

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

⁴ Adrian Kazakov^a, Anna Milillo^a, Alessandro Mura^a, Stavro
⁵ Ivanovski^b, Valeria Mangano^a, Alessandro Aronica^a, Elisabetta De
⁶ Angelis^a, Pier Paolo Di Bartolomeo^a, Luca Colasanti^a, Miguel
⁷ Escalona-Morán^c, Francesco Lazzarotto^d, Stefano Massetti^a, Martina
⁸ Moroni^a, Raffaella Noschese^a, Fabrizio Nuccilli^a, Stefano Orsini^a,
⁹ Christina Plainaki^e, Rosanna Rispoli^a, Roberto Sordini^a, Mirko
¹⁰ Stumpo^a, Nello Vertolli^a

¹¹ ^a INAF-IAPS, via Fosso del Cavaliere 100, 00133, Rome, Italy

¹² ^b INAF-Osservatorio Astronomico di Trieste, via Giambattista Tiepolo 11, 34143, Trieste, Italy

¹³ ^c Augmented Intelligence Lab, Rua Lugo 2, 36470, Salceda de Caselas, Spain

¹⁴ ^d INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122, Padova, Italy

¹⁵ ^e ASI - Italian Space Agency, Via del Politecnico, 00133, Rome, Italy

¹⁶ E-mail: adrian.kazakov@inaf.it

¹⁷ **Abstract.** The surface information derived from exospheric measurements at planetary bodies
¹⁸ complements the surface mapping provided by dedicated imagers. Indeed, these measurements
¹⁹ offer critical insights into surface release processes, dynamics of various interactions within
²⁰ the planetary environment, and the effects of erosion, space weathering, and, ultimately, the
²¹ planet's evolution. This study explores a tentative proxy method for deriving the elemental
²² composition of Mercury's regolith from in-situ measurements of its neutral exosphere using
²³ deep neural networks (DNNs). We present a supervised feed-forward DNN architecture—a
²⁴ network of fully-connected neural layers, the so-called multilayer perceptron (MLP). This
²⁵ network takes exospheric densities and proton precipitation fluxes, derived from a simulated

orbital run through Mercury's exosphere, as inputs and predicts the chemical elements of the surface regolith below. It serves as an estimator for the surface-exosphere interaction and the processes leading to exosphere formation, which, in our simulated setup, include micrometeoroid impact vaporization, ion sputtering, photon-stimulated desorption, and thermal desorption. Extensive training and testing campaigns demonstrate the MLP DNN's ability to accurately predict and reconstruct surface composition maps from simulated exospheric measurements. These results not only affirm the algorithm's robustness but also illuminate its extensive capabilities in handling complex data sets for the creation of estimators for modeled exospheric generation. Furthermore, the tests reveal substantial potential for further development, suggesting that this method could significantly enhance the analysis of complex surface-exosphere interactions and reduce uncertainties in models that generate planetary exospheres. This work anticipates the analysis of data from the SERENA (Search for Exospheric Refilling and Emitted Natural Abundances) instrument package aboard the Mercury Planetary Orbiter, part of the BepiColombo space mission to Mercury, with its nominal phase starting in 2026.

Keywords: Mercury, Exosphere, Surface composition, Deep neural networks

1. Introduction

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

The active processes in the formation of Mercury's tenuous atmosphere are widely discussed in the literature (Mura et al., 2007; Wurz et al., 2010; Killen and Burger, 2019; Grava et al., 2021; Milillo et al., 2023). Four predominant processes release atoms and molecules from the surface into the exosphere: micrometeoroid impact vaporization (MIV), sputtering after solar wind and heavy ion impacts (SP), thermal desorption (TD), and photon-stimulated desorption

(PSD). MIV and SP are particularly indicative of the regolith composition below, as they involve higher energy transfers capable of dislodging neutral species from their minerals. Conversely, TD and PSD, being less energetic, tend to release atoms and molecules that are weakly bonded to minerals, such as volatile elements, most of which eventually fall back and are reabsorbed by the surface (Killen et al., 2007; Mura et al., 2009; Gamborino et al., 2019; Leblanc et al., 2023). Once in the exosphere, the released particles undergo further transformations due to interactions with radiation pressure, photons, and charged particles. Such interactions can modify the charge, chemical state, and movement of these exospheric constituents. However, in a first approximation in the sparse exosphere, the atomic and molecular abundances resulting from these actors could be traced back to the planet, connecting the surface properties, like composition, mineralogy, and physical state to the different processes and the dynamics of matter around the planet (Milillo et al., 2020; Rothery et al., 2020). There has already been direct evidence that this is the case for the Magnesium exosphere, which is directly related to the Magnesium-rich surface below, as shown by Merkel et al. (2018).

To gain a deeper understanding of Mercury's exosphere, scientists use sophisticated models to simulate the active processes and their effects on the planetary surface, thereby attempting to replicate the generation of the exosphere. This extensively applied method compares the results of simulations to those measured from space (e.g. Sarantos et al. (2009); Cassidy et al. (2015); Plainaki et al. (2017)) or from Earth (e.g. Wurz and Lammer (2003); Mura et al. (2009); Mangano et al. (2015)). However, the inherent complexity of these interactions, which includes electromagnetic, chemical, mechanical, thermal, and other effects, and the validity of the chosen parameters in their mathematical representation add significant challenges. Some of these effects have not been precisely evaluated for each release process, leading to a broad range of simulated results with considerable uncertainty, depending on the assumptions made at the outset. To address these, a multifaceted approach is required, involving the refinement of models through improved measurements, continual reassessment of the model structure, and advanced statistical methods to better understand and quantify uncertainties.

In parallel, machine learning algorithms, particularly deep neural networks (DNNs), offer a novel approach to capture the relationships between the variables. These algorithms can resolve, to some extent, the data generation mechanisms (Russell and Norvig, 2009; LeCun et al., 2015; Goodfellow et al., 2016), providing a tool to explore in depth the relationships

88 between the components of Mercury's environment. This work will demonstrate how DNNs,
89 especially multilayer perceptrons (MLPs), can be employed within the data analysis of Mercury's
90 exosphere to reconstruct the elemental surface map underneath. Suitable for nonlinear regression
91 tasks, DNNs scale effectively with increasing training data and input parameters (Minsky and
92 Papert, 2017; Hinton, 2007; Ciresan et al., 2010), offering a promising direction for tackling the
93 complexities inherent in modeling Mercury's exosphere.

94 This study builds upon the preliminary work of Kazakov et al. (2020) by extending, refining,
95 and further advancing the application of deep neural networks in predicting Mercury's surface
96 composition from exospheric measurements. It involves predicting the elemental composition of
97 the surface using data from more sophisticated and realistic models that simulate all major
98 processes—MIV, SP, TD, and PSD—and consider the influence of solar radiation pressure
99 and photolysis on the exosphere. The multilayer perceptrons have been extensively optimized
100 through a comprehensive exploration of their building blocks, resulting in the development of
101 a robust predictive algorithm. The methodological advancements showcased in this research
102 pave the way for improved predictive accuracy and reliability in the DNNs employed in this
103 application.

104 In **Section 2**, we introduce and detail the algorithm - the multilayer perceptron deep
105 neural network - outlining its structure for the multivariate regression task of predicting surface
106 composition. This section methodically breaks down each component of the algorithm, providing
107 a comprehensive guide for constructing effective neural network architectures. **Section 3** delves
108 into the mechanisms behind exospheric data generation, encompassing the models of Mercury's
109 surface, its environment, and the processes generating the exosphere. It also elaborates on the
110 creation of the datasets used in the algorithms, including feature selection and data augmentation
111 for the DNN inputs. The findings from an extensive training and testing campaign are explored
112 in **Section 4**. This section details the selection of the architectures' hyperparameters, offering
113 insights into the optimal choices within the DNN hyperparameter space to develop an effective
114 MLP DNN. The testing of the algorithms is presented, showcasing their performance on a variety
115 of surface-exosphere pairs and culminating in the visual demonstration of reconstructed surface
116 elemental composition maps. The paper concludes in **Section 5**, presenting a discussion and
117 future perspectives. This final section outlines the method's potential and the wide scope for
118 further research and development in this field.

¹¹⁹ **2. Method**

¹²⁰ *2.1. Prediction Task and General Characteristics of the Method*

¹²¹ In this study, we develop a machine learning algorithm and apply it to supervised multivariate
¹²² regression of exospheric data at Mercury using a multilayer perceptron deep neural network. The
¹²³ objective of this DNN is to infer the regolith source material, believed to be a primary contributor
¹²⁴ to Mercury's exosphere. Specifically, the DNN predicts elemental surface composition fractions
¹²⁵ from exospheric density measurements, governed by the equation:

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

¹²⁶ where \hat{y}_i is the fraction of an elemental species predicted by the neural network to be present
¹²⁷ in the surface area below the exospheric measurement, and n is the total number of elements in
¹²⁸ the prediction task (Figure 1).

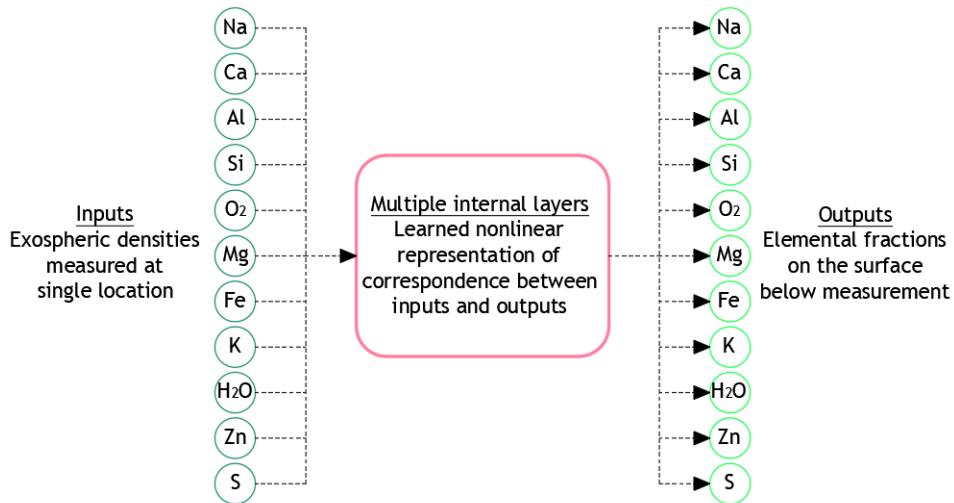


Figure 1. DNN prediction task schematics. The input to the neural network on the left are the exospheric densities at a single location in the exosphere. The output is the relative surface elemental composition as fractions summing up to 1 at a surface area just below the exospheric measurement. The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

¹²⁹ The methodological and algorithmic developments in this study include:

- ¹³⁰ • Devising the algorithm's application for predicting surface elemental composition.

- 131 • Building DNNs capable of operating in a multi-process environment, integrating the four
132 primary active processes (MIV, SP, TD, PSD) for neutral species release.
- 133 • Implementing a data production model with plausible assumptions for the exosphere
134 generation processes.
- 135 • Training the algorithms with augmented datasets.
- 136 • Employing hyperparameter tuning to optimize the DNN design parameters.
- 137 • Investigating the formation of the physical processes parameter space.

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

141 2.2. Deep Neural Network Architecture

142 The MLP, a class of feedforward neural network, excels in multivariate regression by modeling
143 complex nonlinear functions with its multi-layered, fully connected structure and nonlinear
144 activation functions (Minsky and Papert, 2017; Rumelhart et al., 1986a; Kingma and Ba,
145 2014; LeCun et al., 2015). This architecture, combined with optimization techniques like
146 backpropagation, allows MLPs to identify intricate patterns in high-dimensional data, making
147 them ideal for robust predictive modeling.

148 Training involves preprocessing data for network suitability, building the model using the
149 Keras framework with TensorFlow (Abadi et al., 2015; Chollet et al., 2015), and iteratively
150 tuning the network through backpropagation to minimize error (Rumelhart et al., 1986b). This
151 process ensures effective and reproducible model performance.

152 The architecture of a multilayer perceptron enables complex data processing through
153 a structured network of layers: an input layer, multiple hidden layers for nonlinear
154 transformations, and an output layer for predictions. The network's effectiveness hinges
155 on key components like the loss function, which guides accuracy improvements, and the
156 regularizer, which ensures generalizability. Efficiently chosen optimization algorithms and precise
157 hyperparameter tuning further enhance the network's performance. Figure 2 illustrates this
158 interplay, crucial for tasks like analyzing Mercury's exosphere, with subsequent sections detailing
159 each component's role in predictive capabilities. The inner connectiveness of the MLP DNN
160 neural units is shown on Figure 3.

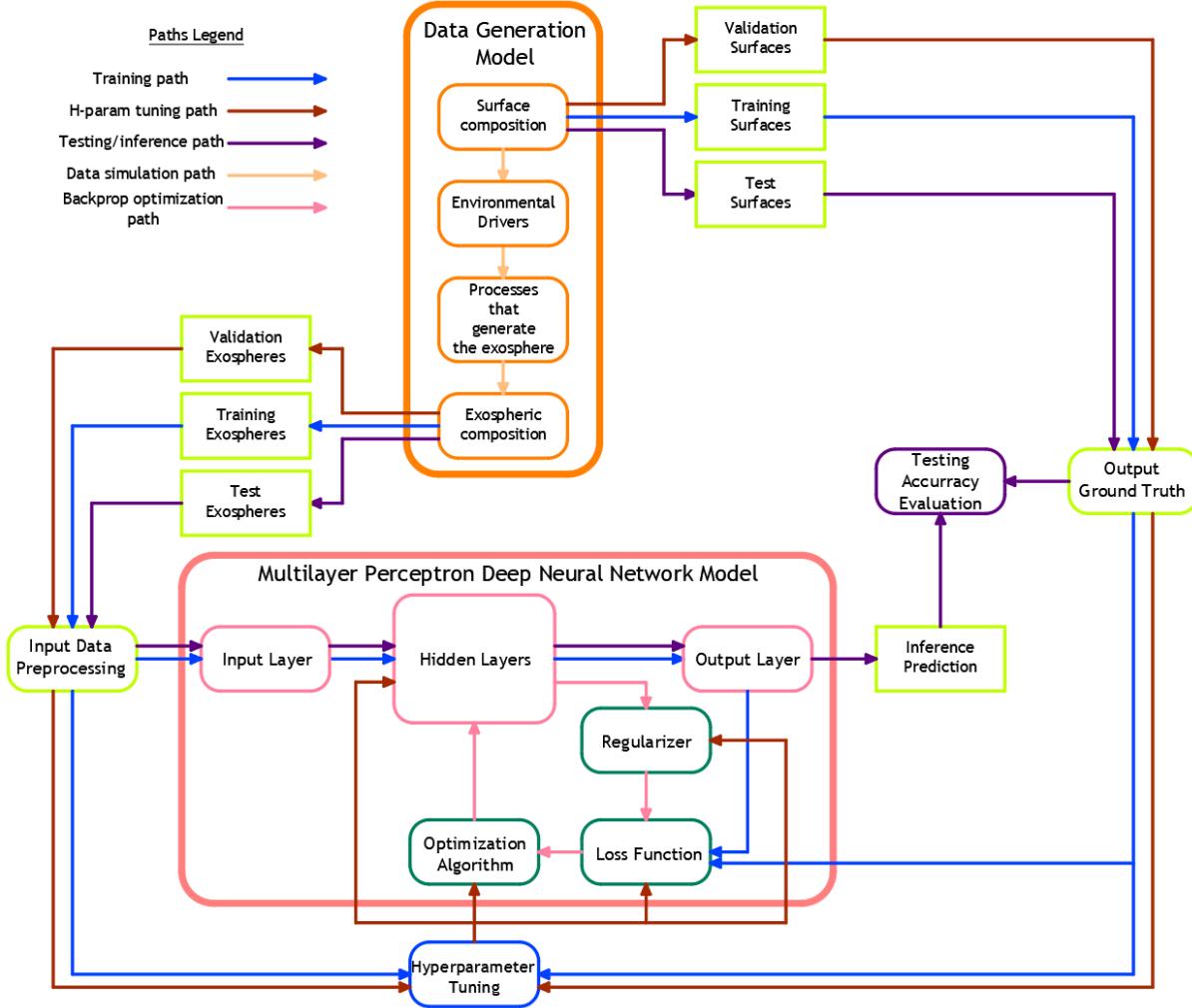


Figure 2. MLP DNN architecture overview. The data generation model produces both the inputs and the outputs for training, validating, and testing the algorithm. This data is passed through the MLP DNN model in the training, hyperparameter tuning and testing phases, respectively. The backpropagation optimization 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.

161 *Input Layer*

162 The input layer of the MLP introduces data, in our case Mercury's exospheric density
 163 measurements, into the network, with each neuron representing a distinct data feature. For
 164 example, distinct elemental density measurements are represented by separate neurons. Before
 165 entry, data undergo normalization to ensure uniform influence on the learning process, thereby

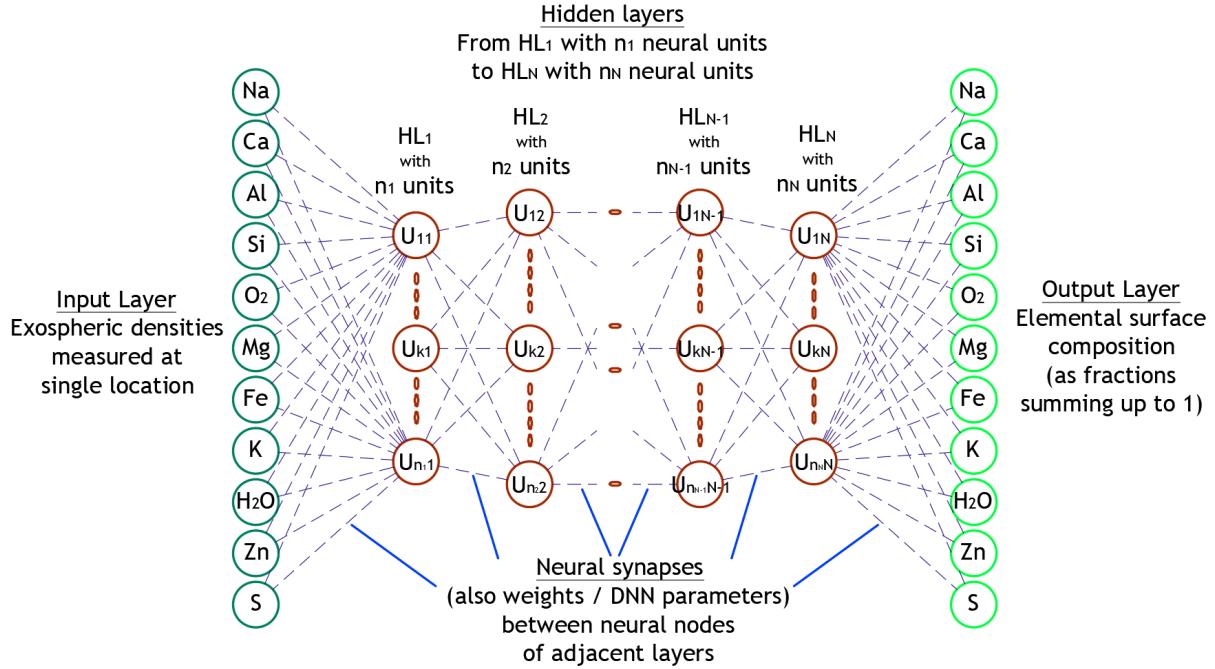


Figure 3. MLP DNN basic structure schematics. Exospheric densities form the input layer on the left. The output layer is formed from the relative surface elemental composition as fractions summing up to 1. There are N number of hidden layers with n_1 to n_N number of neural units (neurons). This structure 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} .

¹⁶⁶ preventing bias (Goodfellow et al., 2016). This involves standardizing each feature to zero mean
¹⁶⁷ and unit variance as per the equation:

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

¹⁶⁸ where \mathbf{x} is the standardized vector of input features, \mathbf{x}_{orig} is the original vector of input
¹⁶⁹ features, $\boldsymbol{\mu}$ is the vector of means of the feature values, and $\boldsymbol{\sigma}$ is the standard deviations vector.
¹⁷⁰ Such standardization enhances the efficiency and stability of the network's learning process.

¹⁷¹ Hidden Layers

¹⁷² The hidden layers form the core of the MLP architecture, where the actual processing and
¹⁷³ learning occur (Minsky and Papert, 2017; Hinton, 2007). Positioned between the input and
¹⁷⁴ output layers, they transform input data into a form usable for predictions. Each hidden layer is
¹⁷⁵ 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 by the application of the ReLU activation function, enabling the network to learn and represent complex nonlinear relationships.

Finally, the output from the hidden layers is passed on to the output layer, where the final prediction is made. The architecture and depth of the hidden layers are critical and typically determined through empirical methods and hyperparameter tuning. This ensures the network has the requisite complexity for effective learning while avoiding overfitting to the empirical distribution present in the training data.

198 Output Layer

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

204 The activation function used in the output layer is crucial and depends on the nature of
 205 the prediction task. In our case, where the output is a set of continuous values that sum
 206 to 1 (representing fractions), the softmax function is used (Joachims, 2002). The softmax
 207 function converts the raw output of the network into a probability distribution, ensuring that
 208 the predicted fractions are non-negative and sum up to one, aligning perfectly with the physical
 209 reality of our task.

210 The formula for the softmax function is as follows:

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

211 where \hat{y}_i is the predicted fraction for the i -th element, and z represents the raw output
 212 values from the final hidden layer. This configuration allows the network to deliver accurate,
 213 meaningful predictions of Mercury's surface composition, synthesizing the representation insights
 214 gained from all previous layers.

215 *Loss Function*

216 The loss function plays a pivotal role in guiding the optimization process, quantifying
 217 the discrepancy between the network's predictions and the actual target values to gauge
 218 model accuracy. In our most successful MLP tests on predicting Mercury's surface elemental
 219 composition, the Kullback-Leibler (KL) divergence (Cover and Thomas, 2006) has proven
 220 particularly effective. It measures how one probability distribution, representing the predicted
 221 elemental composition (the output from the MLP), diverges from the actual distribution (the
 222 true elemental composition). The formula for KL divergence is:

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

223 where P represents the true distribution of the fraction of element i in the data, and Q is the
 224 predicted distribution from the MLP. Other loss functions like Mean Absolute Error (MAE) and
 225 Mean Squared Error (MSE) were also considered. However, the KL divergence was preferred
 226 for our regression task because it aligns better with the probabilistic requirements, focusing on
 227 relative proportions rather than absolute quantities of elements.

228 *Regularizer*

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

234 The L2 regularization is mathematically represented as:

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

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

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

246 Optimization (Learning) Algorithm

247 The training of our multilayer perceptron for predicting Mercury’s surface composition
 248 employs the Adam optimization algorithm, a refinement of stochastic gradient descent known
 249 for its effectiveness with large-scale data and complex models (Kingma and Ba, 2014). The
 250 fundamental mechanism of Adam involves dynamically and adaptively updating the weights of
 251 the combination matrices for each hidden layer to minimize the total error as indicated by the loss
 252 function. This is achieved through backpropagation optimization (Rumelhart et al., 1986a,b),
 253 where the weights are adjusted following their gradients with respect to the loss function:

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

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

258 *Hyperparameter Tuning*

259 Key hyperparameters of our MLP DNN are:

- 260 • **Learning Rate:** This parameter governs the size of the steps taken during the
261 backpropagation optimization algorithm along the weight gradients of the loss function.
262 A well-balanced learning rate is critical—it must be large enough to navigate plateaus in
263 the loss function’s parameter space, yet sufficiently small to converge to (or remain near)
264 the minimum of the error.
- 265 • **Mini-Batch Size:** This refers to the size of the random subset of examples used in each
266 training iteration, impacting both the speed and stability of the learning process.
- 267 • **Number of Hidden Layers and Neurons:** These parameters determine the depth and
268 width of the neural network, influencing its ability to model complex relationships in the
269 data.
- 270 • **L2 Regularization Coefficient:** This defines the degree of penalty imposed on large
271 weight values, helping to prevent overfitting by controlling model complexity.

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

279 *Performance Metrics*

280 To evaluate the performance of our machine learning model, we utilize both customized
281 and standard metrics to ensure precise and insightful quantitative assessments. Our primary
282 metric, the Euclidean similarity 4 (ES4), integrates elements of Euclidean distance and cosine

283 similarity, providing a nuanced measure of prediction accuracy by considering both magnitude
 284 and directionality in multidimensional space:

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

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

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

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

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

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

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

303 3. Data Generation Model

304 For the development of our MLP DNN, we employ a data generation model consisting of
 305 three components: a surface composition model, an exosphere generation model, and a model
 306 for the distributions of the drivers of surface release processes. The surface model in our study
 307 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.

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.

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, considers the full fraction of volatile species (such as Na, K, H_2O , S, and O_2) as being readily available for release into the exosphere as they are loosely bound to the regolith grains. The list of elements resulting from the mineral break down is presented in Table 2. The elemental composition resulting from this process represents the 'actual' or

356 '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 ₂ O	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.

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

362 Firstly, the surface in our model is represented as a grid comprised of 36×18 surface tiles in a
 363 modified Mercator projection. Each tile measures 10°×10°, which, at the equator, translates to
 364 approximately 425 km×425 km. This averaging of composition inevitably reduces the complexity
 365 of the parameter space, as it results in the loss of finer details in the spatial relations of the
 366 spread of the different species - elemental and mineral - on the planetary surface. However, it is
 367 important to note that the potential resolution of surface composition maps reconstructed from
 368 exospheric measurements by an orbiting spacecraft cannot be significantly higher.

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

372 3.2. Environmental Drivers

373 In our model definition, we incorporate the environmental conditions and various factors that

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

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

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

Furthermore, the dust environment around Mercury is considered for particles smaller 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 20 km/s in Mercury's vicinity, in agreement with the modal impact velocity reported by Cintala (1992). This is compared to a planet velocity at perihelion of 59 km/s. However, our model does not differentiate between the origins of these dust particles—whether they come from the Main Belt Asteroids, Jupiter Family Comets, Oort Cloud Comets, or Halley Type Comets—nor does it consider the full ranges and exact distributions of particle sizes and velocities as in Pokorný et al. (2018). Additionally, no large meteorite impacts or increases of fluxes due to particularly dense cometary streams, such as from comet Encke (Plainaki et al., 2017), are considered. Grain size distribution influence of the dust particles is also not represented in our physical parameter space.

The environmental conditions on the planet itself present a diverse range of parameters due to varying exposure to sunlight and shadow, as well as differences in particle fluxes on the planet's leading and trailing sides due to its high orbital velocity. Our model incorporates detailed

maps that illustrate solar incidence angles and planetary velocity incidence angles at Mercury's perihelion (Figure 4). It's crucial to recognize Mercury's unique orbit-spin resonance, which alternates the sides facing the Sun at the same true anomaly angle in successive orbits, a fact that we have taken advantage of later in our study.

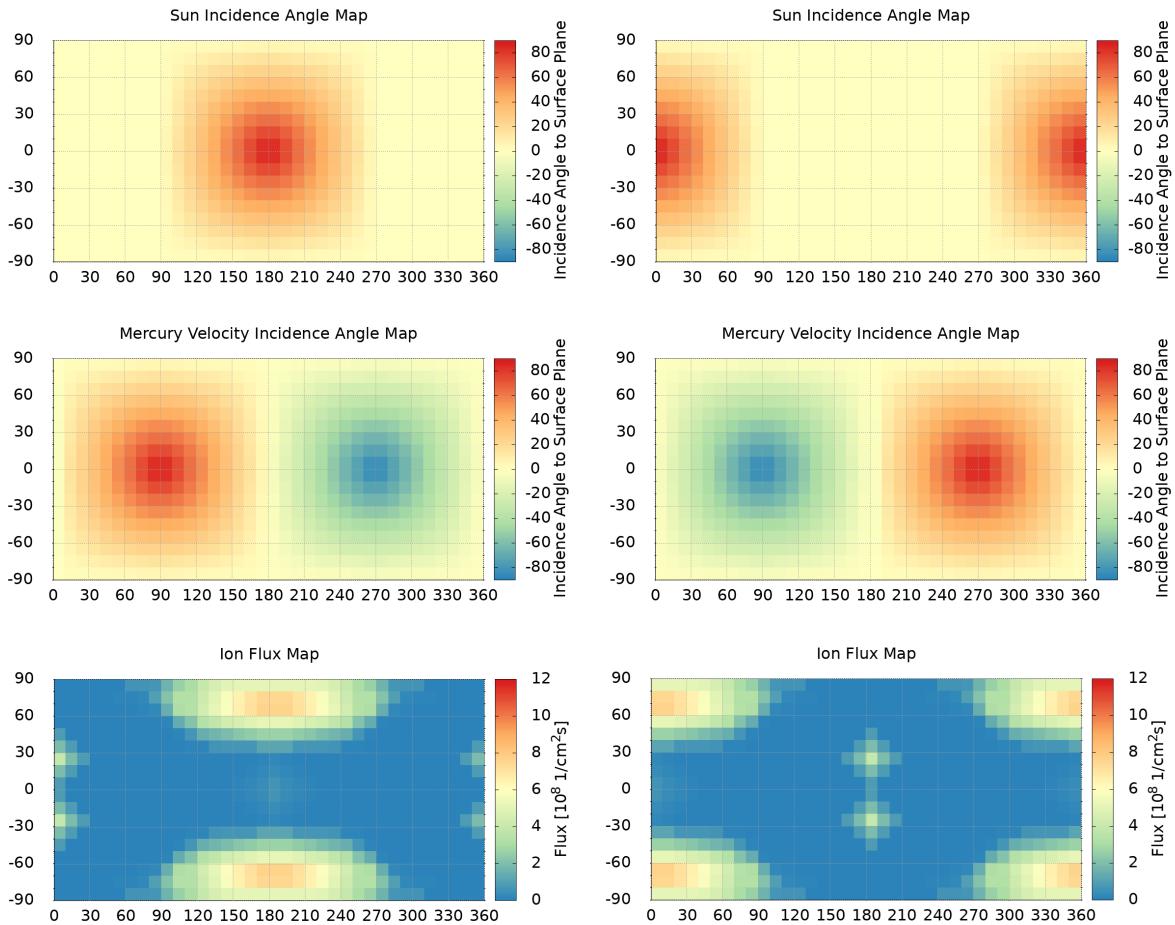


Figure 4. Maps of incidence angles due to planet orientation at two consecutive perihelia. Perihelion 1 on the left and perihelion 2 on the right. The top panels represent the incidence maps of solar photons on the surface. The middle panels represent the angles between the planet's velocity and the 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

3.3 Exosphere Generation Model

3 DATA GENERATION MODEL

413 through magnetic field cusps directly onto the planet's surface at specific locations known as cusp
 414 footprints (Massetti et al., 2003). The impact of this process is significant, as it concentrates ion
 415 bombardment in particular areas, altering the surface composition and influencing the generation
 416 of the exosphere. We have delineated the shape and relative impact areas of these cusp footprints
 417 in Figure 4, highlighting the regions on Mercury's surface that are most affected by the ion
 418 funneling. These zones are another variable that is dynamically changing and in many cases in
 419 very short timeframes. In our study, we have considered a fairly wide zones of ion precipitation.
 420 A more dynamic dependence on the magnetic reconnection rate driven by the Interplanetary
 421 Magnetic Field (IMF) strength and orientation is neglected.

422 A summary of the environmental conditions considered in our study and their contribution
 423 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
	Photon flux at Earth	1/cm ² s	3.0 x 10 ¹⁵	Equivalent photon flux	PSD
Comets and Asteroids	Dust particle size	μm	0-100	Micrometeorite flux	MIV
	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
Magnetic fields	Cusp footprints size	Map		Ion precipitation zone	SP
	Cusp footprints location	Map		Ion precipitation zone	SP
	Cusp footprints ion flux distribution	Map		Ion flux	SP
Planet	Cusp footprints area coefficient	-	0.4	Ion flux	SP
	Planet velocity	km/s	59	Micrometeorite flux	MIV
	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.

424 3.3. Exosphere Generation Model

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

434 exospheres for each species.

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

444 *3.3.1. Temperature map and Thermal desorption (TD)*

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

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

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

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

452 TD is considered only as a direct thermal ejection of species from the surface i.e. sticking
 453 coefficient equal to 1. The flux of atoms or molecules resulting from TD is given by Mura et al.
 454 (2007):

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

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

458 *3.3.2. Photon flux and Photon-stimulated desorption (PSD)*

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.

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

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

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

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

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

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

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

471 The PSD process' energy distribution is modeled using a formula adapted from Johnson et al.
 472 (2002):

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

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

475 *3.3.3. Micrometeorite fluxes and impact vaporization*

476 MIV is a highly energetic surface-release process capable of releasing the entire surface
 477 material in a given volume after a micrometeoroid impact (not only the volatile species). In the
 478 Mura et al. (2007) model, the MIV exosphere is simulated starting from a map of surface release
 479 particles at a given release energy, represented as a Maxwellian distribution at 4000 K (Wurz
 480 and Lammer, 2003).

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

487 The distribution of this flux onto Mercury's surface is influenced by the velocity of Mercury
 488 and its projection onto the surface area where the flux is calculated, as illustrated in Figure
 489 4. We employ a simple relationship between the angle of incidence and the modification of the
 490 mean flux onto the surface, defined as:

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

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

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

where V_x and V_P are the volumes of the released vapor and the impactors, respectively, v is the velocity of the impactors, and c , d , and e are constants. This simplification of the parameter subspace for this complex vaporization process is deemed sufficient for our study.

This approach yields outflows of surface matter ranging from 3.5×10^{-16} to 7×10^{-16} $\text{gcm}^{-2}\text{s}^{-1}$. These values are conservatively estimated to be marginally smaller than those suggested by Cintala (1992) and two orders of magnitude smaller than those proposed by Pokorný et al. (2017), fitting within the parameter space of interest to modelers without overly emphasizing this omnipresent process. Additionally, this assumption is considered to present a more challenging scenario for the algorithm due to the inherent representability of the surface composition by the part of the exosphere generated due to the MIV.

3.3.4. Ion precipitation and Ion sputtering (SP)

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

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

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

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

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

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

where Y is the yield of the process, c the surface relative abundance of the species, Φ_I the ion flux, Φ_n the neutral flux emitted from the surface, E_i the impact energy, E_e the energy of the ejected particles, and f_S an empirical model for the energy distribution of ejected particles,

521 defined as:

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

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

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

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

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

532 3.3.5. Dynamics of the Exosphere

533 The dynamics of the exosphere, as simulated in the Mura et al. (2007) model, encompass the
 534 movement and behavior of particles after they have been released into the exosphere, including
 535 their interactions, trajectories, and eventual fate. Factors such as gravitational influences,
 536 electromagnetic forces, and collisions are examined to understand how they shape the structure
 537 and composition of the exosphere.

538 Once in the exosphere, each elemental species follows ballistic trajectories under the influence
 539 of Mercury's gravity. These particles are also subject to conditions that define their mean
 540 lifetime in the exosphere before impacting the surface or undergoing photoionization. When an
 541 exospheric particle is ionized, it is no longer simulated, making the ionization process a net loss
 542 to the exosphere.

543 Another relevant effect included in the model is solar radiation pressure, which tends push
 544 neutral elements away from the direction of incoming sunlight, effectively propelling them toward
 545 the night side of the planet. This movement is not uniform across all species; it varies depending

546 on the physical properties of the particles, such as their mass and effective cross-section, which
547 influence how much momentum they absorb or reflect from solar photons. The parameters for
548 each of the elemental species used in our exospheric model are listed in Table 4.

549 *3.4. Generation of the Datasets*

550 We conduct separate simulation runs for each of the four primary surface release processes,
551 reproduced for each distinct species. Following these individual simulations, we aggregate
552 the outcomes to compose the overall exosphere. This process essentially involves summing
553 the resulting individual exospheres generated for each species, without considering interactions
554 between the various processes. For instance, we do not account for potential competition among
555 processes for a finite pool of particles at the surface. Similarly, the exospheres for different
556 species are treated as non-interacting entities.

557 The number of particles launched in the Monte Carlo simulation significantly affects the
558 accuracy of the results. A higher number of particles better captures the statistical behavior
559 of the populations. However, the need for more computational resources and time increases
560 with the number of particles. For our purposes, considering the coarse resolution and the high
561 number of simulations required, we launch 50,000 particles for each simulation run.

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

566 This section further describes the comprehensive process of dataset generation, detailing how
567 we simulate measurements and observations that mimic real-world exospheric data.

568 *3.4.1. Measurements and Observations Creation*

569 As explained above, the creation of our datasets starts from a randomly generated surface
570 elemental composition based on realistic mineralogies, establishing the 'ground truth' for each
571 dataset. We then simulate environmental effects to craft a static representation of the exosphere
572 at a specific moment, taking into account the necessary physical and chemical processes.

573 Next, we strategically define various positions within the exosphere to place a virtual sensor,
574 aiming to comprehensively provide "measurements" of its constituents. This conceptual sensor
575 operates without considering the complexities and potential inaccuracies introduced by real-

576 world sensor characteristics such as detector noise and sensitivity limitations. To create
577 a comprehensive and static snapshot of the exosphere, measurements across the simulated
578 exosphere are conducted simultaneously.

579 The positions in the exosphere are selected to maximize data diversity and relevance,
580 considering factors like altitude, latitude, and environmental conditions. To balance the
581 required spatial variability and facilitate the reconstruction of complete surface maps from DNN
582 predictions, we align our virtual sensor positioning in the exosphere directly above each surface
583 grid tile's center. These measurements collectively form what we refer to as an observation or a
584 data subset. Each observation/subset comprises 648 measurement data points per altitude level.
585 Every data point within an observation is a vector that encapsulates the measured densities of
586 all neutral species present in the exosphere as per our simulation's setup. To each data point
587 there is also the corresponding 'ground truth' vector formed from the elemental fractions of the
588 surface tile below.

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

597 *3.4.2. Training, Validation and Test Datasets*

598 We produce three distinct types of datasets: training, hold-out validation, and testing
599 datasets. This subdivision is a fundamental practice in machine learning, ensuring that the
600 algorithm is trained on a diverse set of data, validated for accuracy, and finally tested for
601 generalization to unseen data. Different subsets of generated exospheres are used for training,
602 validating, and testing the DNN. For the training datasets, multiple observations/subsets
603 representing different exospheres are aggregated, enhancing the dataset's size, complexity, and
604 variability. Conversely, the hold-out validation and testing datasets are each derived from
605 individual observations/subsets to maintain representability of particular planetary surface-

606 exosphere pairs.

607 Given the mission-specific orbital characteristics of the BepiColombo/Mercury Planetary
608 Orbiter and the enhanced measurement capabilities of the STROFIO mass spectrometer within
609 the SERENA instrument suite at lower altitudes, we define a baseline constant altitude subset at
610 500 km. This altitude, corresponding to the periherm (closest approach) of the MPO, is chosen
611 for its potential to yield accurate measurements with an improved signal-to-noise ratio. In our
612 training campaign we examine both the baseline dataset type with examples only at altitudes of
613 500 km (648 examples per subset) and the augmented dataset type (4,536 examples per subset),
614 the latter of which provides the MLP DNN with the capability to make predictions at diverse
615 altitudes.

616 The training phase demands a complex strategy that captures a larger part of the data
617 distribution to train the algorithm to approximate the relationships between the employed
618 physical processes accurately. To address this, we aggregate multiple data subsets or observations
619 to form the training dataset, with the number of included subsets ranging from 10 baseline
620 observations (yielding a total of 6,480 data points) to 300 augmented observations (resulting in
621 a staggering 1,360,800 data points). The multiple training sets used in our study vary in size
622 and complexity and present different empirical distributions to the DNN. Training the DNN on
623 each set produces different estimators of the processes, each tailored to the particular empirical
624 distribution of its training set. The construction of a representative training dataset plays a
625 crucial role in ensuring the resulting estimator of physical processes closely approximates the
626 actual data distribution. Additionally, increasing the training set size reduces the probability of
627 the algorithm becoming biased towards a non-representative smaller dataset distribution.

628 For the validation and testing phases of our DNN algorithms, we adopt a different strategy.
629 Hold-out validation involves using a single validation dataset to monitor the trained network
630 during each training epoch. This process helps identify when the algorithm begins to overfit the
631 training data, thereby reducing its generalization capabilities. The validation set is also crucial
632 during hyperparameter tuning to evaluate the algorithm's ability to generalize effectively. At
633 the same time, the test datasets consist of examples (data points) that are never seen during
634 training, ensuring that the MLP DNN algorithm's performance evaluation is not biased by
635 improvements in accuracy due to evaluation on previously seen examples.

636 For validation and test dataset purposes, employing single data subsets allows for

637 straightforward prediction and reconstruction of surface maps corresponding to individual
638 surface-exosphere simulations. Our hold-out validation dataset is selected at the MPO's
639 periherm altitude of 500 km (baseline subset) to maintain consistency in evaluation conditions.

640 The test datasets are designed to thoroughly assess the algorithm's predictive capabilities
641 under varied conditions. To ensure a thorough assessment of the algorithm's performance, we
642 generate a total of 135 test datasets derived from 15 distinct exosphere simulations, each at a
643 constant altitude ranging from 200 km to 1500 km. This diversified testing ground allows us
644 to explore the algorithms' responsiveness to altitude variations. Such an approach facilitates a
645 comprehensive analysis and evaluation across a broader spectrum of exospheric altitudes and
646 conditions, enabling us to more accurately gauge the algorithm's efficacy and robustness in
647 predicting the surface elemental composition. A more detailed description of the datasets and
648 their underlying surfaces is given in Appendix A.

649 *3.4.3. Feature Selection and Engineering*

650 The final step is to supplement the datasets with additional metadata, or features, from these
651 observations, such as geometrical information, to serve as auxiliary inputs for the algorithms.
652 The effectiveness of deep neural network algorithms in modeling complex relationships within
653 data is significantly influenced by the selection and engineering of input features. Carefully
654 chosen or engineered, these features, describing each data point within all three types of datasets,
655 can enhance the algorithm's ability to discern patterns and relationships, thereby improving its
656 overall performance.

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

659 • **Altitude of Measurement:** This feature is critical for capturing altitude-specific
660 dynamics, enabling the algorithm to identify how the distribution of neutral species changes
661 with altitude relative to their source points on the surface. Both the actual altitude and a
662 logarithm of the altitude were tested as features. The latter engineered feature is intended
663 to highlight non-linear altitude effects on the measured parameters, providing another layer
664 of depth to the altitude-related analysis.

665 • **Logarithm of Exospheric Density:** By applying the base 10 logarithm to the exospheric
666 density of each species, we introduce a constraint that aids the algorithm in exploring

nonlinear relationships, acknowledging the exponential decrease in density with altitude.

• **Subsolar Angle:** Represented as either the value of the angle ϕ directly, or as $\cos(\phi - 180)$, this feature helps differentiate between exospheric populations on the dayside, nightside, and the transitional terminator regions, enhancing the model's spatial awareness.

• **Latitude Dependency:** Using $\sin(\gamma)$, this feature allows the algorithm to account for latitude-specific phenomena, such as ion sputtering, which vary across different latitudinal zones.

• **Proton Flux Virtual Data:** Integrating virtual measurements akin to those from the MIPA ion detector of the SERENA suite, this feature hints at sputtering effects induced by proton precipitation on the surface, offering a proxy for understanding underlying ion induced physical processes.

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

4. Results

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

4.1. Training Phase and DNN Finalization

An extensive training campaign was undertaken to explore both the empirical distribution represented in the training datasets, and the hyperparameter space of the neural network architecture. This effort aimed to develop an accurate estimator that demonstrates optimal

generalization capabilities by closely approaching the true data generating distribution. In this section we outline the most important findings of this investigation phase to ultimately refine and finalize the components of the multilayer perceptron deep neural network. The complete training campaign is detailed in Appendix B.

700 Eliminating Skewed Predictions

Initial analysis showed that accuracy metrics for predicting surface elemental composition were skewed by the high prevalence of oxygen (O_2) and silicon (Si). To address this, we excluded these elements from the prediction vector, adjusting the model to focus on the normalized proportions of the remaining nine elements. This adjustment improved the model's relevance and performance by aligning with our study's objectives more effectively.

706 Training Set Size and Data Augmentation

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

714 Learning Curve Examination for Optimal Training Duration

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

721 Selected Feature Set

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

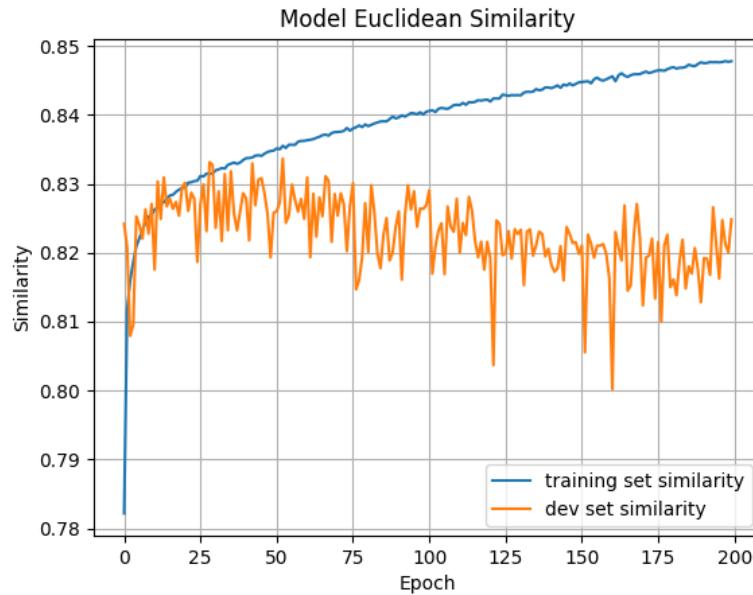


Figure 5. Learning curve for the MLP DNN training. The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (300 subsets, 1,360,800 data points) and the hold-out validation dataset (1 subset, 648 data points) respectively.

- 724 • Logarithmic transformations of elemental species exospheric densities.
- 725 • Logarithm of the altitude at which measurements were taken.
- 726 • Sun incidence angle.
- 727 • Presence of H⁺ ions arriving through open field lines.
- 728 • Local time.
- 729 • Latitude.

730 *Hyperparameter Optimization and DNN Structural Components Finalization*

731 An extensive hyperparameter optimization effort resulted in the selection of the final MLP
 732 DNN architecture, consisting of a four-layer structure with 600, 500, 350, and 250 neurons in
 733 each layer respectively (Figure 6). The regularization coefficient was optimized to a higher value
 734 of 1.0×10^{-5} to improve generalization, while the learning rate was finely tuned to 0.5×10^{-4} .
 735 Training was conducted in mini-batches of 512 examples, identified as near-optimal through our
 736 optimization process.

737 The combined efforts of hyperparameter exploration, architectural fine-tuning, and dataset
 738 augmentation have significantly propelled our model's performance and its ability to generalize

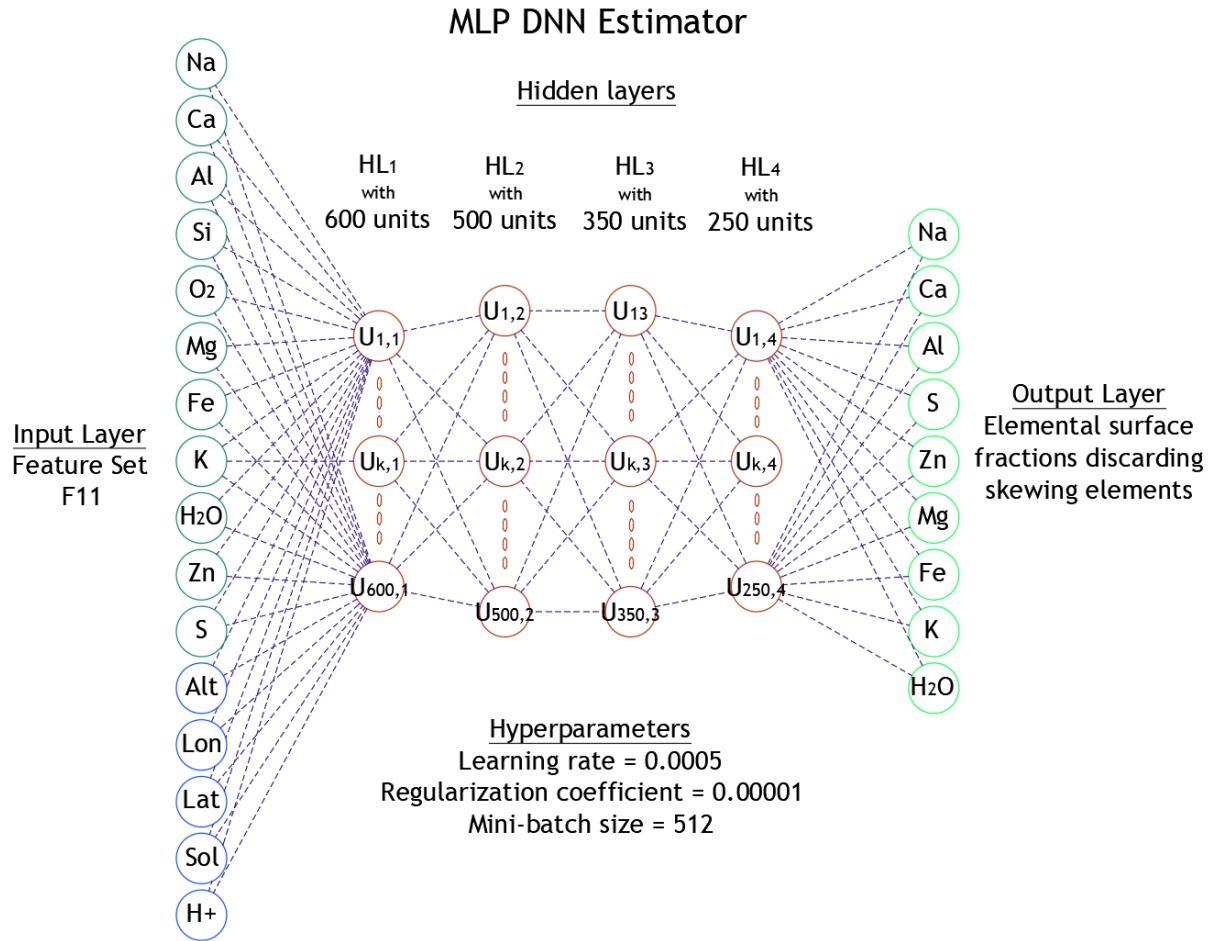


Figure 6. Finalized multilayer perceptron deep neural network. The input layer consists of the features collected in feature set F11, the output layer is adjusted to remove the skewing predictions towards better estimating the O₂ and Si species. There are four hidden layers with 600, 500, 350, and 250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

from the empirical distribution to the true underlying data-generating distribution—by as much as a total of 10% in ES4 and up to 30% in R-squared. This optimized architecture, along with structural parameters and algorithmic characteristics refined during our comprehensive training campaign, ensures that the MLP DNN is a robust model for our sophisticated predictive tasks.

4.2. Testing Phase

The ultimate evaluation of our multilayer perceptron deep neural network algorithm's performance hinges on its ability to accurately predict surface compositions and reconstruct elemental surface maps using datasets it has not previously encountered. Our research incorporates two distinct test campaigns designed to assess the MLP network's predictive

748 prowess. These campaigns were structured to apply the final network, fine-tuned with an
749 extensive training set comprising 300 augmented subsets, across test datasets derived from a
750 variety of altitudes not previously seen during training.

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

757 *4.2.1. Preliminary Test Campaign*

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

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

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

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

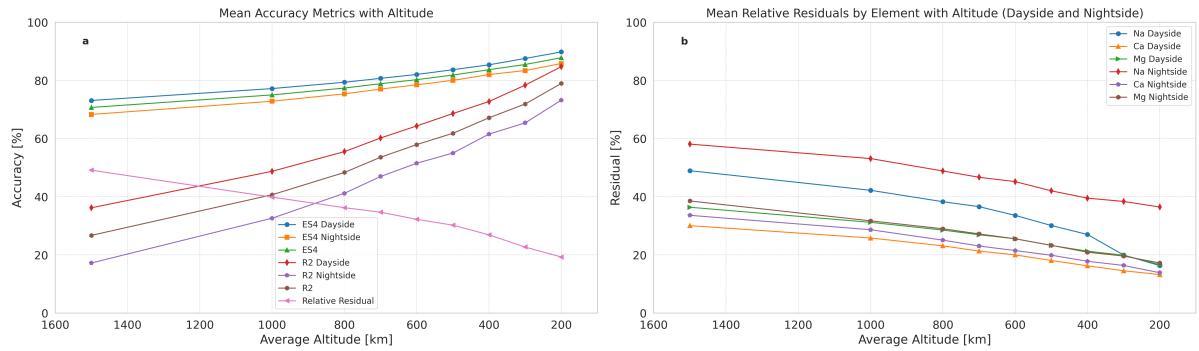


Figure 7. Preliminary test campaign. Panel **a** shows the mean accuracy (dayside, nightside, and overall) and relative residuals metrics of the MLP DNN predictions on the 15 test surfaces. In panel **b** are displayed the dayside and nightside mean relative residuals for the elements Na, Ca, and Mg.

778 4.2.2. Main Test Campaign

779 We note that Mercury has a unique rotational and orbital dynamics, particularly its 3:2 orbit-
 780 spin resonance. This results in the nightside hemisphere becoming the dayside in every following
 781 orbit. Thus, by using a combination of data from two consecutive years, we can fill the gap in the
 782 predictions of the volatile species, as leveraged in the main test campaign. This approach utilized
 783 double-simulation compound predictions, focusing on the same 15 surface compositions from
 784 the preliminary campaign but observed at two consecutive perihelia. During these two periods,
 785 different halves of Mercury's surface were illuminated by the Sun, allowing for comprehensive
 786 daylight observation of the entire planet over the two simulations. For this campaign, predictions
 787 specifically targeted sunlit surface tiles, enabling an in-depth analysis of surface compositions
 788 that were previously on the nightside in the initial test phase. Measurements for this campaign
 789 were again taken at a range of altitudes from 200 to 1500 km.

790 A significant outcome of the combined odd-even orbit campaign was the improved accuracy
 791 in predicting volatile species' fractions, aligning more closely with the refractory species'
 792 predictions observed in the preliminary campaign. This enhancement in predictive accuracy for
 793 volatiles under daylight conditions underscores the importance of solar illumination in accurately
 794 assessing surface compositions.

795 The main campaign demonstrated a notable increase in overall prediction and map
 796 reconstruction accuracy, with the average ES4 metric reaching approximately 89.70% and the
 797 average R-squared metric reaching 83.41% at the lowest altitude of 200 km (Figure 8). This

4.2 Testing Phase

4 RESULTS

accuracy diminished at higher altitudes, attributed to the exosphere's dynamic nature and the increased complexity in tracing back exospheric particles to their originating surface tiles. There was, however, a marked improvement compared to the preliminary findings, which highlights the efficacy of considering Mercury's solar exposure in enhancing predictive models' accuracy. By focusing solely on the dayside observations across two perihelia, the campaign effectively capitalized on optimized conditions for surface composition reconstruction. The box statistical plots shown on Figure 8 suggest also reduction in the range of prediction accuracies and residuals with decrease in altitude.

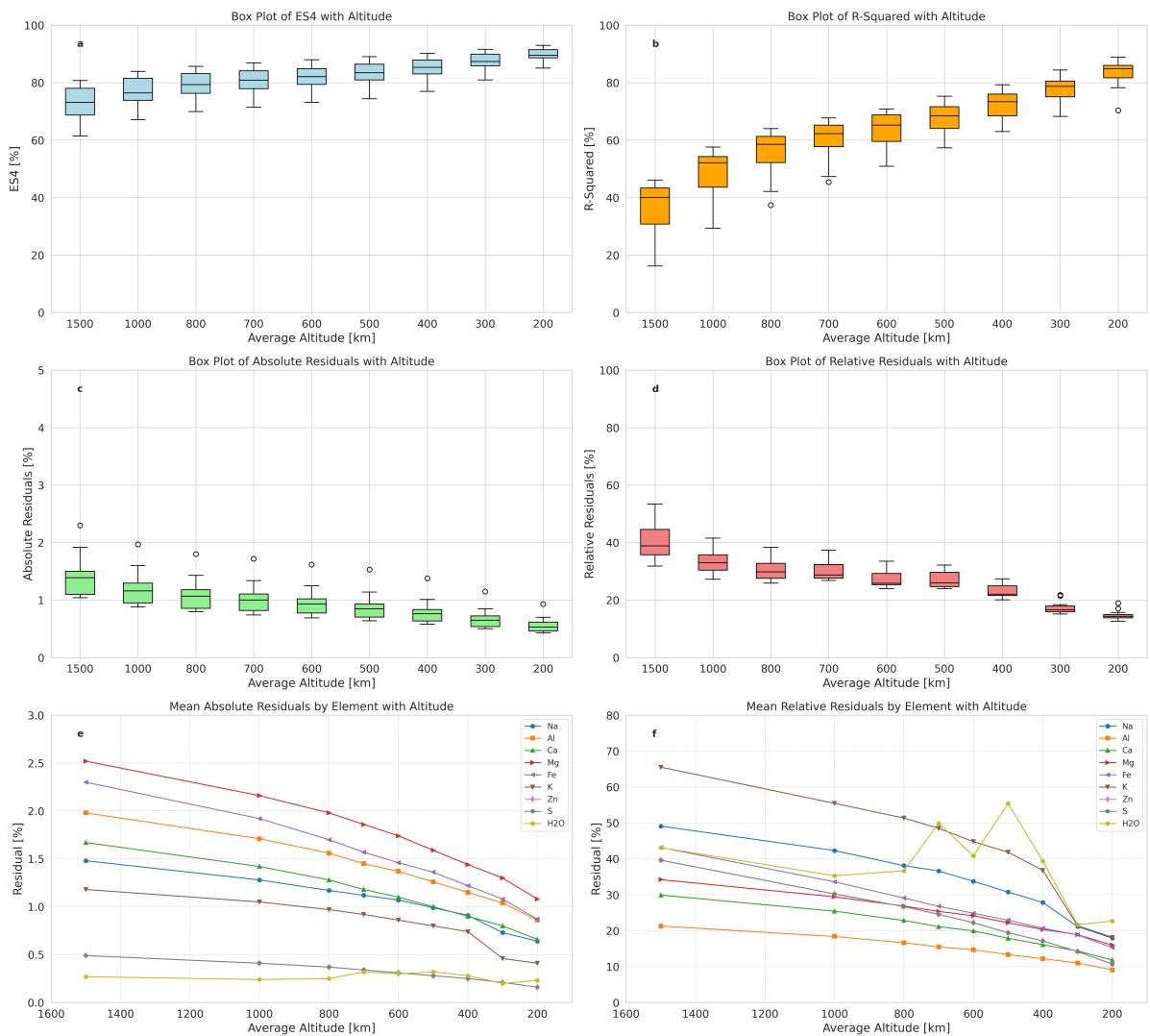


Figure 8. Box plots of the accuracy metrics (panels a and b) and the mean residuals (panels c and d) of the MLP DNN predictions on the 15 test surfaces of the main test campaign. Panels e and f give respectively the absolute and relative residuals for all predicted elements averaged over the 15 predicted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

806 Going deeper in the detailed statistics of the predictions by the MLP DNN for each elemental
 807 species, we can observe its tendencies in the box plots of the absolute and relative predictions for
 808 Aluminium, Calcium, and Sodium (Figure 9). Box plots for the remaining elements are provided
 809 in Appendix C. Almost all elements have good prediction statistics at the lowest altitude, with
 810 the median relative residual of Aluminium particularly impressive at only 8.69% at 200 km.
 811 The algorithm has more difficulties with Sodium, with its median relative residual at 17.86% at
 812 200 km, which is nevertheless a good result. The robustness in the prediction of the refractive
 813 elements is present throughout the altitudes, even up to 1500 km, where the median relative
 814 residuals of Aluminium is 21.18%, while that for Calcium is 30.76%. While the range of the
 815 prediction errors for the different types of surface elements is impressive through the altitudes.

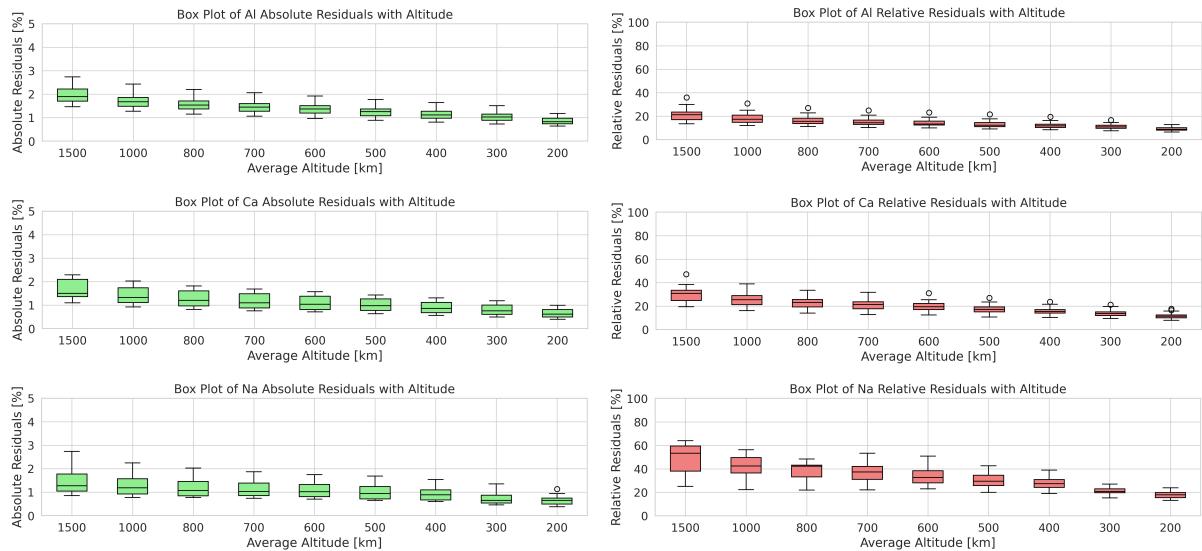


Figure 9. Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the elements Aluminium, Calcium, and Sodium.

816 Additionally, our visual comparisons from the map reconstructions (Figures 10 to
 817 12) highlighted the algorithm's strengths and weaknesses in predicting different elemental
 818 distributions.

819 The discrepancy in predictive accuracy between refractory species, such as Aluminium (Figure
 820 10) and Calcium (Figure 11), and volatile ones, like Sodium (Figure 12), is still present in the
 821 predictions of the particular test set 2 shown on the figures. However, there is a noticeable
 822 improvement by the daylight only predictions of the main testing campaign, compared to the
 823 one from the preliminary campaign, which included the night side prediction. This is shown

4.2 Testing Phase

4 RESULTS

on the map reconstructions of Sodium (test set 2), where the bottom-most panels in Figure 12 show that increased errors on the night side (longitudes 0-90 and 270-360) and an average relative residual of 23.73% from the preliminary test, directly 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%.

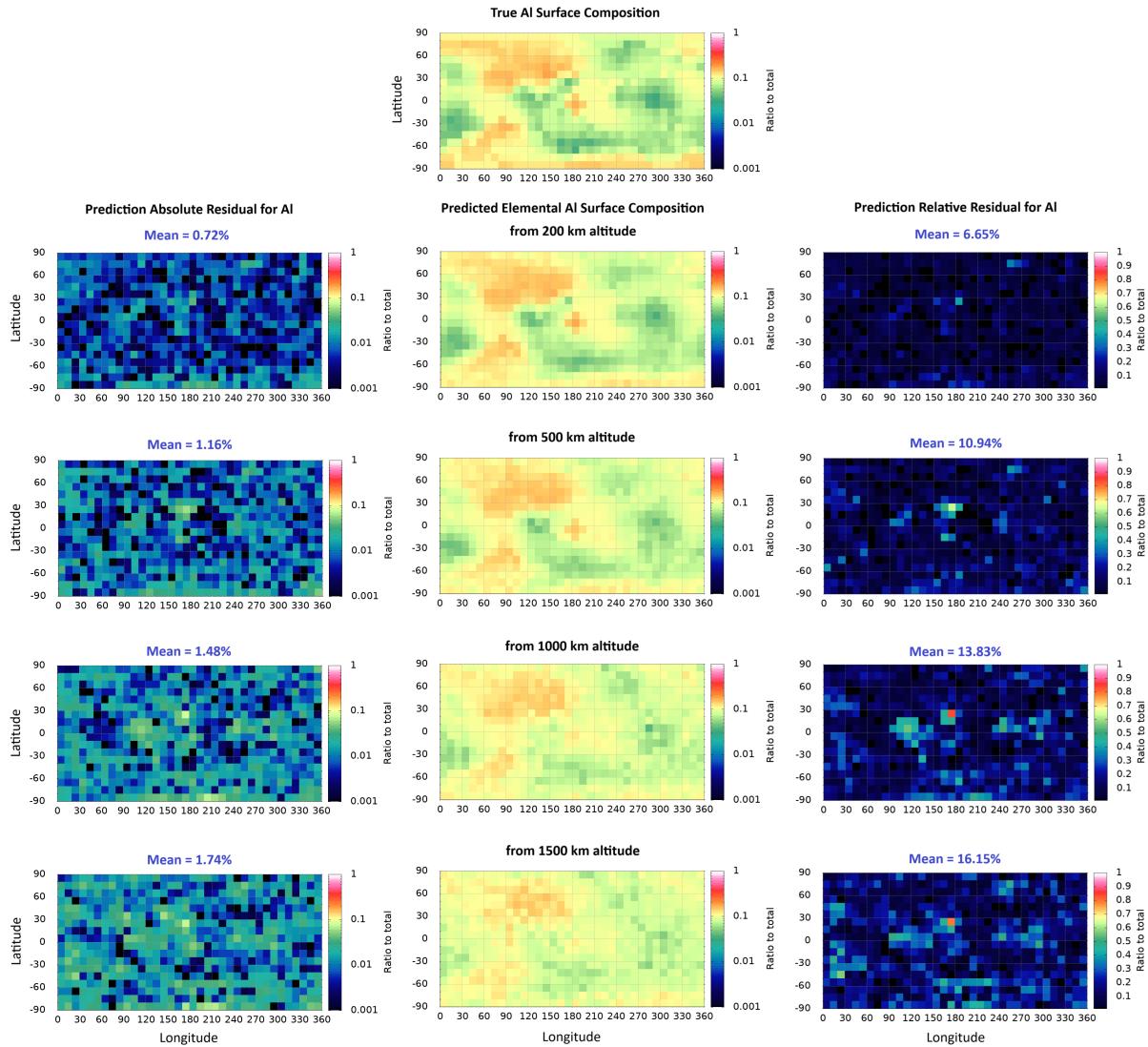


Figure 10. Main test campaign - MLP DNN map reconstructions of the same sample Aluminium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

From a closer examination of the reconstructed maps, we can clearly see that certain large

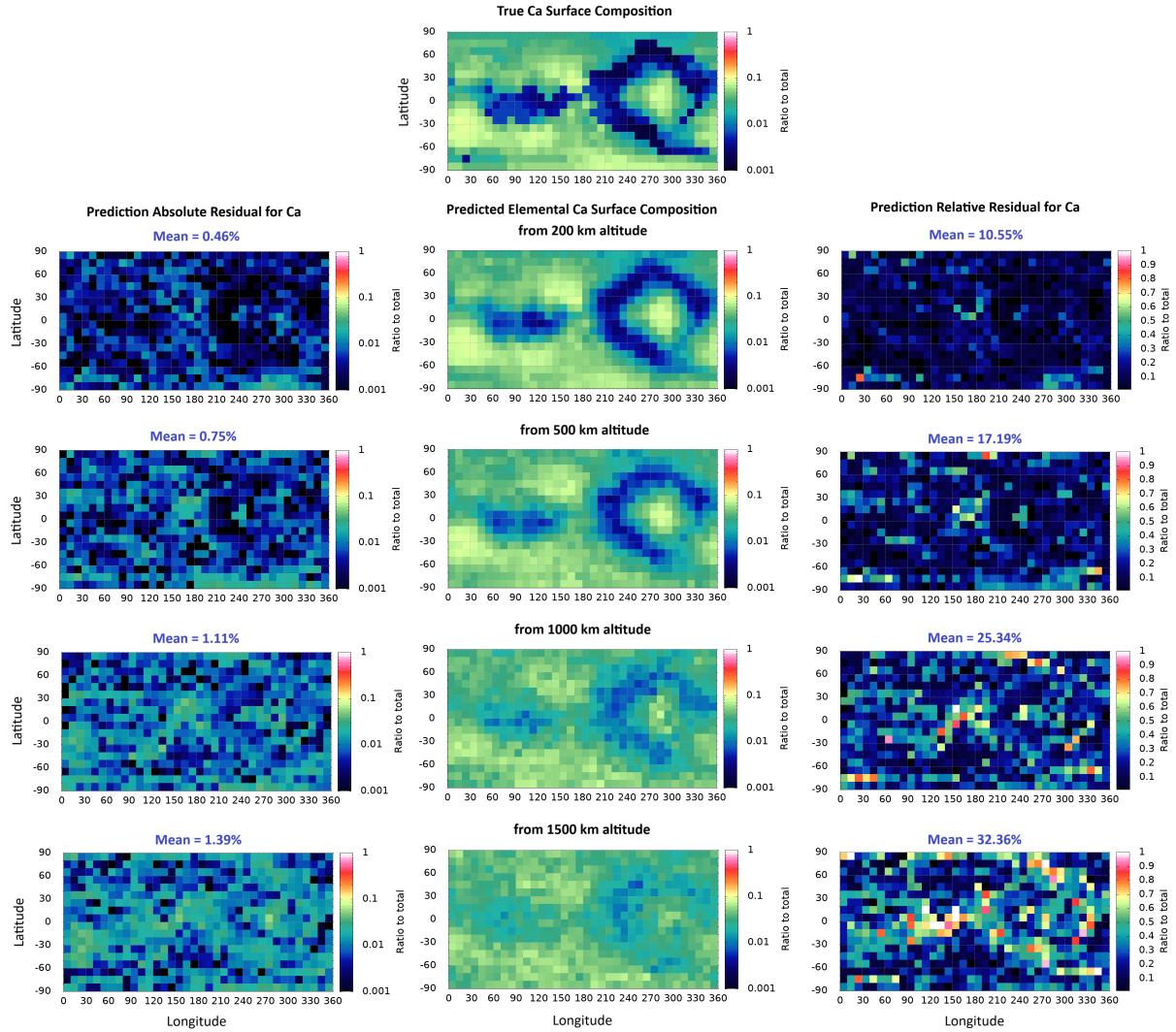


Figure 11. Main test campaign - MLP DNN map reconstructions of the same sample Calcium surface composition (test set number 2). Inputs are coming from the dayside of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map 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.

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

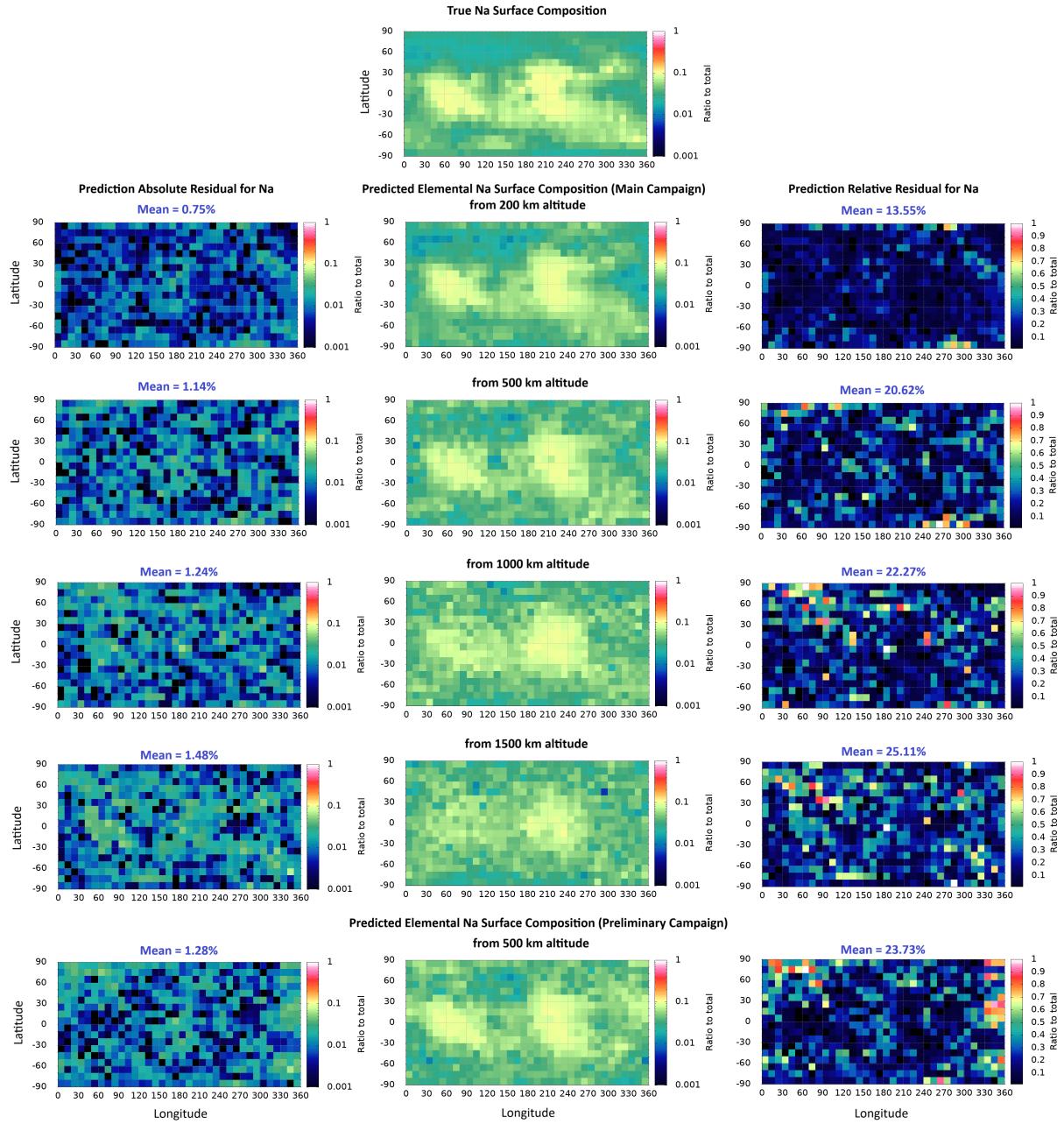


Figure 12. Main test campaign - MLP DNN map reconstructions of the same sample Sodium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

836 C.

837 **5. Discussion and Future Work**

838 This research presents a novel approach for deducing Mercury's surface composition using
839 advanced deep neural networks to analyze exospheric density measurements. The algorithm
840 is adept at predicting relative elemental compositions and reconstructing surface maps, crucial
841 for understanding how neutral atoms transit from the planet's regolith into its exosphere. Key
842 characteristics of the method include the definition and utilization of the simulated model's
843 physical parameter space, the development of a sophisticated multilayer perceptron DNN
844 architecture, the application of Bayesian hyperparameter tuning for optimal configuration, and
845 the integration of domain-specific knowledge into the feature selection process.

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

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

864 The developed algorithms for surface composition reconstruction achieved high fidelity in
865 their predictions, with mean accuracy metrics reaching 89.70% (ES4) and 83.41% (R-squared)

5 DISCUSSION AND FUTURE WORK

866 across 15 test sets at 200 km altitude. Additionally, both mean absolute and relative residuals
867 of elemental predictions showed a robust trend of decreasing with altitude, reaching as low as
868 0.56% and 14.70% respectively. The effectiveness of these algorithms was particularly notable in
869 reconstructing elemental composition maps from low-altitude measurements (200km to 700km),
870 especially for sufficiently represented elements, such as aluminium, calcium, magnesium, and
871 sodium. At the same time, minor elements, such as water, exhibited jumps and anomalies in
872 their predictions, attributed to their low overall fraction in our model and their randomized
873 placement on the surface (not constrained to the poles only). Meanwhile, the MLP was able to
874 capture the coupling of sulfur and zinc in its predictions, even if the input exospheric densities
875 are widely different for the two elements.

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

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

892 Another critical area of future research is testing DNNs trained with data from one region of
893 the physical processes parameter space against data distributions from different regions. This
894 exploration is essential to assess the DNNs' performance when applied to various models of
895 exospheric production and their respective data distributions, potentially constructing another
896 layer for analyzing the surface-exosphere interactions with this innovative tool.

5 DISCUSSION AND FUTURE WORK

897 One potential application of this approach involves utilizing multiple DNNs trained on
898 different regions of parameter space. Input data could be passed through all of the pre-built
899 DNNs, and their predictions juxtaposed in a subsequent layer of the algorithm to estimate
900 the input data's underlying generation mechanism. This would constrain the range of physical
901 parameters to those of the DNNs with the highest accuracy on the input data.

902 Another potentially groundbreaking utilization may explore the mapping between the
903 physical processes space, defined by analytical equations, and the purely data-driven DNN
904 representation, built from the internal weight matrices of the neural nets. This mapping could
905 serve the dual purpose of understanding how a description of reality constructed strictly from
906 data relates to the description by physical equations, and exploring potential synergies between
907 the two in describing the real world.

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

918 The contrast between using simulated data and incorporating real observational data
919 into our algorithm development merits further exploration. While simulations provide a
920 controlled environment for testing various scenarios, they do not capture the full complexity
921 and unpredictability of actual exospheric data. This could lead to significant deviations
922 in the parameter space from those assumed in our simulations, as real processes and their
923 interdependent variables may change over time and are not fully represented in simulations.

924 Shifting our focus from simulated to real physical processes is a bold and potentially
925 transformative step. Developing an estimator capable of effectively processing and analyzing
926 real-world data from Mercury's exosphere would significantly advance our understanding of
927 planetary surfaces and their interactions with their environments. This progress would not only

5 DISCUSSION AND FUTURE WORK

928 deepen our theoretical knowledge but also offer practical insights into the formation, dynamics,
929 and evolution of planetary exospheres.

930 This study is performed in anticipation of the upcoming BepiColombo ESA/JAXA mission.
931 The mission will deploy two spacecraft—the Mercury Planetary Orbiter (MPO) and the Mercury
932 Magnetospheric Orbiter (MMO)—equipped with a suite of instruments aimed at understanding
933 Mercury’s surface, exosphere, and magnetosphere (Benkhoff et al., 2010; Milillo et al., 2010,
934 2020). Particularly, our study targets future utilization of measurements from the SERENA
935 (Search for Exospheric Refilling and Emitted Natural Abundances) suite on the MPO, which
936 includes sensors like STROFIO (STart from a ROtating Field mass spectrOMeter), MIPA
937 (Miniature Ion Precipitation Analyser), PICAM (Planetary Ion CAMera), and ELENA (Emitted
938 Low Energy Neutral Atoms) to analyze exospheric gas composition, plasma fluxes, and ion
939 precipitation (Orsini et al., 2010, 2020; Milillo and Wurz, 2014). Notable instruments aboard
940 BepiColombo, which may provide images of the surface, include MIXS, MGNS, MERTIS, and
941 SIMBIO-SYS (Benkhoff et al., 2010).

942 The application of these methods to the observational data from BepiColombo’s suite of
943 instruments offers a promising path to refine models of exosphere generation. By comparing
944 predicted surface compositions with actual measurements, we can more accurately constrain
945 our models, enhancing our understanding of planetary processes.

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

957 **Acknowledgments**

958 This work was supported by the Italian Space Agency (ASI) - SERENA contract no. 2018-8-

959 HH.1-2022 "Scientific participation in the mission BepiColombo SERENA - Phase E1".

APPENDIX A DETAILED DATASETS

960 **Appendix A. Detailed Datasets**

961 *Appendix A.1. 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.

962 *Appendix A.2. Consolidated Datasets*

Appendix A.2 Consolidated Datasets

APPENDIX A DETAILED DATASETS

Surfaces	Na	Al	Si	O ₂	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						
Surf-Exo Pair Name	Resulting Dataset Type	# of Examples Per Dataset	Mercury TAA	Altitude Range [km]		Total # of Datasets from Surf-Exo pair
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.

963 **Appendix B. Detailed Training Campaign**

964 *Starting Point*

965 Our journey in the big training campaign endeavour commenced with a baseline training
966 dataset, which, being the simplest, was least representative of the true data generation
967 distribution. This dataset comprised 10 subsets with a total of 6,480 data points from the
968 F00 feature set (only exospheric density measurements). The initial architecture of the MLP
969 DNN was modeled closely after the structure proposed in the preliminary study of Kazakov
970 et al. (2020), featuring a four-layered network with 400, 200, 200, and 100 neurons, respectively.
971 Notably, our study expanded the input layer to accommodate a greater number of elements -
972 11 total input elements.

973 The output layer of the network employs softmax units, designed to predict the surface
974 composition of the same 11 elements provided as inputs. Initially, the minibatch size was set to
975 1,024 examples. The regularization L-2 coefficient and the learning rate were chosen as 1.0×10^{-6}
976 and 0.5×10^{-4} , respectively, to balance the trade-off between learning efficiency and the risk of
977 overfitting.

978 *Eliminating Skewed Predictions*

979 The initial analysis of predicting fractionated surface elemental composition revealed that
980 the accuracy metrics were significantly skewed by the prevalence of certain abundant elements,
981 notably oxygen (O_2) and silicon (Si), which are omnipresent in most of the minerals in our model.
982 This skewness, stemming from the algorithm's propensity to more accurately predict these two
983 elements, was addressed by excluding them from the prediction vector and adjusting it to ensure
984 a normalized sum of 1. Consequently, the refined model focuses on predicting the normalized
985 proportions of the remaining nine elements, with a subsequent denormalization process applied
986 for the map reconstruction purposes. This strategic exclusion of the most abundant elements
987 led to a marked enhancement of approximately 4% in the predictive R^2 accuracy for the other
988 nine elements. It is important to note, however, that the input vector maintained its original
989 composition of 11 elements.

990 This decision to modify the output layer by removing two elements was driven by a clear
991 rationale: the omnipresent elements, though significant, held less interest for the objectives
992 of our study compared to the other elements. This approach underscores our commitment to

993 optimizing the model’s performance where it matters most, despite recognizing that alternative
994 configurations of the output layer might exist.

995 *Training Set Size and Learning Curve Examination*

996 Exploring the behavior of the initial MLP DNN involved examining its performance
997 in relation to the expansion of the training dataset size and the extension of training
998 duration. The aim was to demonstrate the algorithm’s nominal operation during both
999 training and inference phases by analyzing its learning curves. This included assessing training
1000 and generalization accuracies across a training dataset and the hold-out validation dataset,
1001 respectively. Additionally, identifying the optimal training duration for inference was crucial to
1002 mitigate the risk of overfitting, in line with the guidance provided by Bengio (2015).

1003 Our investigation spanned training sets ranging from 10 to 200 unaugmented data subsets. We
1004 observed a clear positive relationship between increasing the dataset size and the enhancement
1005 of generalization accuracy.

1006 In parallel, the algorithm’s behavior was monitored in terms of its optimization process over
1007 multiple iterations (epochs) across the entire training dataset, employing stochastic gradient
1008 descent to converge to the minimum of the loss function. Analysis of the learning curves
1009 revealed a maximum in predictive performance on the validation dataset after 40 epochs. This
1010 was in contrast to the outcomes observed at 200 epochs of SGD, despite the training dataset’s
1011 distribution increasingly aligning with each additional training iteration.

1012 The learning curve depicted in Figure B1 also hints at the potential for further enhancements
1013 in training predictions, given the rapid ascent observed towards the training’s culmination.
1014 However, to ensure robust inference capabilities, it’s imperative to diminish the variance. This
1015 could potentially be achieved by incorporating a greater number of training examples and/or
1016 intensifying the regularization measures.

1017 *Feature Sets Examination*

1018 The evaluation of feature sets played a pivotal role in optimizing the performance of the
1019 MLP DNN, particularly through the training of the network with various input features across
1020 the unaugmented 200-subset training dataset. The assortment of the examined feature sets is
1021 detailed in Table B1. The iterative process of enhancing input features resulted in substantial
1022 improvements in prediction accuracy for several modified feature sets with ES4 going up by

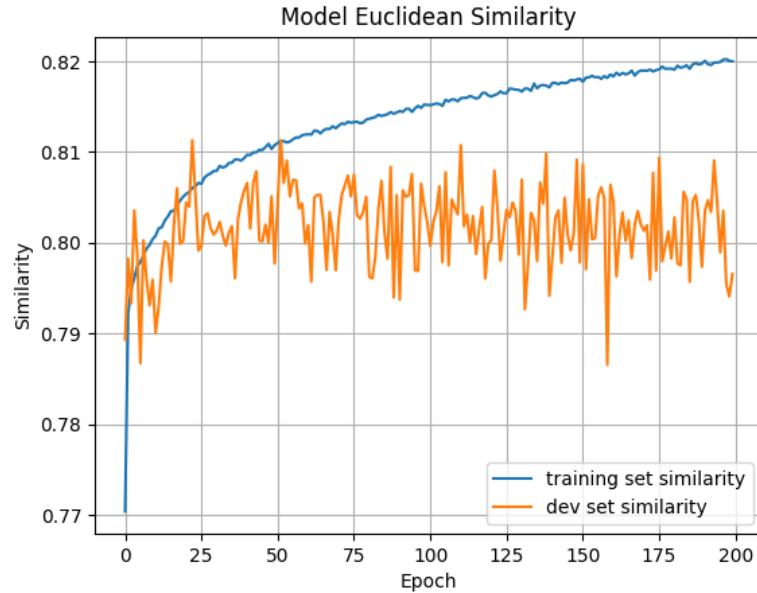


Figure B1. Learning curve for the MLP DNN training. The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (200 subsets, 129,600 data points) and the development hold-out validation dataset (1 subset, 648 data points) respectively.

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.

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.

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

APPENDIX B DETAILED TRAINING CAMPAIGN

1029 transformations of elemental species exospheric densities, which provide a normalized scale for
1030 comparing densities of various elements; (2) the logarithm of the altitude at which measurements
1031 were taken, introducing a scale that accommodates the wide range of altitudes without skewing
1032 the data; (3) Sun incidence angle, accounting for the variation in solar energy impacting the
1033 elemental composition; (4) the presence of H⁺ ions arriving through open field lines, a feature
1034 indicating solar wind interaction with the planetary surface; (5) Cosine of solar time longitude,
1035 offering a representation of the position in solar time longitude; and (6) Sine of latitude, providing
1036 a function to capture latitudinal variations.

1037 The selection of F11 as the final feature set was predicated on its ability to yield the
1038 most reliable and accurate predictions, thereby encapsulating the intricate dynamics and
1039 characteristics vital for elemental composition analysis. This feature set's efficacy highlights
1040 the nuanced approach required in feature selection to enhance neural network performance for
1041 specific predictive tasks. All the future training and testing were performed with this feature
1042 set as inputs to the DNN.

1043 *Hyperparameter Optimization and DNN Structural Components Finalization*

1044 In our quest to fine-tune the multilayer perceptron for optimal performance, a significant
1045 focus was placed on hyperparameter optimization. This process was critically informed by
1046 the parameters outlined in Section 2.2, employing a Bayesian search strategy to navigate the
1047 hyperparameter space efficiently. Our methodology involved running the optimization process
1048 five times, with each iteration spanning 50 cycles and starting from a point incrementally
1049 closer to the previously identified minimum, for a total of 250 cycles. This approach was
1050 instrumental in inching towards the optimal hyperparameter settings, with subsequent iterations
1051 yielding diminishing returns, indicative of approaching a plateau near the optimal values in the
1052 hyperparameter space.

1053 During this campaign, the selection of loss functions emerged as a pivotal consideration, with
1054 our experiments revealing substantial variations in model performance across different functions.
1055 The discerning application of loss functions, particularly the adoption of the KL-divergence for
1056 evaluating probability-like outputs, marked a jump in performance.

1057 The culmination of our hyperparameter optimization efforts led to the finalization of the MLP
1058 DNN architecture, characterized by a four-layered structure with 600, 500, 350, and 250 neural

APPENDIX B DETAILED TRAINING CAMPAIGN

1059 units respectively (Figure 6). An adjustment was made to the regularization coefficient, setting
1060 it to the found higher value of 1.0×10^{-5} to enhance model generalization. Concurrently, the
1061 learning rate was optimized to 0.5×10^{-4} , balancing the trade-off between learning speed and
1062 stability. Training was conducted in mini-batches of 512 examples, a size determined through
1063 our optimization exercises to be close to optimal. This meticulously optimized structure and
1064 parameter set represent the culmination of our comprehensive campaign to refine the MLP DNN,
1065 ensuring it stands as a robust model for our advanced predictive task.

1066 *Augmented Data Study*

1067 In the concluding phase of our training campaign, we embarked on a strategic initiative to
1068 enhance the representability of the empirical distribution, thereby aligning it more closely with
1069 the true data-generating distribution—a target that remains inherently elusive due to limited
1070 direct access. This endeavor was pursued through the deliberate augmentation of our training
1071 datasets, an approach that involved the integration of additional examples derived from the same
1072 exospheric observations that constituted our initial datasets. However, these new inclusions were
1073 distinct in their representation of varying altitudes, thereby enriching the diversity and depth
1074 of our training data.

1075 The initial expansion of our dataset to encompass 200 augmented subsets had already
1076 demonstrated significant promise in enhancing the model’s performance. Motivated by these
1077 preliminary successes, we ambitively expanded our dataset even further to include a total of
1078 300 augmented subsets, culminating in an impressive 1,360,800 examples. This substantial
1079 augmentation effort was driven by the rationale that incorporating measurements from varying
1080 altitudes would not only bolster the dataset’s comprehensiveness but also empower our model
1081 to predict with greater accuracy across a diverse range of altitude-specific inputs.

1082 The outcome of this labor were unmistakably positive, with the augmented datasets markedly
1083 improving the robustness and accuracy of our MLP DNN, increasing further the validation set’s
1084 ES4 to 84.0% (+1.5%) and its R^2 to 63.5% (+3.5%). The strategic inclusion of altitude-varied
1085 examples was particularly impactful, enabling the algorithm to achieve enhanced predictive
1086 precision for inputs across different altitudes.

1087 *Implications and Results of the Training*

1088 As our meticulous exploration of the hyperparameter space culminated in identifying a region

APPENDIX B DETAILED TRAINING CAMPAIGN

1089 that, while not conclusively the ultimate minimum, demonstrates unparalleled accuracy in
1090 inferences on the hold-out validation dataset, we arrived at several pivotal implications and
1091 results from our training campaign. This journey through hyperparameter optimization has
1092 yielded a collection of finely tuned multilayer perceptron deep neural networks, each reflecting
1093 a nuanced understanding of the underlying data-generating processes.

1094 Firstly, one of the outcomes of this campaign is the demonstration of the algorithm’s efficiency,
1095 achieving optimal training within 40 complete epochs. This not only highlights the effectiveness
1096 of our chosen architecture but also underscores the potential for accuracy improvements with
1097 the expansion of the training dataset. Such findings affirm the architectural decisions made in
1098 designing our MLP DNN for the task at hand.

1099 Secondly, our investigation revealed the critical role of specific features in guiding the
1100 algorithm toward more precise predictions of exospheric measurements and surface composition.
1101 The identification of these key features underscores the importance of thoughtful feature selection
1102 in enhancing model performance.

1103 Thirdly, the exploration led to the refinement of the MLP’s internal structure, significantly
1104 bolstered by experiments with various loss functions and output layers, alongside the application
1105 of Bayesian hyperparameter optimization. While acknowledging that the realm of possible
1106 architectural enhancements remains vast, the current configuration stands as a testament to the
1107 robustness and efficacy of our model.

1108 Lastly, the strategic augmentation of our dataset with additional exospheric observations
1109 has unequivocally improved the algorithm’s predictive capabilities. This expansion not only
1110 enriches the model’s training environment but also enhances its ability to generalize across a
1111 broader spectrum of the empirical distribution, thereby moving closer to the elusive true data-
1112 generating distribution.

1113 The combined efforts of hyperparameter exploration, architectural fine-tuning, and dataset
1114 augmentation have significantly propelled our model’s performance. Through this comprehensive
1115 training campaign, we have not only achieved a high degree of accuracy in our predictions
1116 but also laid a solid foundation for future research to build upon, promising even greater
1117 advancements in our understanding and representation of complex data-generating processes.
1118 Through this concerted effort, we have significantly advanced the model’s capacity to generalize
1119 from the empirical distribution to the true underlying data-generating distribution.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

1120 **Appendix C. Box Plots and Map Reconstructions of the Main Test Campaign**
 1121 **Predictions by Species**

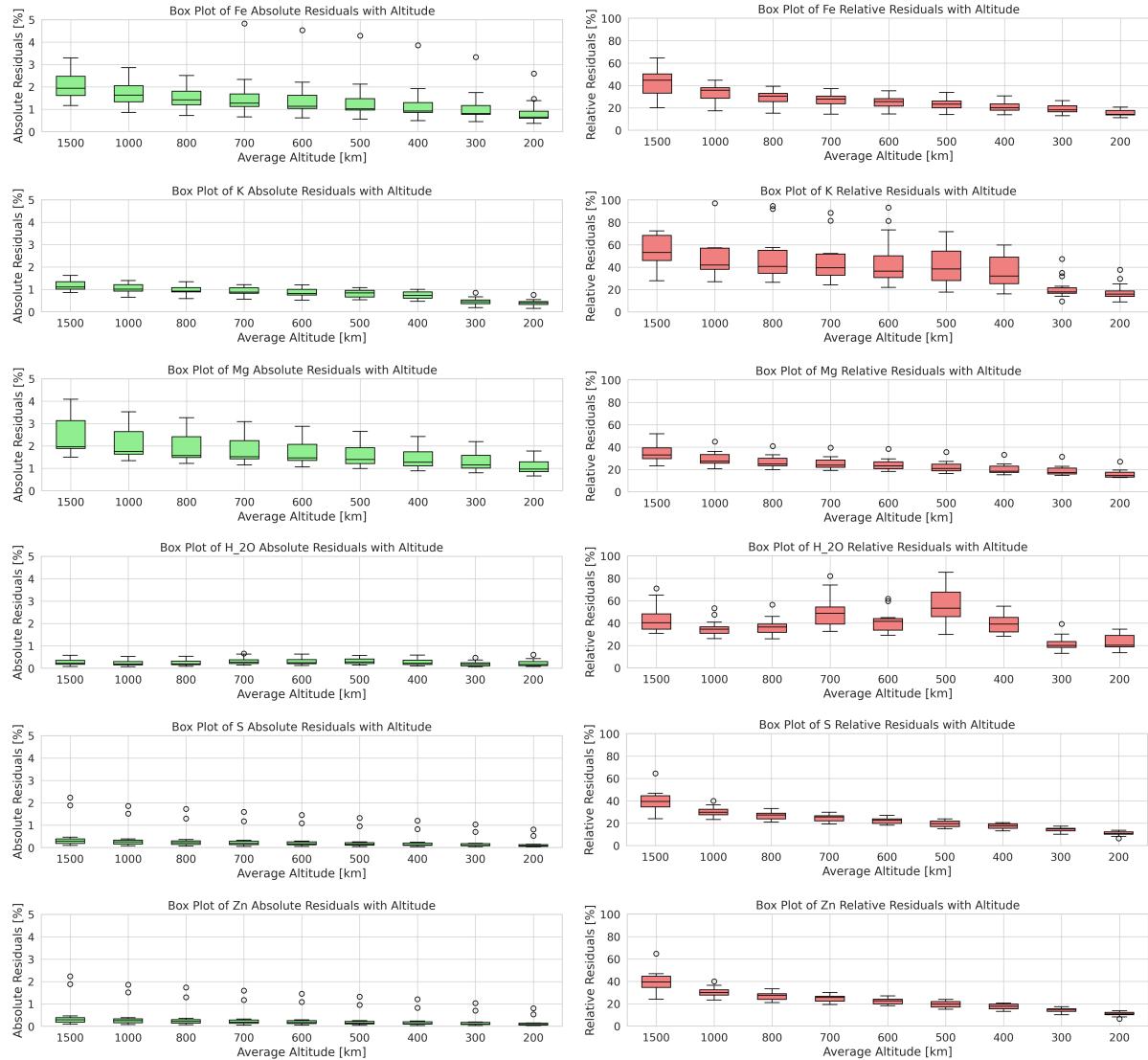


Figure C1. Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Iron, Potassium, Magnesium, Water, Sulfur, and Zinc.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

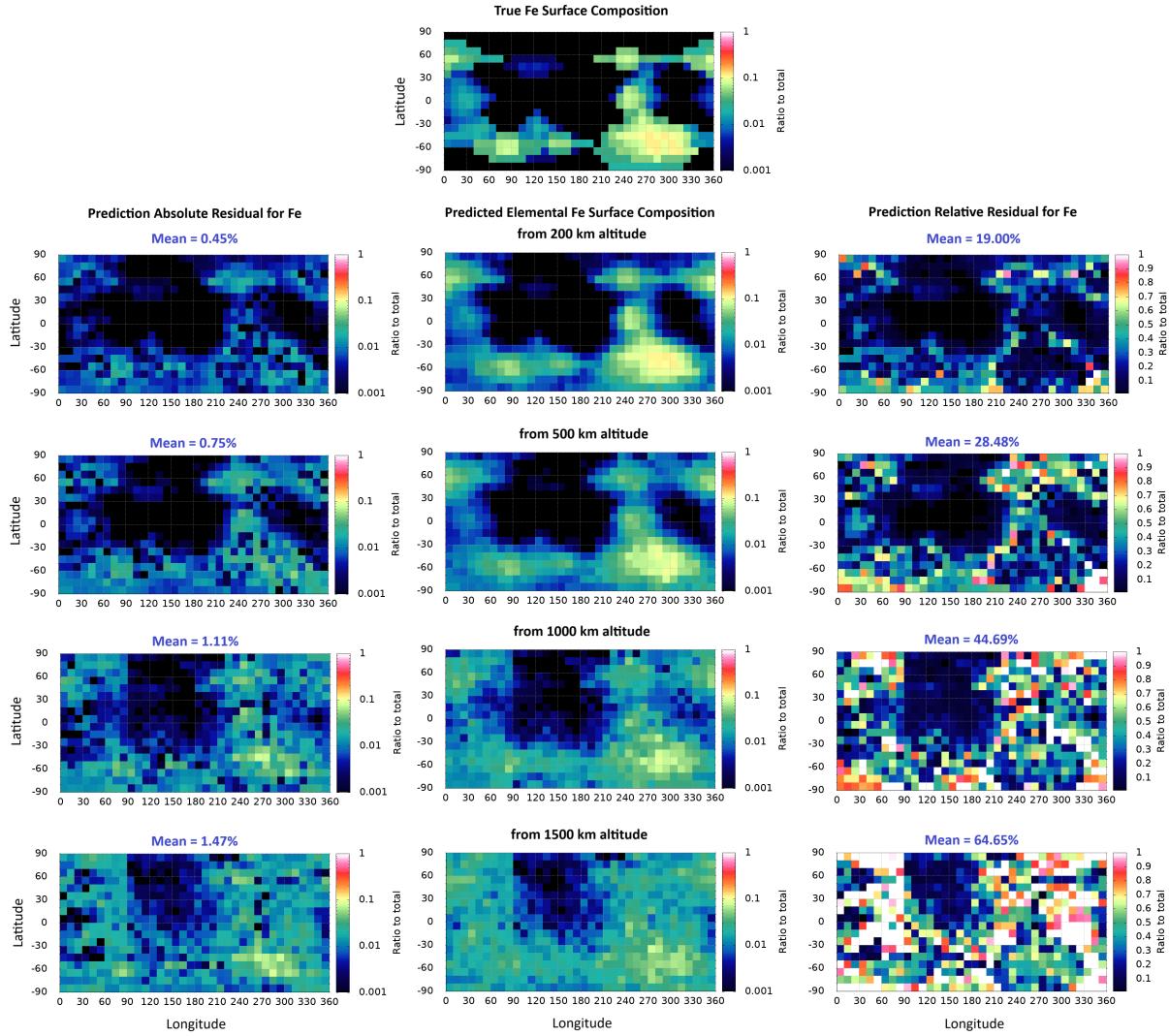


Figure C2. Main test campaign - MLP DNN map reconstructions of the same sample Iron surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

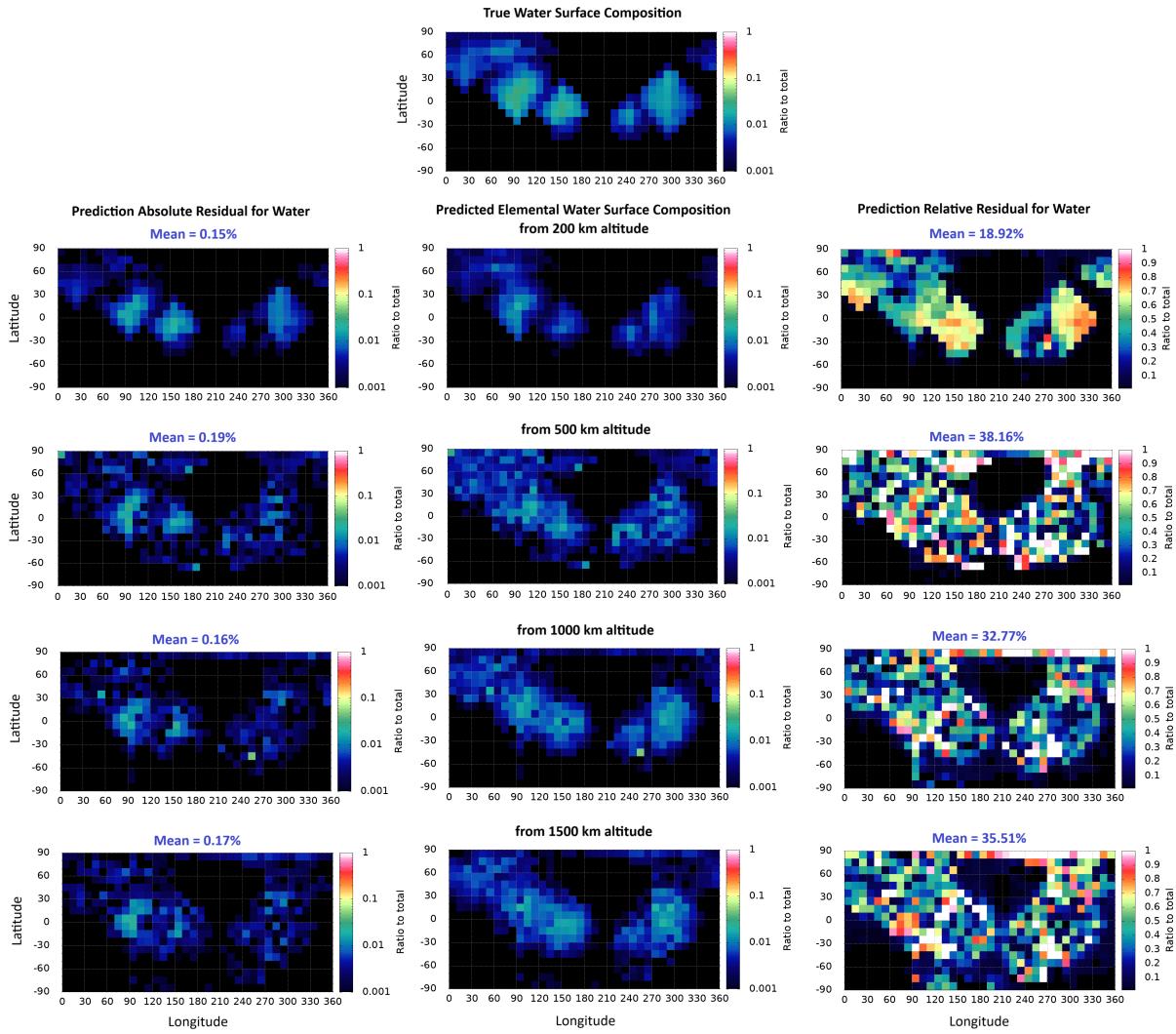


Figure C3. Main test campaign - MLP DNN map reconstructions of the same sample Water surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

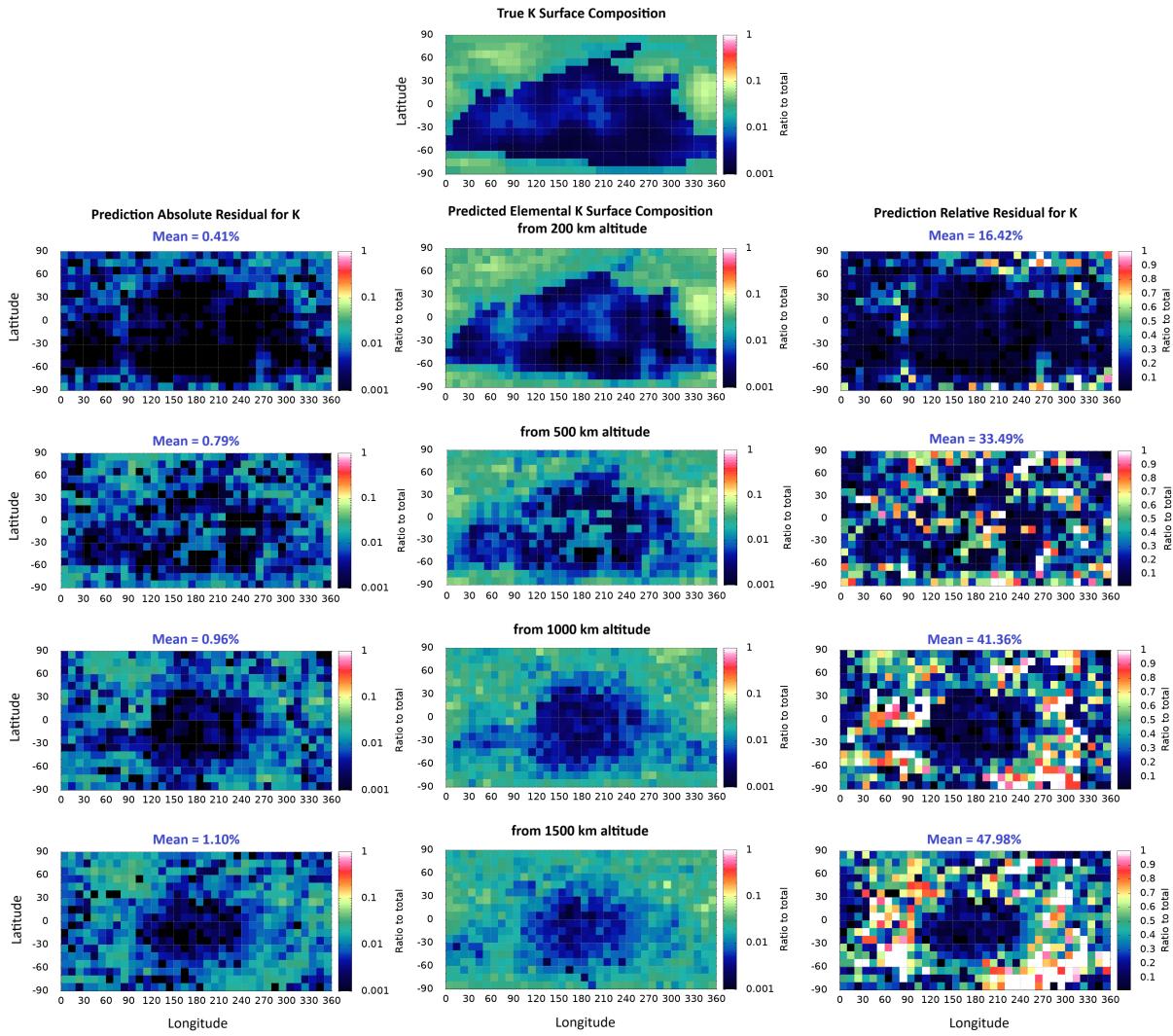


Figure C4. Main test campaign - MLP DNN map reconstructions of the same sample Potassium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

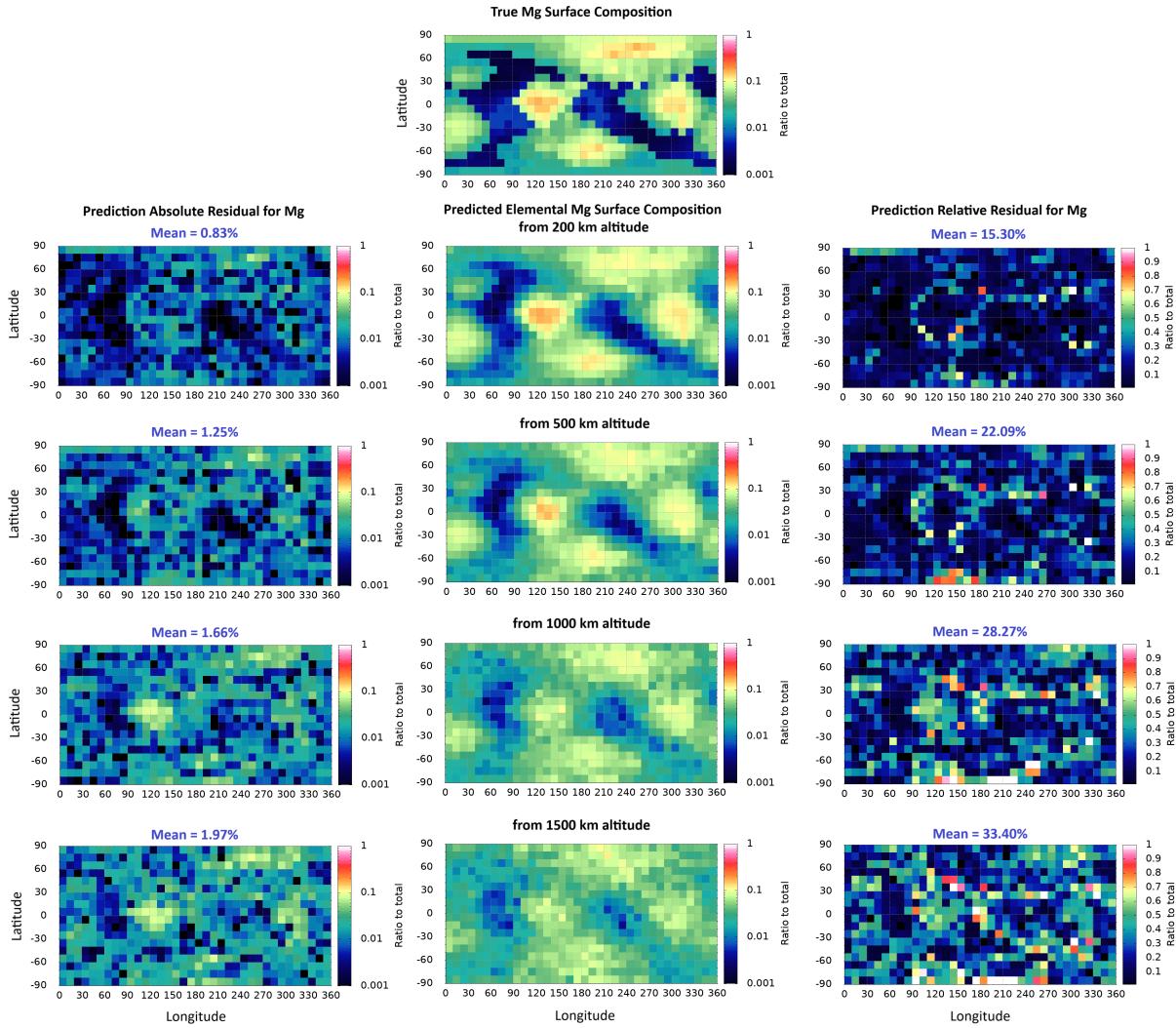


Figure C5. Main test campaign - MLP DNN map reconstructions of the same sample Magnesium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

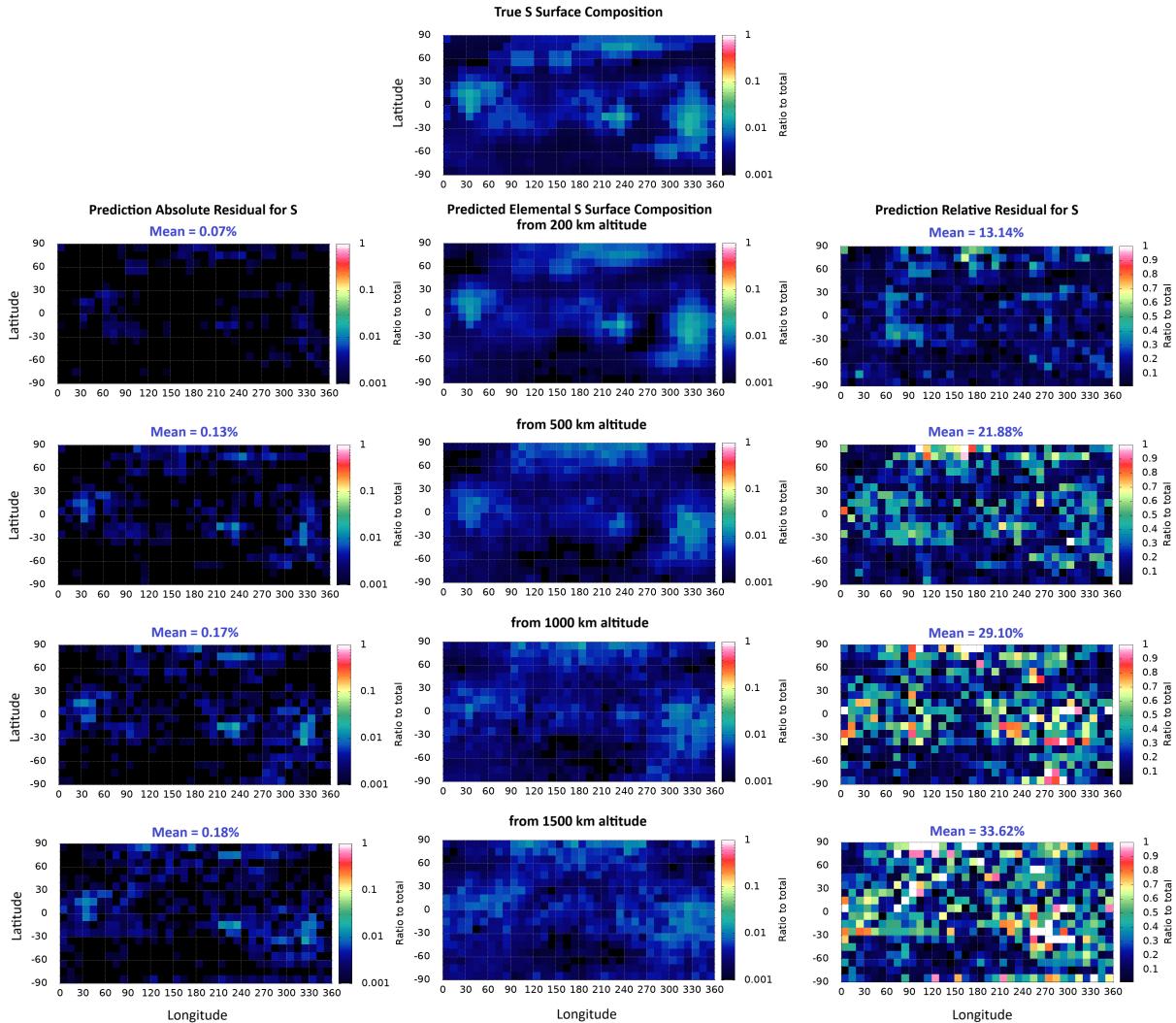


Figure C6. Main test campaign - MLP DNN map reconstructions of the same sample Sulfur surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

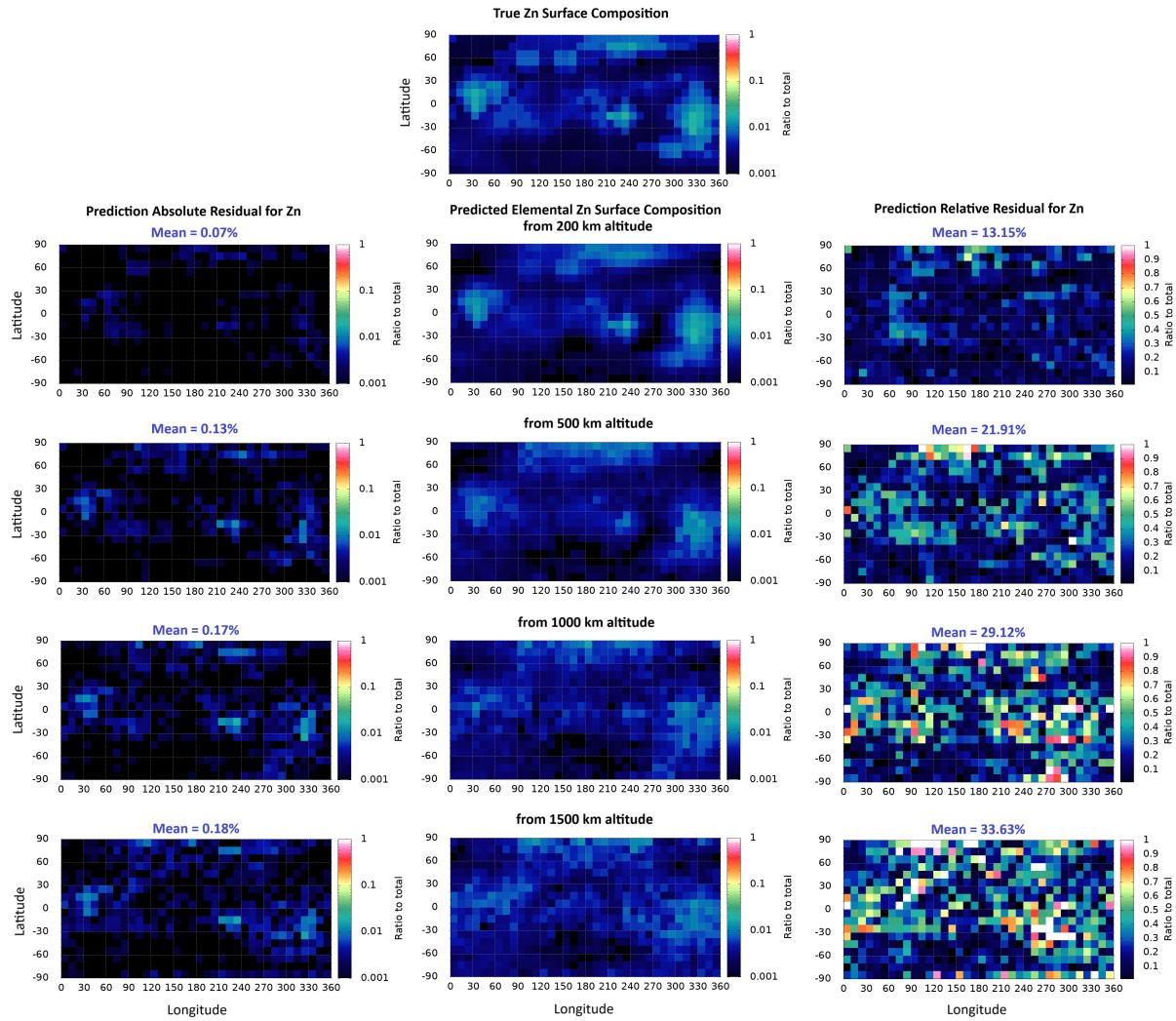


Figure C7. Main test campaign - MLP DNN map reconstructions of the same sample Zinc surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "ground 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.

APPENDIX C BOX PLOTS AND MAP RECONSTRUCTIONS OF THE MAIN TEST CAMPAIGN PREDICTIONS BY SPECIES

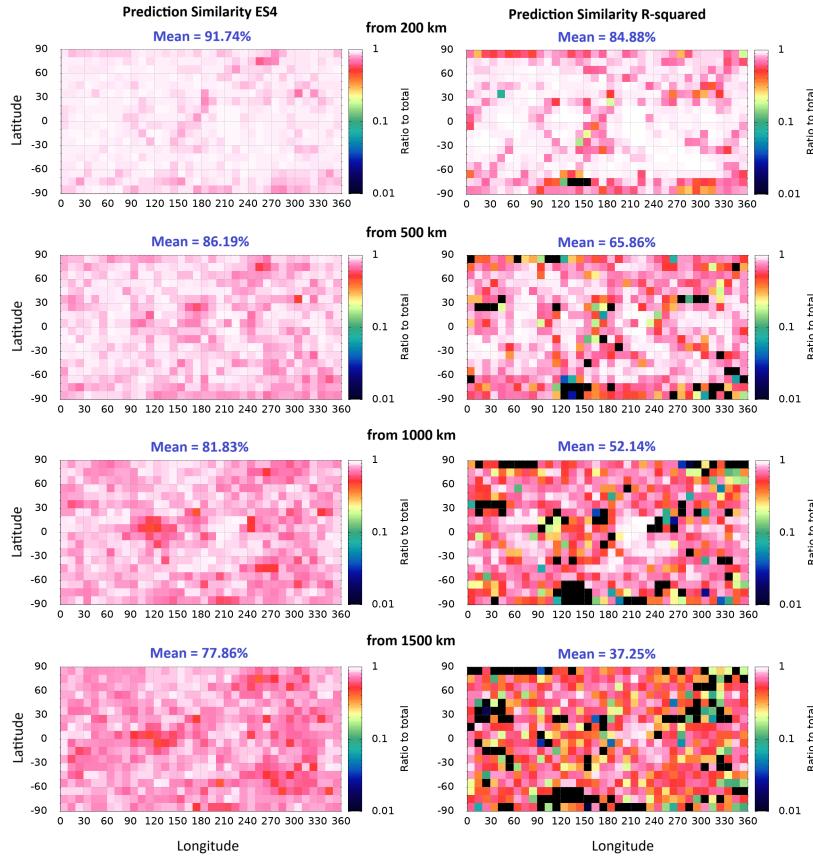


Figure C8. Map representations of the accuracy metrics evaluated for the predictions by the finalized trained MLP DNN of the same sample total surface composition (test set number 2). Dayside only inputs of two simulated exospheres from two consecutive Mercury perihelia at different altitude levels (200,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

- 1122 **References**
- 1123 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis,
1124 A. and Dean, J., Devin, M., et al. (2015). Tensorflow: Large-scale machine learning on
1125 heterogeneous systems. *Software available from tensorflow.org*.
- 1126 Bengio, Y. (2015). Early inference in energy-based models approximates back-propagation.
1127 Technical Report arXiv:1510.02777, Université de Montréal.
- 1128 Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M., Ferri,
1129 P., Middleton, H. R., and Ziethe, R. (2010). Bepicolombo - comprehensive exploration of
1130 mercury: Mission overview and science goals. *Planetary and Space Science*, 58(1):2–20.
- 1131 Berezhnoy, A. (2018). Chemistry of impact events on mercury. *Icarus*, 300:210–222.
- 1132 Bergstra, J., Bardenet, R., Bengio, Y., and Balazs, K. (2011). Algorithms for hyper-parameter
1133 optimization. In *Proceedings of the 24th International Conference on Neural Information
1134 Processing Systems*, pages 2546–2554.
- 1135 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 1136 Cassidy, T. A., Merkel, A. W., Burger, M. H., Sarantos, M., Killen, R. M., McClintock,
1137 W. E., and Vervack, R. J. (2015). Mercury’s seasonal sodium exosphere: Messenger orbital
1138 observations. *Icarus*, 248:547–559.
- 1139 Chollet, F. et al. (2015). Keras. *Github. GitHub repository*.
- 1140 Cintala, M. J. (1992). Impact-induced thermal effects in the lunar and mercurian regoliths. *J.
1141 Geophys. Res.*, 97:947–973.
- 1142 Ciresan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep big simple
1143 neural nets for handwritten digit recognition. *Neural Computation*, 22:1–14.
- 1144 Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-Interscience.
- 1145 Domingue, D. L., Koehn, P. L., Killen, R. M., et al. (2007). Mercury’s atmosphere: A surface-
1146 bounded exosphere. *Space Science Reviews*, 131(1-4):161–186.
- 1147 E. Vander Kaaden, K., M. McCubbin, F., R. Nittler, L., N. Peplowski, P., Z. Weider, S., A.
1148 Frank, E., and J. McCoy, T. (2017). Geochemistry, mineralogy, and petrology of boninitic
1149 and komatiitic rocks on the mercurian surface: Insights into the mercurian mantle. *Icarus*,
1150 285:155–168.

REFERENCES

REFERENCES

- 1151 Gamborino, D., Vorburger, A., and Wurz, P. (2019). Mercury's subsolar sodium exosphere:
1152 An ab initio calculation to interpret mascs/uvvs observations from messenger. *Annales
1153 Geophysicae*, 37:455–470.
- 1154 Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In
1155 *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*,
1156 volume 15, pages 315–323.
- 1157 Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge,
1158 MA.
- 1159 Grava, C., Killen, R. M., Benna, M., Berezhnoy, A. A., Halekas, J. S., Leblanc, F., Nishino,
1160 M. N., Plainaki, C., Raines, J. M., Sarantos, M., Teolis, B. D., Tucker, O. J., Vervack, R. J.,
1161 and Vorburger, A. (2021). Volatiles and refractories in surface-bounded exospheres in the
1162 inner solar system. *Space Science Reviews*, 217(5):61.
- 1163 Head, T. et al. (2018). scikit-optimize/scikit-optimize: v0.5.2. Zenodo.
- 1164 Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*,
1165 11(10):428–434.
- 1166 Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory
1167 and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.
- 1168 Johnson, R., Leblanc, F., Yakshinskiy, B., and Madey, T. (2002). Energy distributions for
1169 desorption of sodium and potassium from ice: the na/k ratio at europa. *Icarus*, 156:136–
1170 142.
- 1171 Kazakov, A. N. et al. (2020). Deep neural networks for analysis of mercury's planetary exosphere.
1172 *Journal of Physics: Conference Series*, 1548:012–014.
- 1173 Killen, R. (2016). Pathways for energization of ca in mercury's exosphere. *Icarus*, 268:32–36.
- 1174 Killen, R. M. and Burger, M. H. (2019). Understanding mercury's exosphere: Models derived
1175 from messenger observations. In *Mercury: The View After MESSENGER*. Cambridge
1176 University Press, London.
- 1177 Killen, R. M., Cremonese, G., Lammer, H., et al. (2007). Processes that promote and deplete
1178 the exosphere of mercury. *Space Science Reviews*, 132(2-4):433–509.
- 1179 Killen, R. M., Potter, A. E., Reiff, P., Sarantos, M., Jackson, B. V., Hick, P., and Giles, B.
1180 (2001). Evidence for space weather at mercury. *J. Geophys. Res.*, 106:20509–20525.

REFERENCES

REFERENCES

- 1181 Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. In *International*
1182 *Conference on Learning Representations*.
- 1183 Leblanc, F., Sarantos, M., Domingue, D., Milillo, A., Savin, D., Prem, P., Benkhoff, J., Zender,
1184 J., Galli, A., Murakami, G., Sasaki, S., Thompson, M., and Raines, J. (2023). How does
1185 the thermal environment affect the exosphere/surface interface at mercury? *The Planetary*
1186 *Science Journal*, 4:227.
- 1187 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.
- 1188 Mangano, V. et al. (2015). Themis na exosphere observations of mercury and their correlation
1189 with in-situ magnetic field measurements by messenger. *Planetary and Space Science*, 115:102–
1190 109.
- 1191 Massetti, S., Orsini, S., Milillo, A., Mura, A., De Angelis, E., Lammer, H., and Wurz, P. (2003).
1192 Mapping of the cusp plasma precipitation on the surface of mercury. *Icarus*, 166:229–237.
- 1193 Merkel, A. W., Vervack, R. J. J., Killen, R. M., Cassidy, T. A., McClintonck, W. E., Nittler,
1194 L. R., and Burger, M. H. (2018). Evidence connecting mercury’s magnesium exosphere to its
1195 magnesium-rich surface terrane. *Geophysical Research Letters*, 45:6709–6797.
- 1196 Milillo, A., Fujimoto, M., Kallio, E., Kameda, S., Leblanc, F., Narita, Y., Cremonese, G., Laakso,
1197 H., Laurenza, M., Massetti, S., McKenna-Lawlor, S., Mura, A., Nakamura, R., Omura, Y.,
1198 Rothery, D., Seki, K., Storini, M., Wurz, P., Baumjohann, W., and Sprague, A. (2010).
1199 The bepicolombo mission: An outstanding tool for investigating the hermean environment.
1200 *Planetary and Space Science*, 58:40–60.
- 1201 Milillo, A., Fujimoto, M., Murakami, G., et al. (2020). Investigating mercury’s environment
1202 with the two-spacecraft bepicolombo mission. *Space Science Reviews*, 216. Article no. 93.
- 1203 Milillo, A., Sarantos, M., Grava, C., Janches, D., Lammer, H., Leblanc, F., Schorghofer, N.,
1204 Wurz, P., Teolis, B., and Murakami, G. (2023). Future directions for the investigation of
1205 surface-bounded exospheres in the inner solar system. *Space Science Reviews*, 219.
- 1206 Milillo, A. and Wurz, P. (2014). *SERENA Science Performance Report*.
- 1207 Milillo, A., Wurz, P., Orsini, S., Delcourt, D., Kallio, E., Killen, R., Lammer, H., Massetti,
1208 S., Mura, A., Barabash, S., Cremonese, G., Daglis, I., De Angelis, E., Di Lellis, A., Livi, S.,
1209 Mangano, V., and Torkar, K. (2005). Surface-exosphere-magnetosphere system of mercury.
1210 *Space Science Reviews*, 117:397–443.

REFERENCES

REFERENCES

- 1211 Minsky, M. and Papert, S. A. (2017). *Perceptrons: An Introduction to Computational Geometry*.
1212 MIT Press, Cambridge, MA.
- 1213 Moroni, M., Mura, A., Milillo, A., Plainaki, C., Mangano, V., Alberti, T., Andre, N., Aronica,
1214 A., De Angelis, E., Del Moro, D., Kazakov, A., Massetti, S., Orsini, S., Rispoli, R., and
1215 Sordini, R. (2023). Micro-meteoroids impact vaporization as source for ca and cao exosphere
1216 along mercury's orbit. *Icarus*, 401:115616.
- 1217 Mura, A., Milillo, A., Orsini, S., and Massetti, S. (2007). Numerical and analytical model of
1218 mercury's exosphere: Dependence on surface and external conditions. *Planetary and Space
1219 Science*, 55:1569–1583.
- 1220 Mura, A., Orsini, S., Milillo, A., Delcourt, D., Massetti, S., and De Angelis, E. (2005). Dayside
1221 h+ circulation at mercury and neutral particle emission. *Icarus*, 175:305–319.
- 1222 Mura, A., Wurz, P., Lichtenegger, H., Schleicher, H., Lammer, H., Delcourt, D., Milillo, A.,
1223 Orsini, S., Massetti, S., and Khodachenko, M. (2009). The sodium exosphere of mercury:
1224 Comparison between observations during mercury's transit and model results. *Icarus*, 200:1–
1225 11.
- 1226 Orsini, S., Livi, S., Lichtenegger, H., et al. (2020). Serena: Particle instrument suite for sun-
1227 mercury interaction insights on-board bepicolombo. *Space Science Reviews*. In Press.
- 1228 Orsini, S., Livi, S., Torkar, K., Barabash, S., Milillo, A., Wurz, P., Di Lellis, A. M., and Kallio,
1229 E. (2010). Serena: A suite of four instruments (elena, strofio, picam and mipa) on board
1230 bepicolombo-mpo for particle detection in the hermean environment. *Planetary and Space
1231 Science*, 58:166–181.
- 1232 Plainaki, C., Mura, A., Milillo, A., Orsini, S., Livi, S., Mangano, V., Massetti, S., Rispoli, R.,
1233 and De Angelis, E. (2017). Investigation of the possible effects of comet encke's meteoroid
1234 stream on the ca exosphere of mercury. *Journal of Geophysical Research: Planets*, 122:1217–
1235 1226.
- 1236 Pokorný, P., Sarantos, M., and Janches, D. (2017). Reconciling the dawn–dusk asymmetry
1237 in mercury's exosphere with the micrometeoroid impact directionality. *The Astrophysical
1238 Journal Letters*, 842(2):L17.
- 1239 Pokorný, P., Sarantos, M., and Janches, D. (2018). A comprehensive model of the meteoroid

REFERENCES

REFERENCES

- 1240 environment around mercury. *The Astrophysical Journal*, 863:31. 31pp, Received 2018 April
1241 3; revised 2018 June 26; accepted 2018 June 27; published 2018 August 8.
- 1242 Rothery, D. A., Massironi, M., Alemanno, G., et al. (2020). Rationale for bepicolombo studies
1243 of mercury's surface and composition. *Space Science Reviews*, 216:66.
- 1244 Rumelhart, D., Hinton, G., and Williams, R. (1986a). Learning representations by back-
1245 propagating errors. *Nature*, 323:533–536.
- 1246 Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning internal representations
1247 by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed
1248 Processing*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge.
- 1249 Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, 3rd edition*.
1250 Prentice Hall Press, USA.
- 1251 Sarantos, M., Slavin, J., Benna, M., Boardsen, S., Killen, R., Schriver, D., and Trávníček, P.
1252 (2009). Sodium-ion pickup observed above the magnetopause during messenger's first mercury
1253 flyby: Constraints on neutral exospheric models. *Geophysical Research Letters*, 36.
- 1254 Wilson III, L. B., Stevens, M. L., Kasper, J. C., Klein, K. G., Maruca, B. A., Bale, S. D.,
1255 Bowen, T. A., Pulupa, M. P., and Salem, C. S. (2018). The statistical properties of solar wind
1256 temperature parameters near 1 au. *The Astrophysical Journal Supplement Series*, 236(2):41.
- 1257 Wurz, P. and Lammer, H. (2003). Monte-carlo simulation of mercury's exosphere. *Icarus*,
1258 164:1–13.
- 1259 Wurz, P., Whitby, J. A., Rohner, U., Martin-Fernandez, J. A., Lammer, H., and Kolb, C. (2010).
1260 Self-consistent modelling of mercury's exosphere by sputtering, micrometeorite impact and
1261 photon-stimulated desorption. *Planetary and Space Science*, 58:1599–1616.