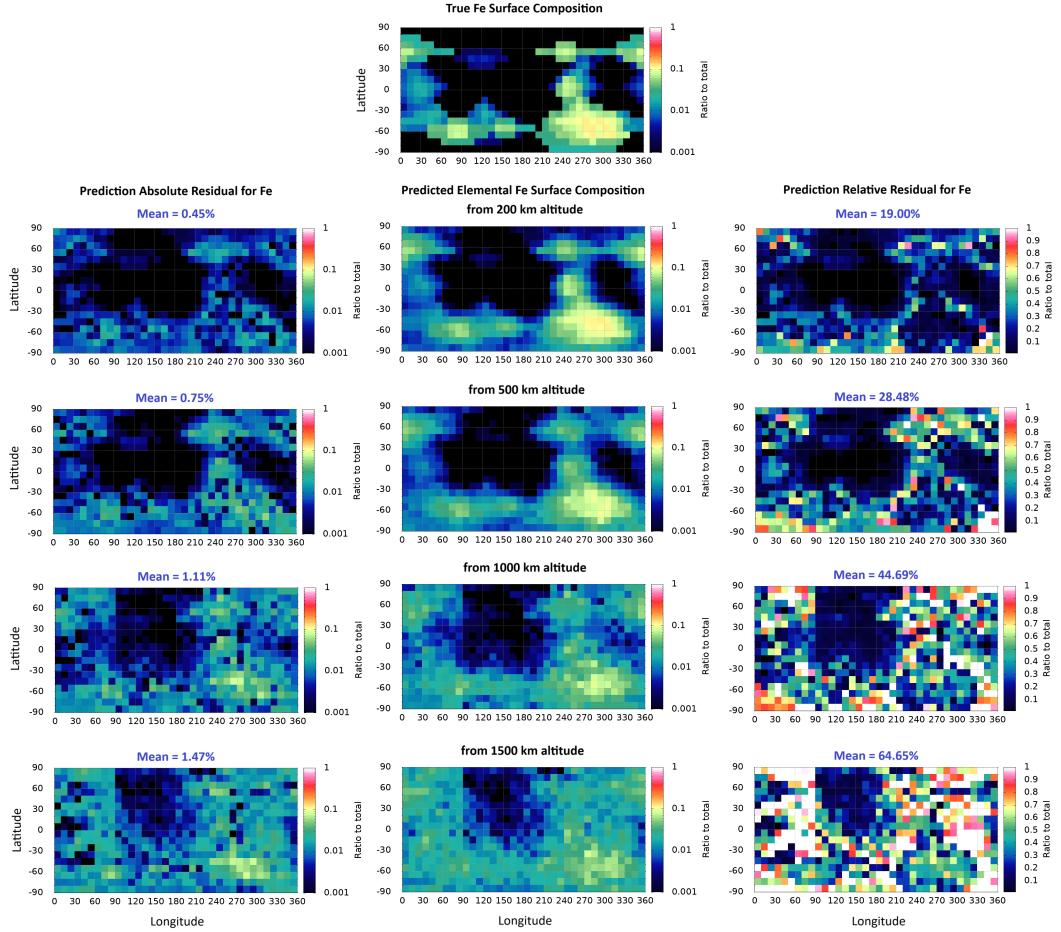


Figure C2. Main test campaign - MLP DNN map reconstructions of the same sample Iron surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

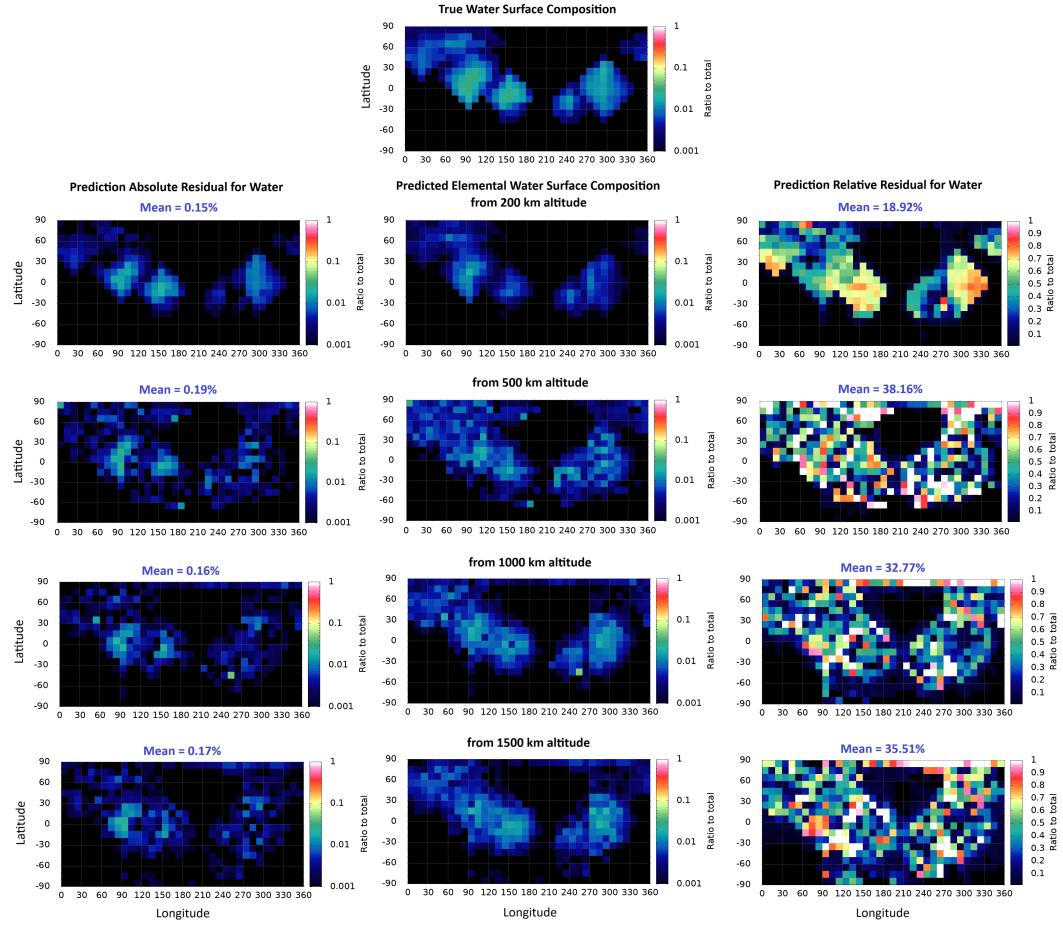


Figure C3. Main test campaign - MLP DNN map reconstructions of the same sample Water surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

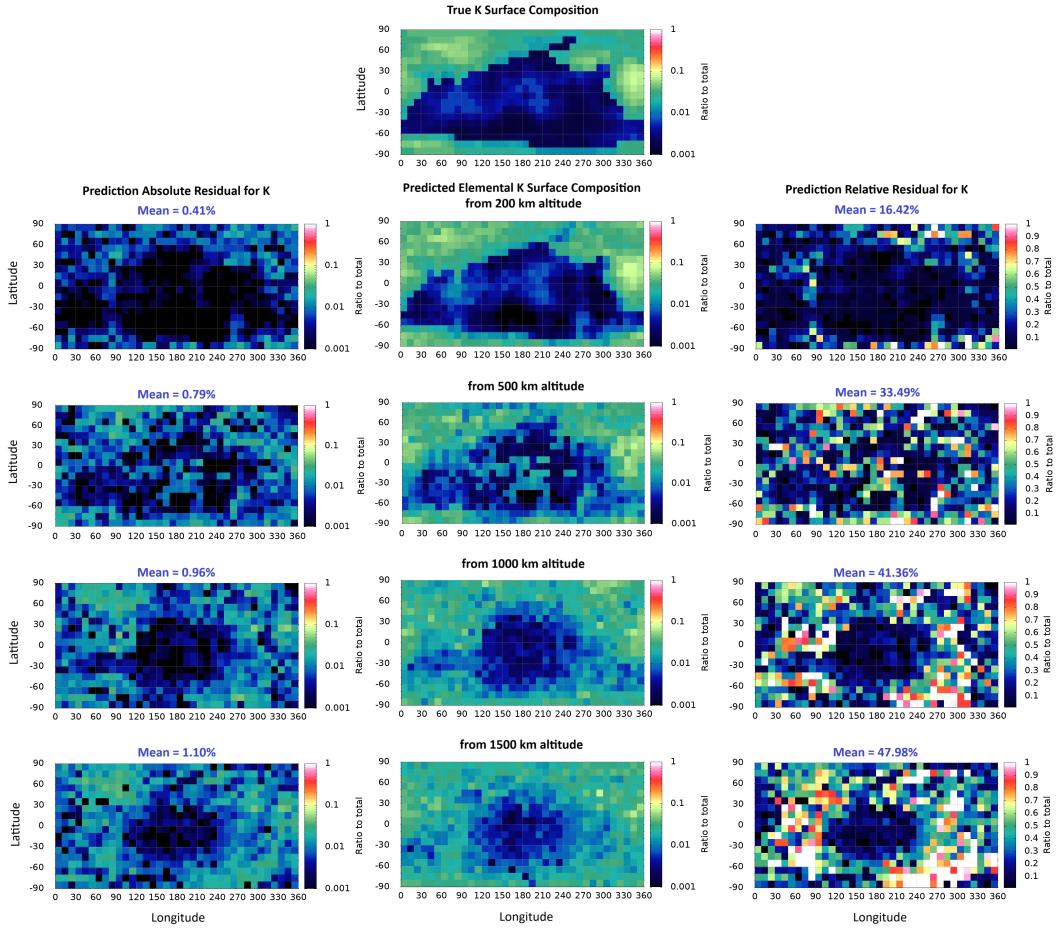


Figure C4. Main test campaign - MLP DNN map reconstructions of the same sample Potassium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

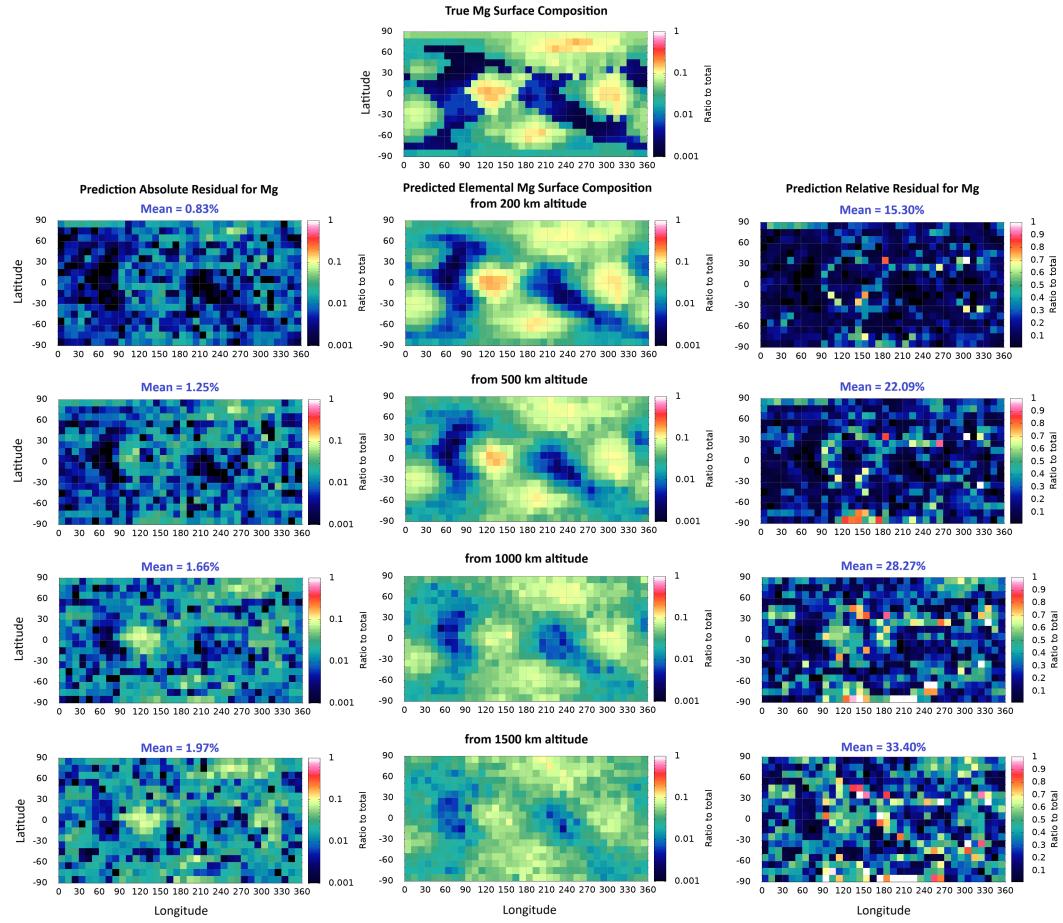


Figure C5. Main test campaign - MLP DNN map reconstructions of the same sample Magnesium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

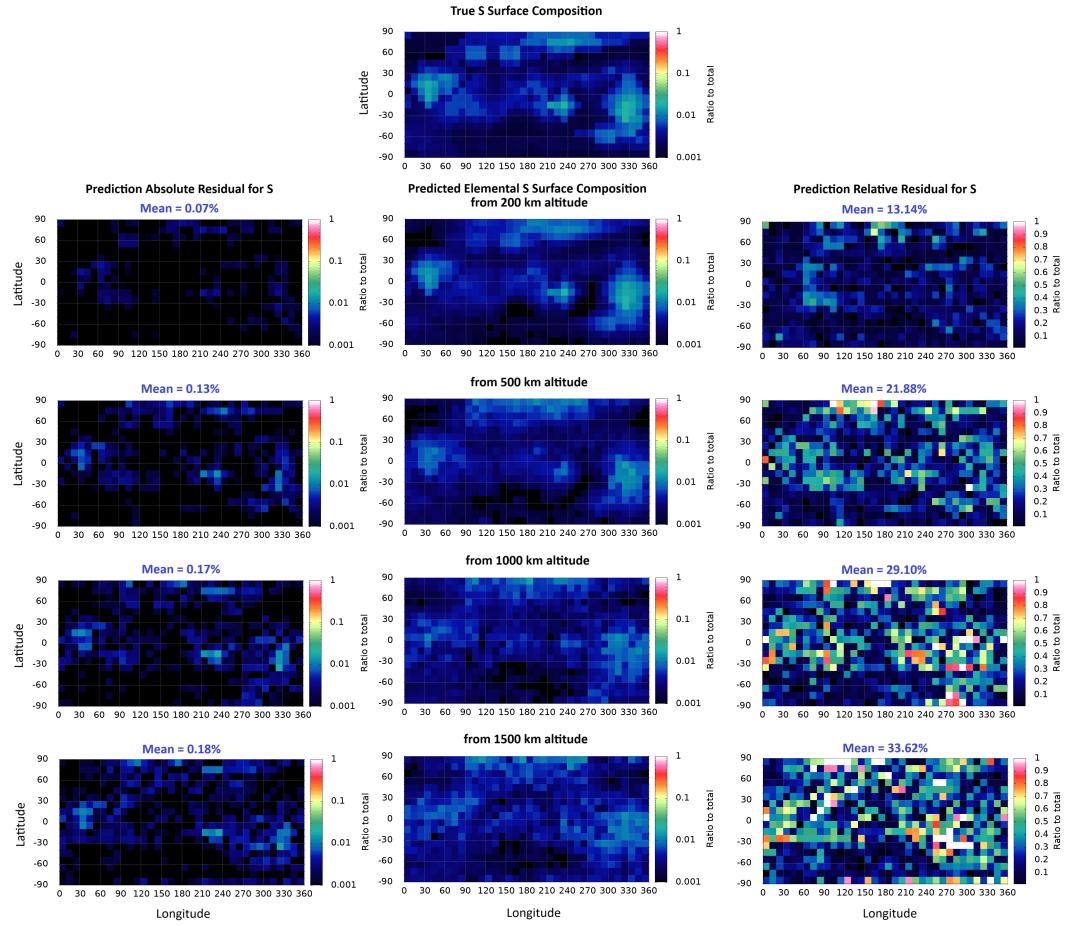


Figure C6. Main test campaign - MLP DNN map reconstructions of the same sample Sulfur surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals.

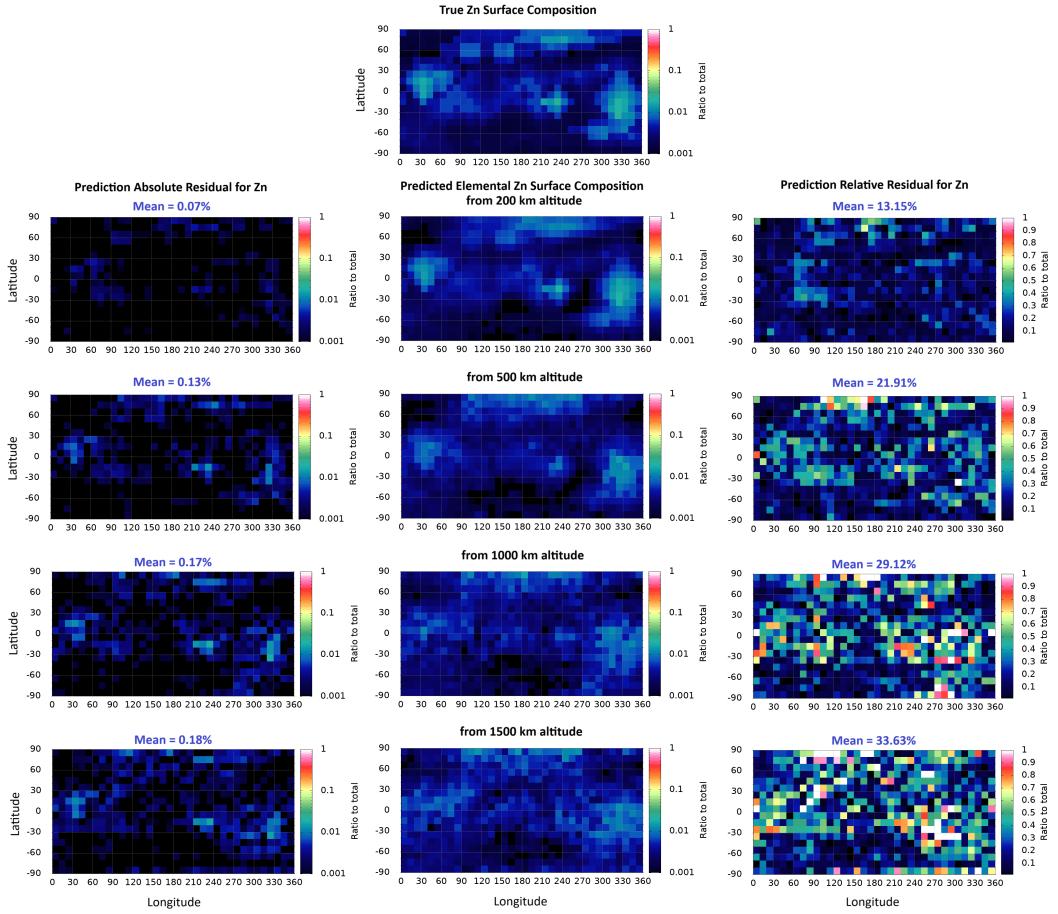


Figure C7. Main test campaign - MLP DNN map reconstructions of the same sample Zinc surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground truth" surface composition. The maps in the middle below it are the predicted fractions for this element. The panels on the left show the absolute residuals to the "ground truth", while on the right are the relative residuals. Notably, the Zinc reconstructed maps correctly follow the Sulfur ones due to the matching source "mineral" for the two elements in our setting.

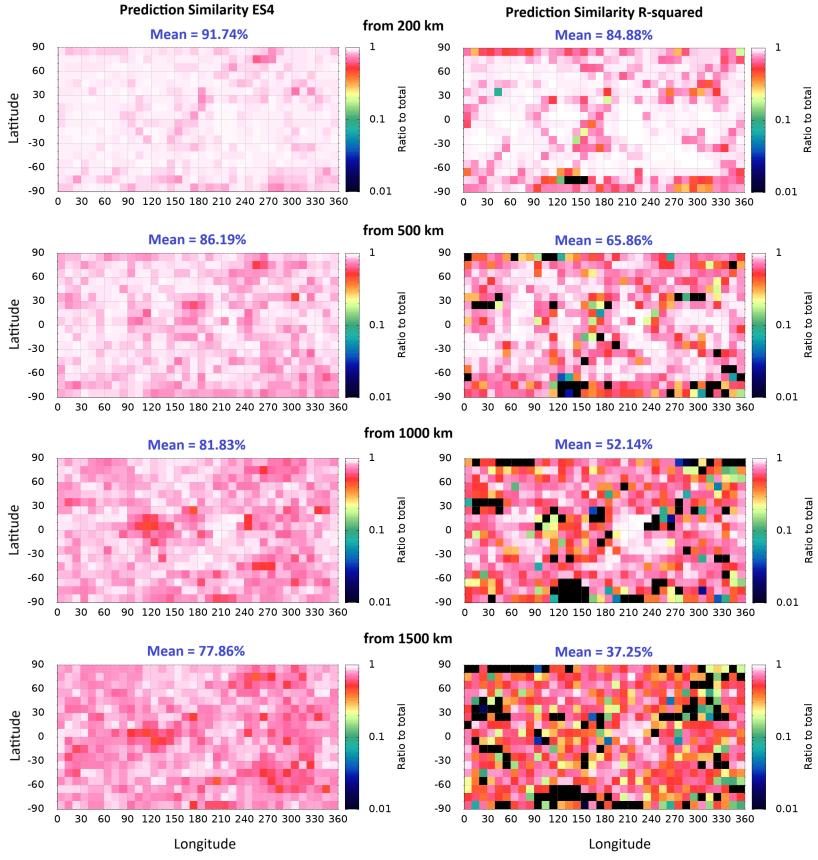


Figure C8. Map representations of the accuracy metrics evaluated for the predictions by the finalized trained MLP DNN of the same sample total surface composition (test set number 2). Dayside only inputs of two simulated exospheres from two consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The panels on the left show the ES4 similarity, while on the right are the R-squared metrics.

1220 **Open Research Section**

1221 The data used for the training and testing the DNN of the main testing campaign
 1222 (both inputs and outputs) are available at Zenodo via doi 10.5281/zenodo.13778487 with
 1223 Creative Commons Attribution license (Kazakov, 2024b). The inputs and outputs data
 1224 to the exosphere generation model are available at Zenodo via doi 10.5281/zenodo.13780529
 1225 with Creative Commons Attribution license (Kazakov, 2024a). The data for all the var-
 1226 ious surfaces used to produce the exospheres and to train and test the algorithm are avail-
 1227 able at Zenodo via doi 10.5281/zenodo.13780740 with Creative Commons Attribution
 1228 license (Kazakov, 2024d).

1229 Version 0.1.0 of the PTF-A-MLP software used to train and test the deep neural
 1230 network is available at GitHub-Zenodo via doi 10.5281/zenodo.13785015 with Creative
 1231 Commons Attribution license (Kazakov, 2024c) and developed openly at [https://github.com/AJKazakov/PTF-](https://github.com/AJKazakov/PTF-A-MLP)
 1232 A-MLP.

1233 The exospheric model used to produce the exosphere is available at: [http://150.146.134.250/cgi-](http://150.146.134.250/cgi-bin/modello-input.pl)
 1234 bin/modello-input.pl.

1235 **Acknowledgments**

1236 This work was supported by the Italian Space Agency (ASI) - SERENA contract no. 2018-
 1237 8-HH.1-2022 "Scientific participation in the mission BepiColombo SERENA - Phase E1".

1238 **References**

- 1239 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... others
 1240 (2015). Tensorflow: Large-scale machine learning on heterogeneous systems.
 1241 *Software available from tensorflow.org*.
- 1242 Bengio, Y. (2015). *Early inference in energy-based models approximates back-*
 1243 *propagation* (Technical Report No. arXiv:1510.02777). Université de Montréal.
- 1244 Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M.,
 1245 ... Ziethe, R. (2010). Bepicolombo - comprehensive exploration of mercury:
 1246 Mission overview and science goals. *Planetary and Space Science*, 58(1), 2-20.
- 1247 Berezhnoy, A. (2018). Chemistry of impact events on mercury. *Icarus*, 300, 210-222.
 1248 doi: 10.1016/j.icarus.2017.08.034
- 1249 Bergstra, J., Bardenet, R., Bengio, Y., & Balazs, K. (2011). Algorithms for hyper-
 1250 parameter optimization. In *Proceedings of the 24th international conference on*
 1251 *neural information processing systems* (p. 2546-2554).
- 1252 Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- 1253 Cassidy, T. A., Merkel, A. W., Burger, M. H., Sarantos, M., Killen, R. M., Mc-
 1254 Clintock, W. E., & Vervack, R. J. (2015). Mercury's seasonal sodium
 1255 exosphere: Messenger orbital observations. *Icarus*, 248, 547-559. Re-
 1256 trieval from <https://doi.org/10.1016/j.icarus.2014.10.037> doi:
 1257 10.1016/j.icarus.2014.10.037
- 1258 Chollet, F., et al. (2015). Keras. *Github. GitHub repository*.
- 1259 Cintala, M. J. (1992). Impact-induced thermal effects in the lunar and mercurian re-
 1260 goliths. *J. Geophys. Res.*, 97, 947-973.
- 1261 Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Deep big
 1262 simple neural nets for handwritten digit recognition. *Neural Computation*, 22,
 1263 1-14.
- 1264 Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Wiley-
 1265 Interscience.
- 1266 Domingue, D. L., Koehn, P. L., Killen, R. M., et al. (2007). Mercury's atmosphere:
 1267 A surface-bounded exosphere. *Space Science Reviews*, 131(1-4), 161-186.
- 1268 E. Vander Kaaden, K., M. McCubbin, F., R. Nittler, L., N. Peplowski, P., Z. Wei-

- 1269 der, S., A. Frank, E., & J. McCoy, T. (2017). Geochemistry, mineralogy, and
 1270 petrology of boninitic and komatiitic rocks on the mercurian surface: Insights
 1271 into the mercurian mantle. *Icarus*, 285, 155-168. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0019103516303116> doi:
 1272 <https://doi.org/10.1016/j.icarus.2016.11.041>
- 1273 Gamborino, D., Vorburger, A., & Wurz, P. (2019). Mercury's subsolar sodium ex-
 1274 osphere: An ab initio calculation to interpret mascs/uvvs observations from
 1275 messenger. *Annales Geophysicae*, 37, 455-470.
- 1276 Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks.
 1277 In *Proceedings of the fourteenth international conference on artificial intelli-
 1278 gence and statistics* (Vol. 15, p. 315-323).
- 1279 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA:
 1280 MIT Press.
- 1281 Grava, C., Killen, R. M., Benna, M., Berezhnoy, A. A., Halekas, J. S., Leblanc, F.,
 1282 ... Vorburger, A. (2021). Volatiles and refractories in surface-bounded ex-
 1283 ospheres in the inner solar system. *Space Science Reviews*, 217(5), 61. doi:
 1284 [10.1007/s11214-021-00833-8](https://doi.org/10.1007/s11214-021-00833-8)
- 1285 Head, T., et al. (2018). *scikit-optimize/scikit-optimize: v0.5.2*. Zenodo. Retrieved
 1286 from <http://doi.org/10.5281/zenodo.1207017>
- 1287 Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive
 1288 Sciences*, 11(10), 428-434.
- 1289 Joachims, T. (2002). *Learning to classify text using support vector machines: Meth-
 1290 ods, theory and algorithms*. Norwell, MA, USA: Kluwer Academic Publishers.
- 1291 Johnson, R., Leblanc, F., Yakshinskiy, B., & Madey, T. (2002). Energy distri-
 1292 butions for desorption of sodium and potassium from ice: the na/k ratio at
 1293 europa. *Icarus*, 156, 136-142.
- 1294 Kazakov, A. (2024a, September). *Inputs and outputs for exospheric simulations
 1295 used in the deep neural network for surface reconstruction from simulated exo-
 1296 spheric measurements*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.13780529> ([Dataset]) doi: 10.5281/zenodo.13780529
- 1297 Kazakov, A. (2024b, September). *Inputs and outputs for the training and test-
 1298 ing of a deep neural network for surface reconstruction from simulated exo-
 1299 spheric measurements*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.13778487> ([Dataset]) doi: 10.5281/zenodo.13778487
- 1300 Kazakov, A. (2024c, September). *PTF-A-MLP*. Retrieved from <https://doi.org/10.5281/zenodo.13785015> ([Software] Version 0.1.0) doi: 10.5281/zenodo.13785015
- 1301 Kazakov, A. (2024d, September). *Surfaces/regoliths used in the training and testing
 1302 of the deep neural network for surface reconstruction from simulated exo-
 1303 spheric measurements*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.13780740> ([Dataset]) doi: 10.5281/zenodo.13780740
- 1304 Kazakov, A., et al. (2020). Deep neural networks for analysis of mercury's plan-
 1305 etary exosphere. *Journal of Physics: Conference Series*, 1548, 012-014. doi: 10
 1306 .1088/1742-6596/1548/1/012014
- 1307 Killen, R. (2016). Pathways for energization of ca in mercury's exosphere. *Icarus*,
 1308 268, 32-36. doi: 10.1016/j.icarus.2015.12.035
- 1309 Killen, R. M., & Burger, M. H. (2019). Understanding mercury's exosphere: Mod-
 1310 els derived from messenger observations. In *Mercury: The view after messen-
 1311 ger*. London: Cambridge University Press.
- 1312 Killen, R. M., Cremonese, G., Lammer, H., et al. (2007). Processes that promote
 1313 and deplete the exosphere of mercury. *Space Science Reviews*, 132(2-4), 433-
 1314 509.
- 1315 Killen, R. M., Potter, A. E., Reiff, P., Sarantos, M., Jackson, B. V., Hick, P., &
 1316 Giles, B. (2001). Evidence for space weather at mercury. *J. Geophys. Res.*,
 1317 106, 20509-20525.
- 1318

- Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. In *International conference on learning representations*.
- Leblanc, F., Sarantos, M., Domingue, D., Milillo, A., Savin, D., Prem, P., ... Raines, J. (2023, 12). How does the thermal environment affect the exosphere/surface interface at mercury? *The Planetary Science Journal*, 4, 227. doi: 10.3847/PSJ/ad07da
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444. Retrieved from <https://doi.org/10.1038/nature14539> doi: 10.1038/nature14539
- Mangano, V., et al. (2015). Themis na exosphere observations of mercury and their correlation with in-situ magnetic field measurements by messenger. *Planetary and Space Science*, 115, 102-109.
- Massetti, S., Orsini, S., Milillo, A., Mura, A., De Angelis, E., Lammer, H., & Wurz, P. (2003). Mapping of the cusp plasma precipitation on the surface of mercury. *Icarus*, 166, 229–237.
- Merkel, A. W., Vervack, R. J. J., Killen, R. M., Cassidy, T. A., McClintonck, W. E., Nittler, L. R., & Burger, M. H. (2018). Evidence connecting mercury's magnesium exosphere to its magnesium-rich surface terrane. *Geophysical Research Letters*, 45, 6709-6797. doi: 10.1029/2018GL078407
- Milillo, A., Fujimoto, M., Kallio, E., Kameda, S., Leblanc, F., Narita, Y., ... 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
- Milillo, A., Fujimoto, M., Murakami, G., et al. (2020). Investigating mercury's environment with the two-spacecraft bepicolombo mission. *Space Science Reviews*, 216. Retrieved from <https://doi.org/10.1007/s11214-020-00712-8> (Article no. 93) doi: 10.1007/s11214-020-00712-8
- Milillo, A., Sarantos, M., Grava, C., Janches, D., Lammer, H., Leblanc, F., ... Murakami, G. (2023, 09). Future directions for the investigation of surface-bounded exospheres in the inner solar system. *Space Science Reviews*, 219. doi: 10.1007/s11214-023-00994-8
- Milillo, A., & Wurz, P. (2014). *Serena science performance report*.
- Milillo, A., Wurz, P., Orsini, S., Delcourt, D., Kallio, E., Killen, R., ... Torkar, K. (2005). Surface-exosphere-magnetosphere system of mercury. *Space Science Reviews*, 117, 397-443. doi: 10.1007/s11214-005-3593-z
- Minsky, M., & Papert, S. A. (2017). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
- Moroni, M., Mura, A., Milillo, A., Plainaki, C., Mangano, V., Alberti, T., ... Soridini, R. (2023). Micro-meteoroids impact vaporization as source for ca and cao exosphere along mercury's orbit. *Icarus*, 401, 115616.
- 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.
- Mura, A., Orsini, S., Milillo, A., Delcourt, D., Massetti, S., & De Angelis, E. (2005). Dayside h+ circulation at mercury and neutral particle emission. *Icarus*, 175, 305-319.
- Mura, A., Wurz, P., Lichtenegger, H., Schleicher, H., Lammer, H., Delcourt, D., ... 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
- Orsini, S., Livi, S., Torkar, K., Barabash, S., Milillo, A., Wurz, P., ... Kallio, E. (2010). Serena: A suite of four instruments (elenia, strofio, picam and mipa) on board bepicolombo-mpo for particle detection in the hermean environment. *Planetary and Space Science*, 58, 166-181.
- Orsini, S., Livi, S. A., Lichtenegger, H., Barabash, S., Milillo, A., De Angelis, E., ...

- 1379 Zampieri, S. (2021, January). Serena: Particle instrument suite for determining
 1380 the sun-mercury interaction from bepicolombo. *Space Science Reviews*,
 1381 217(1), 11. Retrieved from <https://doi.org/10.1007/s11214-020-00787-3>
 1382 doi: 10.1007/s11214-020-00787-3
- 1383 Plainaki, C., Mura, A., Milillo, A., Orsini, S., Livi, S., Mangano, V., ... De Angelis,
 1384 E. (2017). Investigation of the possible effects of comet encke's meteoroid
 1385 stream on the ca exosphere of mercury. *Journal of Geophysical Research: Planets*, 122, 1217-1226. doi: 10.1002/2017JE005304
- 1387 Pokorný, P., Sarantos, M., & Janches, D. (2017). Reconciling the dawn-dusk asymmetry in mercury's exosphere with the micrometeoroid impact directionality.
 1388 *The Astrophysical Journal Letters*, 842(2), L17.
- 1389 Pokorný, P., Sarantos, M., & Janches, D. (2018). A comprehensive model of the
 1390 meteoroid environment around mercury. *The Astrophysical Journal*, 863, 31.
 1391 (31pp, Received 2018 April 3; revised 2018 June 26; accepted 2018 June 27;
 1392 published 2018 August 8) doi: 10.3847/1538-4357/aad051
- 1394 Rothery, D. A., Massironi, M., Alemanno, G., et al. (2020). Rationale for bepi-
 1395 colombo studies of mercury's surface and composition. *Space Science Reviews*,
 1396 216, 66. doi: 10.1007/s11214-020-00694-7
- 1397 Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by
 1398 back-propagating errors. *Nature*, 323, 533-536.
- 1399 Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal rep-
 1400 resentations by error propagation. In D. E. Rumelhart & J. L. McClelland
 1401 (Eds.), *Parallel distributed processing* (Vol. 1, p. 318-362). Cambridge: MIT
 1402 Press.
- 1403 Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach, 3rd edi-*
 1404 *tion*. USA: Prentice Hall Press.
- 1405 Sarantos, M., Slavin, J., Benna, M., Boardsen, S., Killen, R., Schriver, D., &
 1406 Trávnícek, P. (2009). Sodium-ion pickup observed above the magnetopause
 1407 during messenger's first mercury flyby: Constraints on neutral exospheric
 1408 models. *Geophysical Research Letters*, 36. doi: 10.1029/2008GL036207
- 1409 Wilson III, L. B., Stevens, M. L., Kasper, J. C., Klein, K. G., Maruca, B. A., Bale,
 1410 S. D., ... Salem, C. S. (2018). The statistical properties of solar wind tem-
 1411 perature parameters near 1 au. *The Astrophysical Journal Supplement Series*,
 1412 236(2), 41. doi: 10.3847/1538-4365/aab71c
- 1413 Wurz, P., & Lammer, H. (2003). Monte-carlo simulation of mercury's exosphere.
 1414 *Icarus*, 164, 1-13.
- 1415 Wurz, P., Whitby, J. A., Rohner, U., Martin-Fernandez, J. A., Lammer, H., & Kolb,
 1416 C. (2010). Self-consistent modelling of mercury's exosphere by sputtering,
 1417 micrometeorite impact and photon-stimulated desorption. *Planetary and Space
 1418 Science*, 58, 1599-1616.