

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

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

<sup>11</sup> <sup>a</sup> INAF-IAPS, via Fosso del Cavaliere 100, 00133, Rome, Italy

<sup>12</sup> <sup>b</sup> INAF-Osservatorio Astronomico di Trieste, via Giambattista Tiepolo 11, 34143, Trieste, Italy

<sup>13</sup> <sup>c</sup> Augmented Intelligence Lab, Rua Lugo 2, 36470, Salceda de Caselas, Spain

<sup>14</sup> <sup>d</sup> INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122, Padova, Italy

<sup>15</sup> <sup>e</sup> ASI - Italian Space Agency, Via del Politecnico, 00133, Rome, Italy

<sup>16</sup> E-mail: [adrian.kazakov@inaf.it](mailto:adrian.kazakov@inaf.it)

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

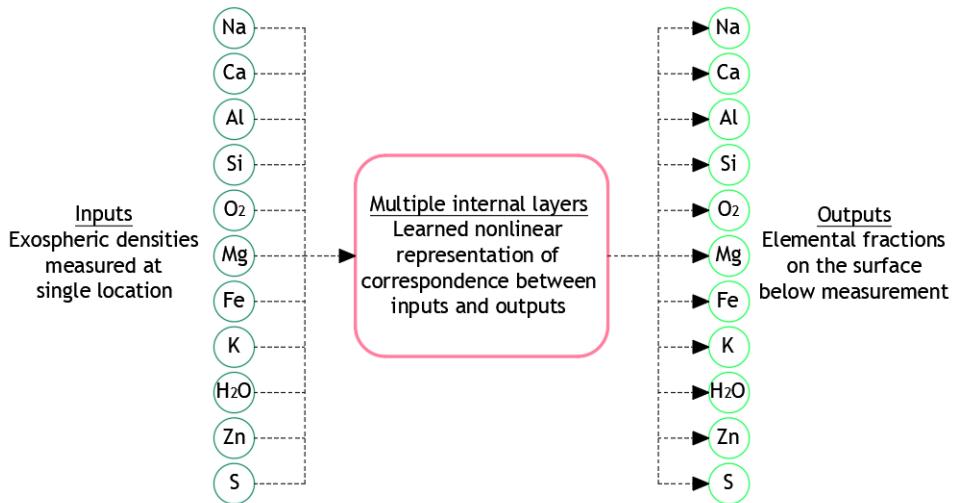
<sup>119</sup> **2. Method**

<sup>120</sup> *2.1. Prediction Task and General Characteristics of the Method*

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

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

<sup>126</sup> where  $\hat{y}_i$  is the fraction of an elemental species predicted by the neural network to be present  
<sup>127</sup> in the surface area below the exospheric measurement, and  $n$  is the total number of elements in  
<sup>128</sup> 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.

<sup>129</sup> The methodological and algorithmic developments in this study include:

- <sup>130</sup> • 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 more sophisticated and realistic data production model with plausible  
134 assumptions for exosphere generation processes.
- 135 • Defining input features to enhance algorithmic inference, accommodating variations in  
136 altitude and illumination conditions.
- 137 • Training the algorithms with augmented datasets.
- 138 • Employing hyperparameter tuning to optimize the DNN design parameters.
- 139 • Conducting a comprehensive exploration of the DNN hyperparameter space.
- 140 • Investigating the formation of the physical processes parameter space.

141 Ultimately, the goal of the MLP DNN is to encapsulate the complex relationships between  
142 various surface processes and their impact on the generation of the exosphere. Through rigorous  
143 training, the network is finely tuned to accurately reflect the dynamics between the surface and  
144 the exosphere. Post-training, the algorithm should be capable of:

- 145 • Capturing the intricate relationships between inputs (exospheric densities) and outputs  
146 (surface elemental fractions).
- 147 • Reverse engineering the exospheric generation processes, encompassing process yields,  
148 chemistry, energy, and distribution characteristics.
- 149 • Predicting the elemental composition of surface tiles.
- 150 • Reconstructing a comprehensive elemental fraction map of Mercury's surface.
- 151 • Formulating an estimator for surface processes contributing to the exosphere.

152 *2.2. Deep Neural Network Architecture*

153 The MLP, a class of feedforward neural network, excels in multivariate regression by modeling  
154 complex nonlinear functions with its multi-layered, fully connected structure and nonlinear  
155 activation functions (Minsky and Papert, 2017; Rumelhart et al., 1986a; Kingma and Ba,  
156 2014; LeCun et al., 2015). This architecture, combined with optimization techniques like  
157 backpropagation, allows MLPs to identify intricate patterns in high-dimensional data, making  
158 them ideal for robust predictive modeling.

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

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

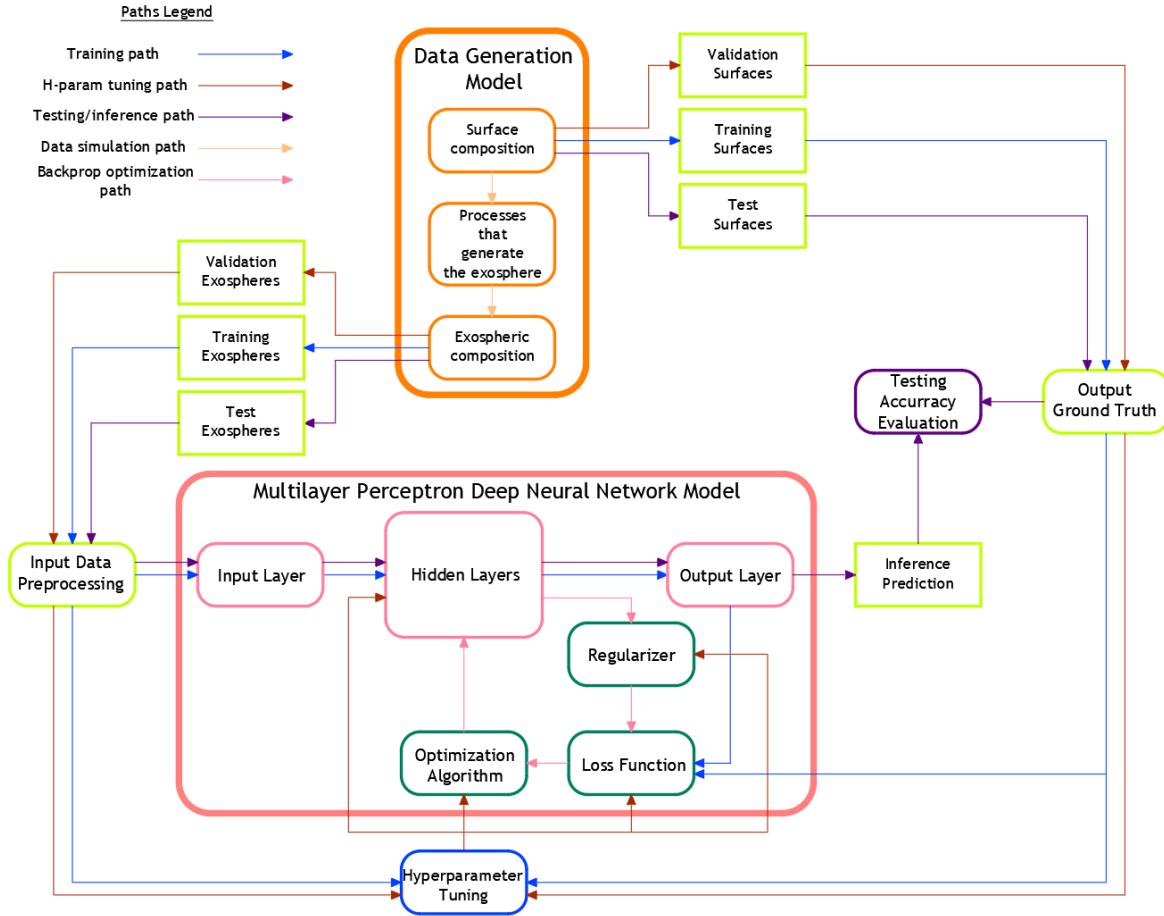
172 *Input Layer*

173 The input layer of the MLP introduces data, in our case Mercury's exospheric density  
 174 measurements, into the network, with each neuron representing a distinct data feature. For  
 175 example, distinct elemental density measurements are represented by separate neurons. Before  
 176 entry, data undergo normalization to ensure uniform influence on the learning process, thereby  
 177 preventing bias (Goodfellow et al., 2016). This involves standardizing each feature to zero mean  
 178 and unit variance as per the equation:

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

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

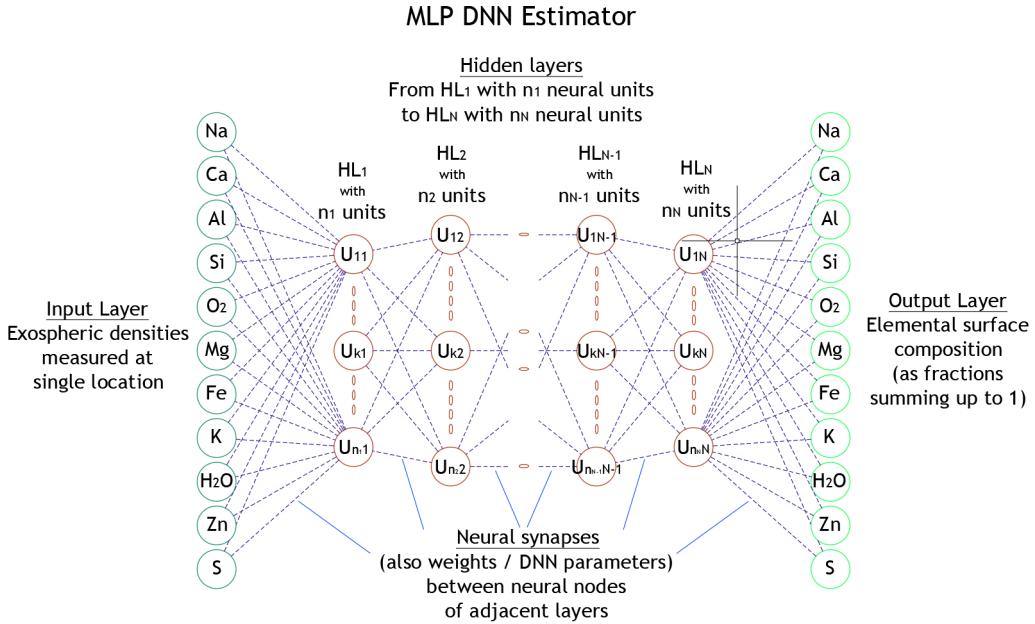
182 Expanding the scope of input features to include contextual information like measurement  
 183 altitude and sun-illumination angle enhances the MLP's predictive accuracy by accounting for  
 184 environmental conditions. These adjustments, detailed in Section 3, optimize data processing  
 185 and distribution to subsequent layers, enabling the MLP to efficiently process and more closely  
 186 represent complex interactions within Mercury's exosphere.



**Figure 2.** MLP DNN architecture overview. The data generation model produces both the inputs and the outputs for training, validation 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.

#### 187      *Hidden Layers*

188      The hidden layers form the core of the MLP architecture, where the actual processing and  
 189      learning occur (Minsky and Papert, 2017; Hinton, 2007). Positioned between the input and  
 190      output layers, they transform input data into a form usable for predictions. Each hidden layer is  
 191      composed of a set of neural units - neurons - and each neuron in these layers is fully connected  
 192      to all neurons in the preceding and succeeding layers, creating a dense network of synaptic  
 193      connections. The structure of these layers is represented mathematically by combination



**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  $W_{N+1}$ .

194 matrices or weight matrices, which, along with the activation function applied at each neuron,  
 195 helps form an estimation of the relationships among the processes acting between the input layer  
 196 and the output layer.

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

203 The transformation within each hidden layer then follows the equation:

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

204 where  $\mathbf{W}^T$  represents the weight matrix,  $\mathbf{x}$  is the input vector to the hidden layer (input

205 features or activations from a previous hidden layer), and  $\mathbf{b}$  is the bias vector of the affine  
 206 transformation. This equation encapsulates the affine transformation followed by the application  
 207 of the ReLU activation function, enabling the network to learn and represent complex nonlinear  
 208 relationships.

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

214 *Output Layer*

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

220 The activation function used in the output layer is crucial and depends on the nature of the  
 221 prediction task. For regression tasks like ours, where the output is a set of continuous values  
 222 that sum to 1 (representing fractions), the softmax function is often used (Joachims, 2002).  
 223 The softmax function converts the raw output of the network into a probability distribution,  
 224 ensuring that the predicted fractions are non-negative and sum up to one, aligning perfectly  
 225 with the physical reality of our task.

226 The formula for the softmax function is as follows:

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

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

231 *Loss Function*

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

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

where  $P$  represents the true distribution of the fraction of element  $i$  in the data, and  $Q$  is the predicted distribution from the MLP.

Other metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE) were considered but are less suited for probabilistic outputs as they do not account for the distribution's nature. MAE measures the average magnitude of errors directly, and MSE, sensitive to outliers, emphasizes larger errors. However, the KL divergence is preferred for our regression task because it aligns better with the probabilistic requirements, focusing on relative proportions rather than absolute quantities of elements.

#### 247 Regularizer

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

The L2 regularization is mathematically represented as:

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

where  $J(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y})$  is the original loss function,  $\lambda$  is the regularization coefficient, and  $\theta$  denotes the vector of all weight parameters (unfolded from the  $\mathbf{W}$  matrices). The right choice of  $\lambda$  is critical. If  $\lambda$  is too large, it can lead to underfitting, where the model is overly simplified

and fails to capture the underlying trends in the data. Conversely, a very small  $\lambda$  might not effectively prevent overfitting.

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

#### 265      *Optimization (Learning) Algorithm*

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

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

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

In our MLP model, Adam’s proficiency in managing sparse gradients and large parameter spaces, and dynamically adjusting the learning rate is invaluable, especially given the complexity of predicting surface elemental composition. The optimizer’s parameters, including the initial learning rate, beta values for the moment estimates, and epsilon (a small number to avoid division by zero), are selected based on empirical evidence and fine-tuned through experimentation. Adam’s ability to swiftly converge and handle non-stationary objectives makes it an excellent choice for this multivariate regression task, improving model performance and reliability.

284     *Hyperparameter Tuning*

285     While the core architecture of the neural network lays the foundation for predicting Mercury's  
286     surface composition, the model's overall efficacy and robustness are provided through the  
287     meticulous tuning of its hyperparameters. Key hyperparameters include:

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

300     To fine-tune these hyperparameters, we employed a Bayesian optimization strategy using the  
301     Gaussian Process (GP) approach, as outlined in Bergstra et al. (2011). The tuning process was  
302     facilitated by the scikit-optimize library (Head et al., 2018), which utilizes a prior probability  
303     distribution function to identify the hyperparameter configuration that minimizes the total loss  
304     on a cross-validation dataset.

305     This systematic adjustment of parameters such as learning rate, network architecture, and  
306     regularization coefficients via Bayesian optimization not only enhances learning capabilities and  
307     overall performance but also optimizes the balance between model complexity and efficiency.  
308     This approach ensures our MLP is well-suited for representing the intricate relationships of  
309     Mercury's surface-exosphere interaction.

310     *Performance Metrics*

311     To evaluate the performance of our machine learning model, we utilize both customized  
312     and standard metrics to ensure precise and insightful quantitative assessments. Our primary

metric, the Euclidean similarity 4 (ES4), integrates elements of Euclidean distance and cosine similarity, providing a nuanced measure of prediction accuracy by considering both magnitude and directionality in multidimensional space:

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

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

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

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

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

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

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

### 3. Exospheric Data

With the BepiColombo mission's nominal operations set to commence in 2026, the anticipated data from the SERENA instrument remains unavailable for our current analysis. Consequently, to apply our methods for reverse engineering a model of the exosphere in order to reconstruct

338 Mercury's surface composition, we have undertaken the task of generating our own datasets  
339 through simulations based on this model.

340 This section delves into our comprehensive approach to modeling and generating simulated  
341 exospheric data. We provide an in-depth explanation of the data generation model, outline the  
342 various processes influencing the exosphere, and discuss the methodologies we have implemented  
343 for creating and refining the datasets that are integral to the functionality of our deep learning  
344 algorithm. Our focus here extends beyond the inherent complexities of Mercury's exosphere to  
345 include a thorough description of the region of the physical parameter space we have used to  
346 represent the actual processes behind exospheric formation. This is pivotal for understanding the  
347 nature of the surface-environment-exosphere relations that our deep neural network is trained  
348 to estimate, setting the foundation for further exploration of this parameter space.

349 *3.1. Data Generation Model*

350 A variety of environmental models have been used historically to simulate the exosphere,  
351 especially data obtained during the measurements of orbiting spacecraft (Sarantos et al., 2009;  
352 Cassidy et al., 2015; Plainaki et al., 2017; Mura et al., 2007; Mangano et al., 2015; Wurz and  
353 Lammer, 2003). Our approach inversely reconstructs the processes that create the exosphere,  
354 using a specific model to train our deep neural networks for predicting surface compositions.

355 Modeling the interaction between planetary surfaces and exospheres encompasses a wide  
356 range of physical processes relevant to it. Central to every model is the multidimensional  
357 physical processes parameter space, which encapsulates the variables and constants necessary  
358 to simulate these intricate processes. The parameter space is an important element in our  
359 simulation approach, allowing us to explore and represent a variety of planetary environment  
360 conditions. It can be further subdivided to two subspaces, resulting from two subsets of  
361 parameters. The parameter subspace constructed from those physical interactions which are  
362 explicitly or implicitly considered in our model is captured in the vector  $P_{rep}$  (or set  $\mathbb{P}_{rep}$  - the  
363 subset of represented parameters - a subset of  $\mathbb{P}$  - the full set of physical parameters). Meanwhile,  
364 those aspects omitted due to lack of computational resources, understanding, or other factors,  
365 are captured in the vector  $\overline{P_{rep}}$  (or subset  $\neg\mathbb{P}_{rep}$ , complement of  $\mathbb{P}_{rep}$  to  $\mathbb{P}$ ).

366 Our method is designed to be adaptable, not limited to a single model but capable  
367 of reconstructing various exosphere generation models and predicting surface compositions

368 accordingly. The models successfully trained on our generated datasets can be termed 'base  
369 nodes.' These nodes define the constraints of the representation of the real exospheric generation  
370 processes within the physical parameter space.

371 Our study primarily revolves around one such 'base node' model and its parameter space.  
372 In this context, the data generation model, utilizing an exospheric simulation by Mura et al.  
373 (2007), assumes an important role in simulating conditions within Mercury's exosphere. While  
374 the model itself is not the central focus of our investigation, it is crucial for generating the  
375 datasets that underpin our analytical algorithms. This model comprehensively describes the  
376 physical processes shaping the exosphere, allowing us to define specific regions in the parameter  
377 space. These regions govern the data distribution our multilayer perceptron deep neural network  
378 aims to estimate.

379 It's important to recognize that the selected representation in the parameter space is not  
380 of single points, but rather of a region. This conceptual shift acknowledges that real-world  
381 conditions are dynamic, encompassing a spectrum of values rather than static points, and aims  
382 to capture a more realistic range of exospheric scenarios. Therefore, the simulation model  
383 generates data by simulating a distribution of points around central values in this parameter  
384 space. This methodology reflects the inherent variability and uncertainties characteristic of  
385 physical processes.

386 The objective of this work is to show that our trained MLP DNN is able to closely approach  
387 the most representative point or region in this parameter space and present itself as an estimator  
388 of the data generation mechanism. This approach represents a fundamental stride in leveraging  
389 the capabilities of deep neural networks to navigate the complexities of exospheric data analysis.

### 390 3.1.1. Surface and Regolith

391 Modeling the surface of Mercury is an intricate task, involving complex relationships between  
392 its components and the environment. For our modeling, we need to consider some of the  
393 characteristics of the planetary surface and regolith (the loose, heterogeneous material covering  
394 solid rock), which are part of our simulation model. These include the influence of surface  
395 composition, texture, physical, chemical and thermal properties, all in the context of forming  
396 the modeling parameter space and defining a region within that space.

397 Firstly, the surface in our model is represented as a grid comprised of  $36 \times 18$  surface tiles in a

398 modified Mercator projection. Each tile measures  $10^\circ \times 10^\circ$ , which, at the equator, translates to  
399 approximately  $425 \text{ km} \times 425 \text{ km}$ . This averaging of composition inevitably reduces the complexity  
400 of the parameter space, as it results in the loss of finer details in the spatial relations of the  
401 spread of the different species - elemental and mineral - on the planetary surface. In addition  
402 to the surface grid sizes in longitude and latitude, other relationships are captured implicitly in  
403 the represented subset of parameter space.

404 The next subset of parameters for our simulation of Mercury's surface and regolith is derived  
405 from its mineralogical composition. We select a specific set of minerals, as detailed in Table 1,  
406 believed to be present on Mercury's surface (Wurz et al., 2010). These minerals are assumed  
407 to exist in varying proportions, contributing to the regolith's overall mineral composition. This  
408 selection process inherently defines another parameter subspace within our model, capturing  
409 some of the complex coexistence relationships among different minerals. These relationships  
410 delineate zones characterized by dominant primary minerals and their secondary counterparts,  
411 enforcing the presence of some of the minerals on the surface. Additionally, constraints on the  
412 minimal fractions of specific minerals and the presence of water ice further refine this parameter  
413 subspace, offering an implicit subset of complexity to our model. From the included minerals,  
414 hedenbergite (a mineral deposited primarily from meteorites), sphalerite (a mineral resulting  
415 from volcanic activity), and water ice are considered as rare minerals, and their presence  
416 is reduced by 80% in the random surface generation, compared to the other six minerals.  
417 Furthermore, in the randomized creation of the surface, they are not allowed to be distributed  
418 everywhere on the surface. The overall mean fractions of the minerals in our datasets are  
419 reported in Table 1, while a more detailed description for the different types of datasets is  
420 shown in the Appendix.

421 Once the mineralogical composition of each surface tile within the datasets is established, the  
422 next step involves determining the decomposition process of these minerals into their constituent  
423 elemental species (atoms or molecules). In our approach, we adopt the classical additive method  
424 for breaking down the regolith minerals into their elemental components (Wurz et al., 2010).  
425 This implies an assumption that the surface, on average, encompasses a complete pool of atoms  
426 and molecules derived from these minerals, which are then subjected to external environmental  
427 forces. This approach, as an approximation, considers the full fraction of volatile species (such as  
428 Na, K, H<sub>2</sub>O, S, and O<sub>2</sub>) as being readily available for thermal and photon-stimulated desorption

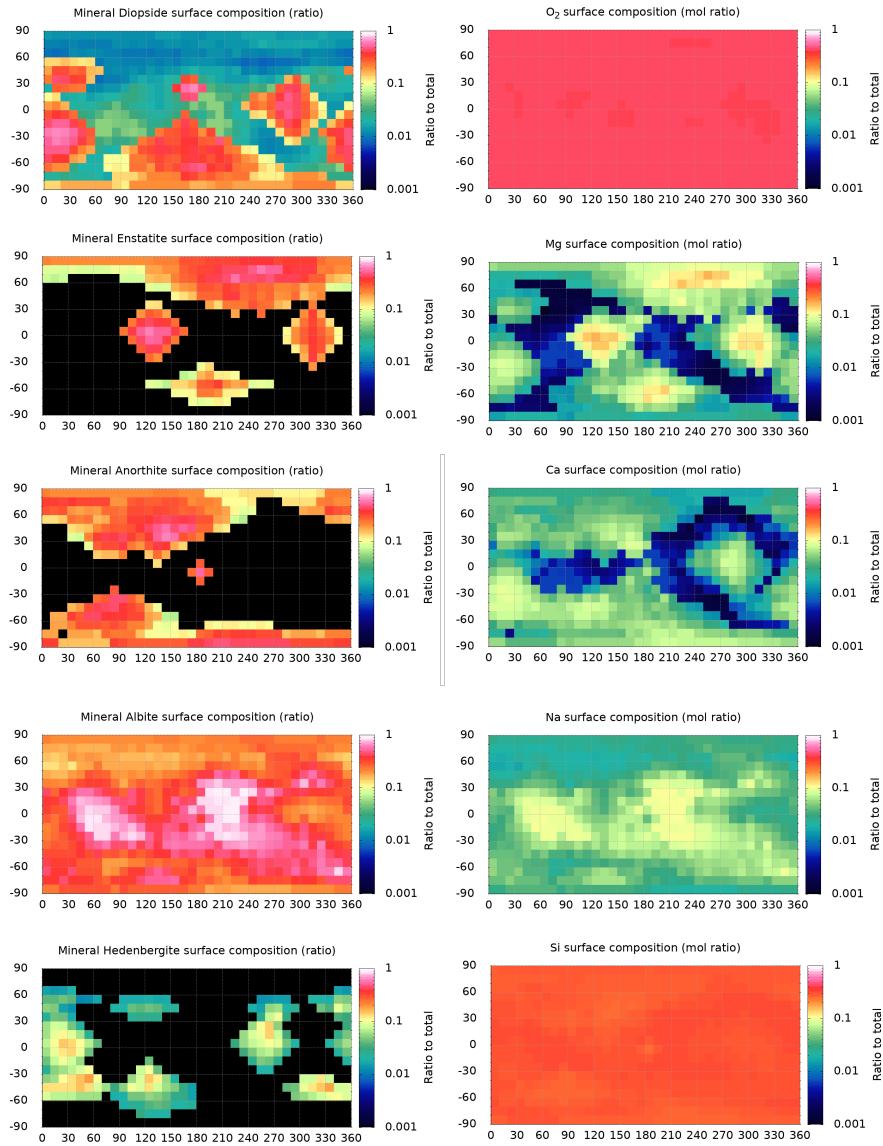
Mineralogy and Relationships (Baseline Model)				
Mineral Name	Chemical Formula	Decomposed to	Rarity	Mean Fraction
Anorthite	$\text{CaAl}_2\text{Si}_2\text{O}_8$	$\text{Ca}, 2\text{Al}, 2\text{Si}, 4\text{O}_2$	-	0.134
Albite	$\text{NaAlSi}_3\text{O}_8$	$\text{Na}, \text{Al}, 3\text{Si}, 4\text{O}_2$	-	0.140
Orthoclase	$\text{KAlSi}_3\text{O}_8$	$\text{K}, \text{Al}, 3\text{Si}, 4\text{O}_2$	-	0.134
Enstatite	$\text{Mg}_2\text{Si}_2\text{O}_6$	$2\text{Mg}, 2\text{Si}, 3\text{O}_2$	-	0.137
Diopside	$\text{MgCaSi}_2\text{O}_6$	$\text{Mg}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	-	0.141
Ferrosilite	$\text{Fe}_2\text{Si}_2\text{O}_6$	$2\text{Fe}, 2\text{Si}, 3\text{O}_2$	-	0.137
Hedenbergite	$\text{FeCaSi}_2\text{O}_6$	$\text{Fe}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	Rare	0.065
Sphalerite	ZnS	Zn, S	Rare	0.069
Water Ice	$\text{H}_2\text{O}$	$\text{H}_2\text{O}$	Rare	0.044

**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.

processes, as they are loosely bound to the regolith grains. The list of elements resulting from the mineral break down is reported in Table 2, and an example split of two mineral surface maps to elemental surface maps is shown on Figure 4. The elemental composition resulting from this process represents the 'actual' or 'ground truth' data that we compare the predictions to within our algorithms.

Elements and Relationships (Baseline Model)				
Element Name	Designation	From Mineral	Rarity	Mean Fraction
Aluminium	Al	Anorthite, Albite, Orthoclase	-	0.072
Calcium	Ca	Anorthite, Diopside, Hedenbergite	-	0.047
Iron	Fe	Ferrosilite, Hedenbergite	-	0.048
Sodium	Na	Albite	-	0.019
Oxygen	$\text{O}_2$	All, except Sphalerite, Water Ice	Dominant	0.423
Sulfur	S	Sphalerite	Rare	0.012
Water Vapor	$\text{H}_2\text{O}$	Water Ice	Rare	0.008
Zinc	Zn	Sphalerite	Rare	0.012
Silicium	Si	All, except Sphalerite, Water Ice	Dominant	0.282
Potassium	K	Orthoclase	-	0.018
Magnesium	Mg	Enstatite, Diopside	-	0.059

**Table 2.** Elemental composition considered in the baseline surface model. The elements are broken down from minerals in the classical additive method. 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).



**Figure 4.** From mineralogy to elemental composition. The figure depicts the resulting elemental maps for Calcium, Magnesium and Sodium from Albite, Anorthite, Enstatite, Diopside, and Hedenbergite. Shown are also the resulting Silicium and Oxygen maps for abundance comparison. **Improve figure by removing unnecessary duplication of coordinates, bars, etc.**

Conversely, while our model omits certain surface qualities—such as grain sizes, slope angles, and roughness—to focus on the critical aspects of mineral and elemental composition in simulating Mercury’s exosphere, we do incorporate a simplified representation of porosity and the presence of microshadows in the ion-sputtering process acting on the surface , defined as porosity and microshadows coefficients. This implies that parameters representing grain sizes, slope angles, and roughness do not contribute to the defined parameter space, while those

440 representing porosity and microshadows, and their interactions with other relevant processes,  
441 contribute to the defined parameter space in a simplified manner.

442 *3.1.2. Environmental Conditions*

443 References are missing...

444 In our model definition, we incorporate the environmental conditions and various factors that  
445 contribute to changes in the sources or processes for the release of material from the planetary  
446 surface into the exosphere. These sources encompass solar radiation, dust particles, and charged  
447 particles that enable surface material to escape into the exosphere. The environmental conditions  
448 around and on Mercury play a integral role in shaping these effects.

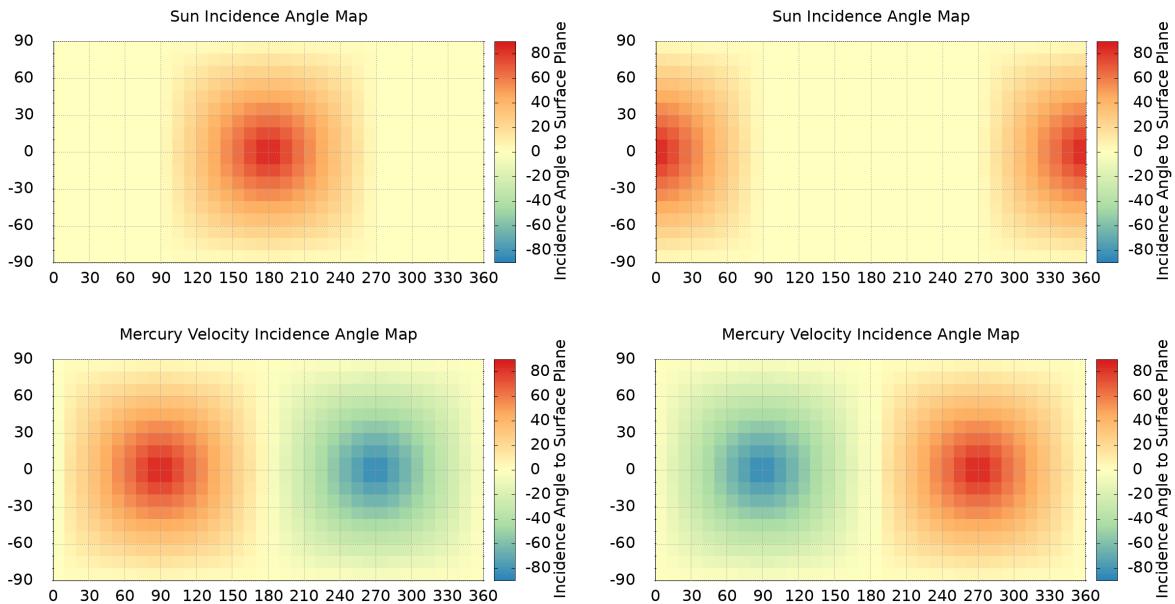
449 Mercury's proximity to the Sun significantly influences its interaction with the surrounding  
450 environment. Its notably eccentric orbit causes substantial variation in distance from the  
451 Sun between perihelion (closest approach) and aphelion (furthest distance). In our model, we  
452 specifically focus on conditions at perihelion, where Mercury is about 0.31 astronomical units  
453 (AU) or 46 million kilometers from the Sun. This close distance markedly impacts the intensity  
454 of the solar influence, thereby affecting the ranges of the effects and processes that contribute to  
455 generating the exosphere. One such influence is the equivalent photon flux, which is the photon  
456 flux at Earth's orbit adjusted to account for Mercury's closer position to the Sun by a factor of  
457  $1/r^2$ , where  $r$  is the distance to the Sun in AU. A photon flux at Earth of  $3.0 \times 10^{15} \text{ cm}^{-2}\text{s}^{-1}$   
458 is considered.

459 Another environmental aspect is the activity level of the Sun, which we have assumed to be  
460 at a moderate level, devoid of extreme events such as coronal mass ejections or solar flares. This  
461 assumption sets the conditions for solar wind velocity at 450 km/s and solar wind density at 60  
462  $\text{cm}^{-3}$  at Mercury's perihelion.

463 Furthermore, the dust environment around Mercury is considered for particles smaller than  
464  $100 \mu\text{m}$  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  
465 Mercury's vicinity in agreement with the modal impact velocity reported by Cintala (1992),  
466 compared to a planet velocity at perihelion of 59 km/s. However, our model does not differentiate  
467 between the origins of these dust particles—whether they come from the Main Belt Asteroids,  
468 Jupiter Family Comets, Oort Cloud Comets, or Halley Type Comets—nor does it consider  
469 the full ranges and exact distributions of particle sizes and velocities. Additionally, no large

470 meteorite impacts or increases of fluxes due to particularly dense cometary streams, such as  
 471 from comet Encke (cite someone), are considered. Grain size distribution influence of the dust  
 472 particles is also not represented in our physical parameter space.

473 The environmental conditions on the planet itself present a diverse range of parameters due  
 474 to varying exposure to sunlight and shadow, as well as differences in particle fluxes on the  
 475 planet's leading and trailing sides due to its high orbital velocity. Our model incorporates the  
 476 true anomaly angle (TAA) of Mercury's orbit around the Sun, producing **detailed** maps that  
 477 illustrate solar incidence angles and planetary velocity incidence angles at Mercury's perihelion  
 478 (Figure 5). It's crucial to recognize Mercury's unique orbit-spin resonance, which alternates the  
 479 sides facing the Sun at the same TAA in successive orbits, a fact that we have taken advantage  
 480 of later in our study.



**Figure 5.** Maps of incidence angles due to planet orientation at two consecutive perihelia. **Add more explanation... Possibly improve/update this figure.**

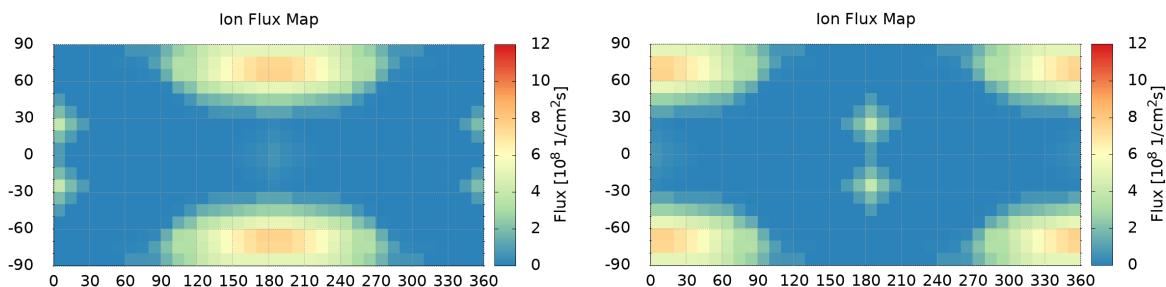
481 In extending our discussion on environmental conditions affecting Mercury, it's important to  
 482 clarify the limitations of our model, especially concerning magnetic field effects. Our study  
 483 does not encompass the full spectrum of influences that the interplanetary magnetic field  
 484 and Mercury's own magnetic field might have on the planet's surface and exosphere. This  
 485 omission includes a variety of magnetic field characteristics such as Mercury's dipole moment, the

486 thickness and strength of the Harris sheet, dipole shift, and other pertinent electric and magnetic  
 487 field parameters. Nor does it engage with complex magnetic field models that could offer a more  
 488 detailed understanding of these interactions. This simplification is considered satisfactory as the  
 489 algorithm will not be taught to learn a shift of the magnetic dipole dependence and dynamical  
 490 effects at this stage.

491 The sole aspect of Mercury's magnetic field that our model takes into account is the presence  
 492 of open magnetic field lines. These lines serve as conduits for charged particles, channeling them  
 493 through magnetic field cusps directly onto the planet's surface at specific locations known as  
 494 cusp footprints. The impact of this process is significant, as it concentrates ion bombardment in  
 495 particular areas, altering the surface composition and potentially influencing the generation of  
 496 the exosphere. We have delineated the shape and relative impact areas of these cusp footprints  
 497 in Figure 6, highlighting the regions on Mercury's surface that are most affected by the ion  
 498 funneling. These zones are another variable that is dynamically changing and in many cases in  
 499 very short timeframes. In our study, we have considered a fairly wide zones of ion precipitation.  
 500 A more dynamic dependence on the magnetic reconnection rate driven by the Interplanetary  
 501 Magnetic Field (IMF) strength and orientation is neglected.

502 **Add about hidden spatial dependence of adjacent surface tiles at various local times and  
 503 latitudes in mixing together populations ejected by sputtering?**

504 This approach allows us to incorporate a critical, though singular, aspect of magnetic field  
 505 influence into our simulation, while acknowledging the broader complexities of magnetic and  
 506 electric field effects that remain beyond the scope of our current model. **Or remain secondary.**



**Figure 6.** Maps of ion fluxes at the surface at two consecutive perihelia. **Add more explanation... Possibly improve/update this figure.**

507 A summary of the environmental conditions considered in our study and their contribution  
 22

508 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 <sup>3</sup>	60	Proton flux	SP
	Photon flux at Earth	1/cm <sup>2</sup> s	3.0 x 10 <sup>15</sup>	Equivalent photon flux	PSD
Comets and Asteroids	Dust particle size	μm	0-100	Micrometeorite flux	MIV
	Mean flux of dust particles	g/cm <sup>2</sup> s	1.0 x 10 <sup>-16</sup>	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.

### 509 3.1.3. Sources and Processes for Generation of the Exosphere

510 Let's now outline the specific processes simulated by our model that contribute to the  
 511 generation of the exosphere, compounded by the source actors that influence the surface to  
 512 eject its constituent materials into space. Each source and process is described in terms of its  
 513 physical basis, the parameters involved, and its relative importance in the overall generation of  
 514 the exosphere **under different scenarios**.

515 The active effects considered here are the four main surface release, namely the micrometeorite  
 516 impact vaporization (MIV), sputtering from protons and heavy ions (SP), thermal desorption  
 517 (TD) and photon-stimulated desorption (PSD). Their respective sources are the micrometeoroid  
 518 fluxes, the precipitating ions through the open field lines of Mercury's magnetic field, the  
 519 temperature effects on the surface, and the solar photons that impact the dayside surface.

520 However, there are quite a few existing gaps in the understanding of these processes, which  
 521 make the problem not fully constrained in terms of what is observed and what model parameters  
 522 correspond to the observations. We, nevertheless, have chosen a range for their values, which  
 523 correspond to some more widely agreed-on observations.

#### 524 *Micrometeorite fluxes and impact vaporization*

525 In our model, micrometeorites are represented as an incoming flux of interplanetary dust  
 526 that expels matter from the surface regolith in relation to the incoming flux and velocity, as  
 527 described by Cintala (1992). The distribution of this flux onto Mercury's surface is influenced

528 by the velocity of Mercury and its projection onto the surface area where the flux is calculated,  
 529 as illustrated in Figure 5. We employ a simple relationship between the angle of incidence and  
 530 the modification of the mean flux onto the surface, defined as:

$$\Delta\phi_{MIV} = \frac{V_{mm} \cos \beta_{Surf}}{V_M}, \quad (10)$$

531 where  $V_{mm}$  is the mean dust velocity,  $V_M$  is the velocity of Mercury and  $\beta_{Surf}$  is the angle  
 532 between Mercury's velocity vector and the surface normal vector. Consequently, the incoming  
 533 flux of dust particles varies between about  $0.7 \times 10^{-16}$  and about  $1.4 \times 10^{-16}$  gcm $^{-2}$ s $^{-1}$  on the  
 534 trailing and leading sides, respectively. At the chosen modal velocity of the incoming flux, a  
 535 constant vapor phase production rate of 5 is assumed, simplifying the parameter subspace for  
 536 this complex vaporization process to an extent deemed sufficient for our study. Notably, surface  
 537 temperature effects are not accounted for in this model of the MIV release process.

538 This approach yields outflows of surface matter ranging from  $3.5 \times 10^{-16}$  to  $7 \times 10^{-16}$   
 539 gcm $^{-2}$ s $^{-1}$ . These values are conservatively estimated to be about a factor of 2 smaller than  
 540 those suggested by Cintala (1992) and two orders of magnitude smaller than those proposed by  
 541 Pokorný et al. (2017), fitting within the parameter space interest to modelers without overly  
 542 emphasizing this omnipresent process. Additionally, this assumption poses a more challenging  
 543 scenario for the algorithm due to the inherent representability of the surface composition by  
 544 the exosphere generated due to the MIV. The vaporized species include larger molecules such  
 545 as CaO, NaOH, NaO, and others, resulting from the complex chemistry within the impact-  
 546 produced cloud (Killen, 2016; Berezhnoy, 2018). However, for our initial DNN analysis iteration,  
 547 we assume these species have very short photolysis lifetimes, quickly breaking down into their  
 548 constituent elements without further energization.

549 The model adopts Maxwellian distributions for the velocities of the excited particles, with  
 550 the vapor temperature averaging 4000 K as per (Wurz and Lammer, 2003).

MIV Source and Process Parameters				
Parameter	Class	Units	Value	Affected Species
Mean flux of dust particles	Source	g/cm $^2$ s	$1.0 \times 10^{-16}$	All
Mean velocity of dust particles	Source	km/s	20	All
Vapor phase production rate	Process	-	5	All
Vapor temperature	Process	K	4000	All

**Table 4.** Source and process parameters for the micrometeoroid impact vaporization.

551     *Ion precipitation and Ion sputtering (SP)*

552     The ion sputtering is initiated by a flux of bombarding ions, predominantly comprising solar  
 553     wind protons, which efficiently ejects atoms/molecules from the surface regolith (Wurz et al.,  
 554     2010; Killen et al., 2007). The ion flux's impact is localized to areas where the open magnetic  
 555     field lines intersect the surface.

556     In our model, the flux impacting Mercury's surface is assumed proportional to the solar  
 557     wind's unperturbed upstream flux of protons, represented as:

$$\phi = C \rho_{sw} v_{sw}, \quad (11)$$

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

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

$$\frac{d\Phi_n}{dE_e} = Y c \int_{E_{\min}}^{E_{\max}} \frac{d\Phi_I}{dE_i} f_S(E_e, E_i) dE_i, \quad (12)$$

563     where  $Y$  is the yield of the process,  $c$  the surface relative abundance of the species,  $\Phi_I$  the  
 564     ion flux,  $\Phi_n$  the neutral flux emitted from the surface,  $E_i$  the impact energy,  $E_e$  the energy of  
 565     the ejected particles, and  $f_S$  an empirical model for the energy distribution of ejected particles,  
 566     defined as:

$$f_S(E_e, T_m) = c_n \frac{E_e}{(E_e + E_b)^3} \times \left[ 1 - \left( \frac{E_e + E_b}{T_m} \right)^{1/2} \right], \quad (13)$$

567     with  $T_m$  as the maximum transmitted energy,  $c_n$  the normalization constant, and  $E_b$  the  
 568     surface binding energy of the ejected species.  $T_m$  is calculated as per Mura et al. (2007) as:

$$T_m = E_i \frac{4m_1 m_2}{(m_1 + m_2)^2}, \quad (14)$$

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

570     For this investigation, we assume a uniform yield efficiency of 0.1 for all species, a figure that  
 571     is relatively high compared to existing literature Schaible et al. (2017). However, we apply a

572 conservative reduction in yield to account for the regolith's porosity (a uniform factor of 0.35) and  
 573 the microshadows within it (a uniform factor of 0.40). This adjustment modestly diminishes the  
 574 sputtering effect, a deliberate choice to complicate the prediction of surface composition by DNN  
 575 algorithms in high-latitude regions receiving solar wind precipitation. The angular distribution  
 576 around the normal direction of the surface is taken as  $\cos^2(\alpha_n)$ .

SP Source and Process Parameters													
Parameter	Class	Units	Al	Ca	Mg	Na	K	Fe	Si	Zn	S	O <sub>2</sub>	H <sub>2</sub> O
Mean ion flux	Source	1/cm <sup>2</sup> s						1.08 x 10 <sup>9</sup>					
Yield efficiency	Process	-						0.1					
Impact energy	Process	eV						1000					
Porosity coefficient	Process	-						0.35					
Microshadows coefficient	Process	-						0.4					
Binding energy	Process	eV	3.36	2.1	1.54	2	0.93	4.34	4.7	1.35	2.88	2	0.5

**Table 5.** Source and process parameters for the ion sputtering.

577 *Temperature map and Thermal desorption (TD)*

578 Thermal desorption becomes notably efficient at temperatures above 400 K, a threshold where  
 579 the vibration of loosely bound atoms and molecules increases significantly (Mura et al., 2007).  
 580 In our model, TD primarily affects a specific population of molecules—namely Na, K, H<sub>2</sub>O,  
 581 S—which are loosely bound to Mercury's surface.

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

$$T_s(\phi, \theta) = T_{\min} + (T_{\max} - T_{\min})(\cos\phi\cos\theta)^{1/4}, \quad (15)$$

586 where  $\phi$  represents the latitude and  $\theta$  the longitude, as outlined in Mura et al. (2007).

587 In our approach, the TD process is assumed to occur without diffusion, focusing solely on  
 588 the direct thermal ejection of species from the surface. The flux of atoms or molecules resulting  
 589 from TD is calculated following the formula provided by Mura et al. (2007):

$$\Phi_n = \nu N c e^{-\left(\frac{U_d}{k_B T}\right)}, \quad (16)$$

590 where  $\nu$  denotes the vibrational frequency of the species,  $N$  the surface density of the regolith,

591  $c$  the fractional presence of the species within the regolith,  $U_d$  the species' binding energy,  $k_B$

592 the Boltzmann constant, and  $T$  the temperature at which desorption occurs. For the efflux  
 593 of particles from the surface, a Maxwellian-Boltzmann distribution is assumed, reflecting the  
 594 statistical nature of the thermal motion contributing to the desorption process.

TD Source and Process Parameters						
Parameter	Class	Units	Na	K	S	H <sub>2</sub> O
Dayside temperature	Source	K		700		
Nightside temperature	Source	K		110		
Surface density	Process	1/cm <sup>2</sup>		7.5 × 10 <sup>14</sup>		
Vibrational frequency	Process	1/s		1.0 × 10 <sup>13</sup>		
Binding energy	Process	eV	2	0.93	2.88	0.5

**Table 6.** Source and process parameters for the thermal desorption.

595 *Photon flux and Photon-stimulated desorption (PSD)*

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

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

$$\Phi_n = N_c \int \Phi_\gamma(E) \sigma(E) dE, \quad (17)$$

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

605 Moreover, the model adjusts for reduced flux at lower incidence angles away from the subsolar  
 606 point, employing the following relation:

$$\Phi_n(\phi, \lambda)^* = \Phi_n \cos(\phi) \cos(\lambda), \quad (18)$$

607 with  $\phi$  representing the longitude in local solar time and  $\lambda$  the latitude, thereby factoring in  
 608 the geometric reduction of flux due to the angle of solar incidence.

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

$$f(E) = \beta(1 + \beta) \frac{EU^\beta}{(E + U)^{2+\beta}}, \quad (19)$$

611 in which  $\beta$  is set to 1 for our study to represent an energy cut-off, and  $U$  denotes the threshold  
 612 energy.

613 What about PSD temperature above surface temperature and cross sections that are inputs  
 614 to Modello?

PSD Source and Process Parameters						
Parameter	Class	Units	Na	K	S	H <sub>2</sub> O
Mean photon flux	Source	1/cm <sup>2</sup> s			3.1 × 10 <sup>16</sup>	
beta coefficient	Process	-			1	
Temperature above regolith	Process	K	200	200	200	0
PSD cross section	Process	1/m <sup>2</sup>	1 × 10 <sup>-25</sup>	1 × 10 <sup>-25</sup>	1 × 10 <sup>-25</sup>	1 × 10 <sup>-22</sup>

**Table 7.** Source and process parameters for the photon-stimulated desorption.

### 615 3.1.4. Dynamics of the Exosphere

616 The dynamics of the exosphere, as simulated in our model, covers the movement and behavior  
 617 of particles after they have been released into the exosphere, including their interactions,  
 618 trajectories, and eventual fate. Factors such as gravitational influences, electromagnetic forces,  
 619 and collisions are examined to understand how they shape the structure and composition of the  
 620 exosphere.

### 621 Photo-ionization

622 Once in the exosphere, each elemental species is subject to a set of conditions that define  
 623 its mean lifetime before it undergoes photo-ionization. This process is primarily driven by the  
 624 intense solar radiation that permeates this planetary layer. The energy from solar radiation is  
 625 sufficient to strip electrons from the outer shell of atoms or molecules, effectively transforming  
 626 them into ions.

627 In our model, once ionization occurs, the recombination of the produced ion with an electron  
 628 is not considered. This assumption is based on the low density of particles in the exosphere,

which makes such recombination events exceedingly rare. Furthermore, the trajectory of these ions around the planet is not tracked post-ionization. As a result, this ionization process is depicted as a net loss to the exosphere.

The mean lifetimes due to photoionization of the elemental species in our exospheric models are listed in Table 8. Explain in more detail how the Mura model implements the photoionization, significant parameters involved, such as mean lifetime...

Exospheric Dynamics Parameters												
Parameter	Units	Al	Ca	Mg	Na	K	Fe	Si	Zn	S	O <sub>2</sub>	H <sub>2</sub> O
Photoionization lifetime	s	600	2500	25000	6000	4000	8000	5000	20000	8000	20000	50
Radiation acceleration	cm/s <sup>2</sup>	5	5	5	15	25	5	5	5	5	5	5

**Table 8.** Exospheric dynamics parameters for the different elemental species.

#### 635 Solar Radiation Pressure

636 Solar radiation pressure, a force exerted by the momentum of photons emitted by the sun, 637 plays a significant role in the movement of neutral particles in the exosphere. As these particles 638 are ejected from the surface of a planetary body, they encounter this radiation pressure, which 639 can alter their trajectories and velocities.

640 The effect of solar radiation pressure is particularly evident as it tends to push neutral 641 elements away from the direction of incoming sunlight, effectively propelling them toward the 642 night side of the planet. This movement is not uniform across all species; it varies depending 643 on the physical properties of the particles, such as their size, mass, and surface properties, or 644 cross-section, which influence how much momentum they absorb or reflect from solar photons.

645 Explain how the Mura model implements the solar radiation pressure, what are the significant 646 parameters, such as cross-section and acceleration...

#### 647 Other Simulation Parameters

648 Other relevant parameters that influence the data generation mechanism include:

- 649 • **The force of gravity** plays a fundamental role in the dynamics of particles within 650 the exosphere. It acts as the anchoring force that determines the trajectory and speed 651 of particles after they are ejected from the planetary surface. This effect is even more 652 significant considering the non-collisional nature of the exosphere. The gravitational term 653 in the state equation is more significant for the heavier species that are less affected by 654 radiation pressure.

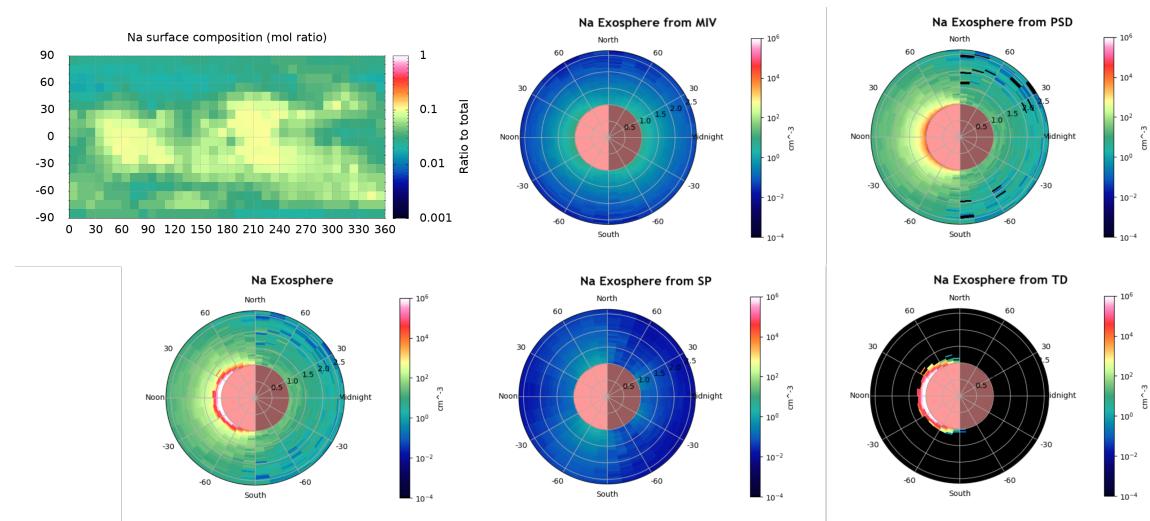
- **The number of particles** included in the simulation significantly affects the accuracy and complexity of the model. A higher number of particles allows for a more detailed and nuanced simulation of the exosphere, capturing better the statistical behavior of particle populations. However, the need for more computational resources and time increases with the number of particles. Balancing detailed simulations with computational efficiency is essential. For our purposes, considering the coarse resolution and the high number of simulations required, we have chosen a particle count of 50,000 for our simulation runs.
- **The exospheric grid** is a virtual representation of space around the planet, divided into discrete cells, serving as the framework for tracking particle positions and movements. This grid enables the simulation to map particle density, velocity, and direction across the exosphere's different regions. The grid's resolution—its fineness or coarseness—directly affects the simulation's precision. A finer grid captures more detailed spatial variations in particle behavior but demands greater computational resources. For our study, we have extended the surface 2D grid in 100 km altitude steps around the planet to a final altitude of **5000 km**.
- **The Coriolis force**, resulting from the planet's rotation, affects the motion of particles in the exosphere. It causes the path of moving particles to curve relative to the planet's surface, influencing their trajectories based on motion direction and latitude. While the Coriolis force doesn't change particle speed, it redirects their paths, creating complex circulation patterns within the exosphere. In our study, we have omitted the Coriolis force effect, thereby simplifying the parameter space.
- **Magnetic field parameters** are used for modeling the behavior of charged particles within the exosphere. The planet's magnetic field interacts with these particles, affecting their trajectories through Lorentz force. However, our model does not account for recombination between ions and electrons or track ions' paths around the planet, thus omitting magnetic field effects from our simulations.

---

#### 681 *Combining the Singular Exospheres*

682 To generate Mercury's exosphere, we ultimately employ a simplified additive approach. This  
683 method entails conducting separate simulations for each of the four primary surface release  
684 processes, reproduced for each distinct species. Following these individual simulations, we

aggregate the outcomes to compose the overall exosphere. This process essentially involves adding together the resulting singular exospheres generated for each species, without considering interactions between the various processes. For instance, we do not account for potential competition among processes for a finite pool of particles at the surface. Similarly, the exospheres for different species are treated as non-interacting entities. An example for the resulting Sodium exosphere is shown on Figure 7.



**Figure 7.** Resulting exosphere from the example Sodium surface composition. The figure depicts the generated singular Sodium exospheres from the MIV, SP, TD and PSD processes, as well as their aggregate. **Improve figure by removing unnecessary duplication of coordinates, bars, etc.**

Add some more sentences to complete this section...

Add example exospheric figures also for Mg and Ca???

### 3.2. Generation of the Datasets

This section describes the comprehensive process of dataset generation, detailing how we simulate measurements and observations that mimic real-world exospheric data. This process is critical for ensuring that the data used to train and validate our analytical algorithm is both representative and robust.

#### 3.2.1. Training, Validation and Test Datasets

We produce three distinct types of datasets: training, hold-out validation, and testing datasets. This subdivision is a fundamental practice in machine learning, ensuring that the

701 algorithm is trained on a diverse set of data, validated for accuracy, and finally tested for  
702 generalization to unseen data.

703 In our study we have built multiple training sets over which the algorithm has performed  
704 different trainings each resulting in a particularly trained and different multilayer perceptron  
705 DNN. The training set size plays a crucial role in the resulting estimator of physical processes'  
706 closeness to the presented data distribution. It needs to be noted that the actual distribution,  
707 in this case coming from the simulated model, is different from the one fed to the algorithm  
708 through the training set. This makes the construction of a representative training dataset very  
709 important in order to make the training data distribution as close as possible to the actual  
710 distribution. Moreover, the increase of training set size will reduce the probability that the  
711 algorithm gets biased towards a non-representative smaller size data distribution.

712 The hold-out validation involves using a single validation dataset to monitor the trained  
713 network during each training epoch. This process helps identify when the algorithm begins to  
714 overfit the training data, thus reducing its generalization capabilities. The validation set is also  
715 crucial during hyperparameter tuning to evaluate the algorithm's ability to generalize effectively.

716 Finally, the test datasets consist of examples (data points) that are never seen during  
717 training, and the MLP DNN algorithm's learning phase has not been influenced by them. This  
718 is important to make sure that the algorithm's performance evaluation is not influenced by  
719 improvement in its accuracy due to evaluation on previously seen examples.

720 *3.2.2. Datasets Generation Procedure*

721 The creation of our datasets adheres to the systematic procedure described below: (1) The  
722 surface elemental composition is randomly generated based on varied mineralogies, establishing  
723 the 'ground truth' for each dataset. (2) We simulate environmental effects to craft a static  
724 representation of the exosphere at a specific moment, taking into account the necessary  
725 physical and chemical processes. (3) Specific points within the exosphere are defined. (4)  
726 Exospheric measurements are taken at these points, and these measurements are compiled  
727 into data subsets, or observations, each representing a comprehensive snapshot of exospheric  
728 conditions from a particular simulated exosphere. (5) For the training datasets, multiple  
729 observations/subsets representing different exospheres are aggregated, enhancing the dataset's  
730 complexity and variability. Conversely, the hold-out validation and testing datasets are each

731 derived from individual observations/subsets to maintain representability of particular planetary  
732 surface/exosphere pairs. (6) Relevant features are selected or engineered from these observations  
733 to serve as inputs for the algorithms.

734 We already discussed points (1) and (2) as part of our environmental and exospheric model  
735 description. The rest of the procedure is described below.

736 To enhance the algorithm's learning potential and adaptability, this dataset generation  
737 process is replicated across a multitude of randomly created surface scenarios. This approach  
738 ensures a comprehensive and diverse training experience for the algorithms.

739 In anticipation of the insights that the BepiColombo mission will provide upon its arrival at  
740 Mercury, particularly through the SERENA suite of particle detectors, our datasets incorporate  
741 simulated measurements of exosphere densities and proton fluxes. These mock-up measurements  
742 are generated to reflect the observational capabilities and perspectives of the Mercury Planetary  
743 Orbiter, blending the orbital motion of the spacecraft with detailed exospheric modeling.

744 *3.2.3. Positioning in the Exosphere*

745 We strategically define various positions within the exosphere to place virtual sensors,  
746 aiming to comprehensively provide measurements of its constituents. These positions were  
747 selected to maximize data diversity and relevance, considering factors like altitude, latitude, and  
748 environmental conditions. In order to balance the required spatial variability and to facilitate  
749 the reconstruction of complete surface maps from DNN predictions, we align our virtual sensor  
750 positioning in the exosphere directly above each surface grid tile's center. This placement  
751 strategy results in a total of 648 measurement data points for each data subset.

752 Given the mission-specific orbital characteristics of the BepiColombo/Mercury Planetary  
753 Orbiter and the enhanced measurement capabilities of the STROFIO mass spectrometer within  
754 the SERENA instrument suite in lower altitudes, we select an optimal baseline altitude of 500  
755 km for our measurements. This altitude, corresponding to the periherm (closest approach) of the  
756 MPO, is chosen for its potential to yield accurate measurements with an improved signal-to-noise  
757 ratio.

758 To enhance our training datasets, we applied data augmentation techniques, incorporating  
759 measurements from a range of altitudes, including 200, 500, 800, 1100, 1400, 1700, and 2000 km.  
760 This approach not only increases the dataset volume but also introduces a variety of exospheric

761 conditions, allowing the algorithm to develop a nuanced understanding of how measurements  
762 at different altitudes correlate with surface compositions. In our training campaign we examine  
763 both the baseline training dataset type with examples only at altitudes of 500 km and the  
764 augmented training dataset type. However, training the MLP DNN on the latter is essential for  
765 its ability to make predictions at diverse altitudes.

766 For the validation and testing phases of our DNN algorithms we adopt a different strategy.  
767 The hold-out validation dataset is set at the MPO’s periherm altitude of 500 km to maintain  
768 consistency in evaluation conditions. On the other hand, to ensure a thorough assessment of the  
769 algorithm’s performance, we generate multiple test datasets, each at a constant altitude ranging  
770 from 200 km to 1500 km. This allows for a comprehensive analysis and evaluation across a  
771 broader spectrum of exospheric altitudes and conditions, enabling us to more accurately gauge  
772 the algorithm’s efficacy and robustness in predicting the surface elemental composition.

773 *3.2.4. Measurement and Observations Creation*

774 The methodology for generating measurements and observations within our simulated  
775 environment focuses on the quantification of particle densities across various elemental species  
776 in the exosphere by a hypothetical, idealized virtual sensor. This conceptual sensor operates  
777 without considering the complexities and potential inaccuracies introduced by real-world sensor  
778 characteristics such as detector noise and sensitivity limitations.

779 For the purpose of creating a comprehensive and static snapshot of the exosphere,  
780 measurements across the simulated exosphere are conducted simultaneously. These  
781 measurements collectively form what we refer to as an observation, or a data subset. Each  
782 observation/subset comprises 648 measurement data points (or 4536 points in the case of  
783 augmented training subsets designed to capture a broader range of altitudinal data). Every  
784 data point within an observation is a vector that encapsulates the measured densities of all  
785 neutral species present in the exosphere as per our simulation’s setup.

786 *3.2.5. Datasets Consolidation*

787 For validation and test dataset purposes, employing single data subsets allows for  
788 straightforward prediction and reconstruction of surface maps corresponding to individual  
789 surface-exosphere simulations. For the hold-out validation dataset, we adopt a more focused  
790 approach by selecting a single baseline subset based on measurements at a 500 km altitude.

791 The test datasets, on the other hand, are designed to thoroughly assess the algorithm's  
792 predictive capabilities under varied conditions. We generate single baseline subsets from 15  
793 distinct exosphere simulations, providing a diversified testing ground. Additionally, to explore  
794 the algorithms' responsiveness to altitude variations, we create multiple test sets from the same  
795 15 surface-exosphere pairs, but with measurements taken at constant altitudes for each altitude-  
796 varied subset ranging from 200 km to 1,500 km.

797 However, the training phase demands a more complex strategy that captures a larger part of  
798 the data distribution, in order to train the algorithm to approximate the relationships between  
799 the employed physical processes accurately. To address this, we aggregate multiple data subsets  
800 or observations to form the training dataset, with the number of included subsets ranging from  
801 10 baseline observations (yielding a total of 6,480 data points) to 300 augmented observations  
802 (resulting in a staggering 1,360,800 data points). It should be noted that utilizing each training  
803 dataset, results in an algorithm trained to represent a different empirical data distribution  
804 derived from this particular dataset. A more detailed description of the datasets and their  
805 underlying surfaces is given in Appendix A.

806 *3.2.6. Feature Selection and Engineering*

807 The effectiveness of deep neural network algorithms in modeling complex relationships within  
808 data is significantly influenced by the selection and engineering of input features. These features,  
809 which describe each data point within all three types of datasets, are pivotal in guiding the  
810 algorithm towards an accurate and meaningful representation of the underlying data structure.  
811 Carefully chosen or engineered features can enhance the algorithm's ability to discern patterns  
812 and relationships, thereby improving its overall performance.

813 In our study, we have carefully considered and incorporated a range of additional features to  
814 enrich our datasets for training and testing the DNN algorithms. These features are designed to  
815 provide the algorithms with nuanced insights into the dynamic interactions within the exosphere,  
816 thereby facilitating a deeper understanding of the data:

- 817 • **Altitude of Measurement:** This feature is critical for capturing altitude-specific  
818 dynamics, enabling the algorithm to identify how the distribution of neutral species changes  
819 with altitude relative to their source points on the surface. Both the actual altitude and a  
820 logarithm of the altitude were tested as features. The latter engineered feature is intended

821 to highlight non-linear altitude effects on the measured parameters, providing another layer  
822 of depth to the altitude-related analysis.

- 823 • **Logarithm of Exospheric Density:** By applying the base 10 logarithm to the exospheric  
824 density of each species, we introduce a constraint that aids the algorithm in exploring  
825 nonlinear relationships, acknowledging the exponential decrease in density with altitude.
- 826 • **Subsolar Angle:** Represented as either the value of the angle  $\phi$  directly, or as  $\cos(\phi - 180)$ ,  
827 this feature helps differentiate between exospheric populations on the dayside, nightside,  
828 and the transitional terminator regions, enhancing the model's spatial awareness.
- 829 • **Latitude Dependency:** Using  $\sin(\gamma)$ , this feature allows the algorithm to account for  
830 latitude-specific phenomena, such as ion sputtering, which varies across different latitudinal  
831 zones.
- 832 • **Proton Flux Virtual Data:** Integrating virtual measurements akin to those from the  
833 MIPA ion detector of the SERENA suite, this feature hints at sputtering effects induced  
834 by proton precipitation on the surface, offering a proxy for understanding underlying ion  
835 induced physical processes.

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

#### 841 4. Results

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

Our approach anticipates the dual avenues of future research in this domain. Firstly, it lays the groundwork for further refining the DNN's architecture and tuning process, aiming at an even more precise representation and understanding of the selected physical processes parameter space. Secondly, it sets the stage for applying the optimized DNN structure to different regions within the processes parameter space, testing its robustness and adaptability in estimating data distributions from novel and varied conditions. This bifurcated strategy underscores our commitment to advancing the field by not only enhancing algorithmic performance in current scenarios but also ensuring their applicability and efficacy in diverse scenarios.

*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 in anticipation of a thorough testing examination of its performance. This endeavor is significant, as mapping the trajectory through hyperparameter space holds considerable promise for future research in this domain. The complete training campaign is detailed in Appendix B.

*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.

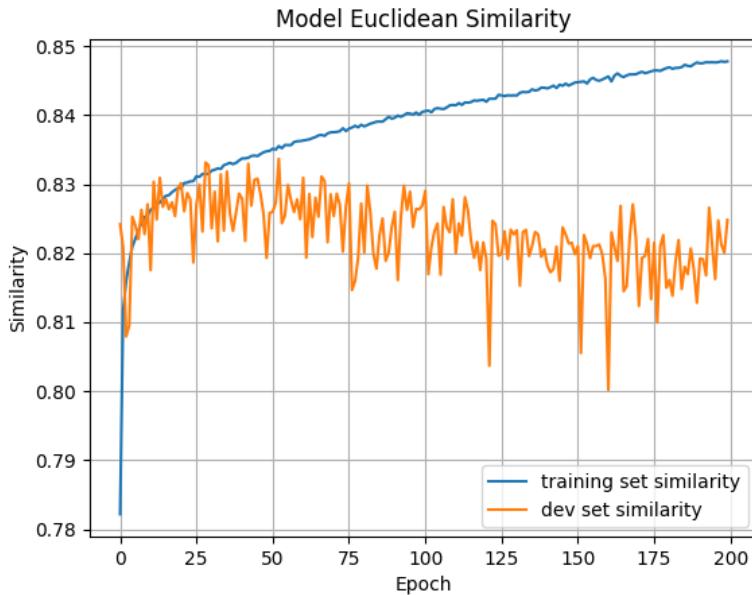
*Training Set Size and Data Augmentation*

We analyzed MLP DNN performance in relation to the expansion of the training dataset size with training sets ranging from 10 to 200 unaugmented 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

879 in altitude, enhancing representability. This strategic choice expanded our dataset to 300  
 880 augmented subsets, totaling 1,360,800 examples, significantly improving the model’s robustness  
 881 and predictive accuracy across altitude-specific inputs.

882 *Learning Curve Examination for Optimal Training Duration*

883 Our examination of the MLP DNN’s learning curves aimed to identify the optimal training  
 884 duration to avoid overfitting, in line with the guidance provided by Bengio (2015). Analysis  
 885 indicated that predictive performance on the validation dataset peaked at 40 epochs, as shown  
 886 in Figure 8. While longer training durations, up to 200 epochs, continued to align the model to  
 887 the training dataset, the best balance between training and inference accuracy was achieved at  
 888 40 epochs, suggesting this as the optimal training duration.



**Figure 8.** 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 development hold-out validation dataset (1 subset, 648 data points) respectively.

889 *Selected Feature Set*

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

- 892 • Logarithmic transformations of elemental species exospheric densities.

- 893 • Logarithm of the altitude at which measurements were taken.
- 894 • Sun incidence angle.
- 895 • Presence of H+ ions arriving through open field lines.
- 896 • Cosine of solar time longitude.
- 897 • Sine of latitude.

898 *Hyperparameter Optimization and DNN Structural Components Finalization*

899 An extensive hyperparameter optimization effort resulted in the selection of the final MLP  
900 DNN architecture, consisting of a four-layer structure with 600, 500, 350, and 250 neurons in  
901 each layer respectively (Figure 9). The regularization coefficient was optimized to a higher value  
902 of  $1.0 \times 10^{-5}$  to improve generalization, while the learning rate was finely tuned to  $0.5 \times 10^{-4}$ .  
903 Training was conducted in mini-batches of 512 examples, identified as near-optimal through our  
904 optimization process.

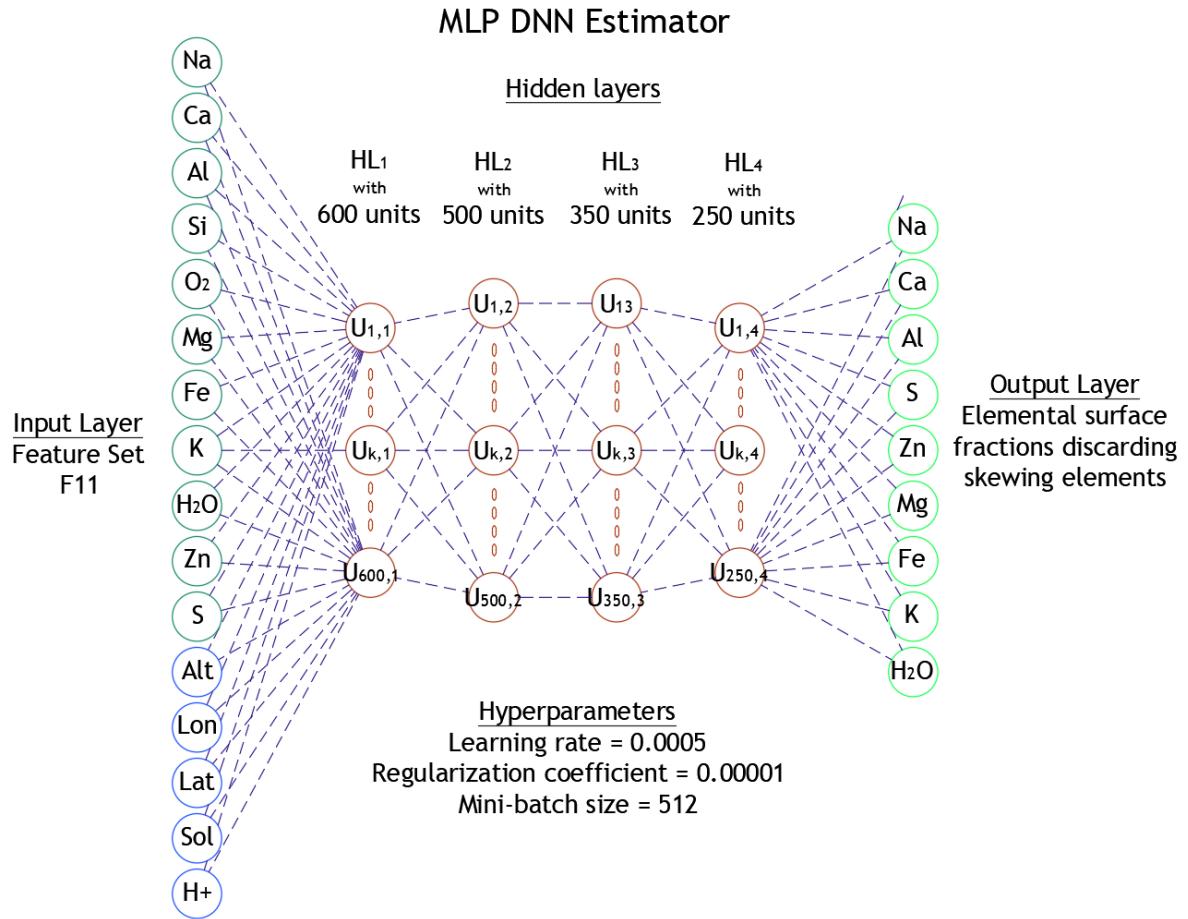
905 The combined efforts of hyperparameter exploration, architectural fine-tuning, and dataset  
906 augmentation have significantly propelled our model’s performance and its ability to generalize  
907 from the empirical distribution to the true underlying data-generating distribution—by as much  
908 as a total of 10% in ES4 and up to 30% in R-squared. This optimized architecture, along with  
909 structural parameters and algorithmic characteristics refined during our comprehensive training  
910 campaign, ensures that the MLP DNN is a robust model for our sophisticated predictive tasks.

911 *4.2. Testing Phase*

912 The ultimate evaluation of our multilayer perceptron deep neural network algorithm’s  
913 performance hinges on its ability to accurately predict surface compositions and reconstruct  
914 elemental surface maps using datasets it has not previously encountered. This phase is crucial  
915 as it tests the algorithm’s generalization capabilities beyond the conditions it was trained under.

916 Our research incorporates two distinct test campaigns, designed to assess the MLP network’s  
917 predictive prowess. These campaigns were structured to apply the final network, fine-tuned with  
918 an extensive training set comprising 300 augmented subsets, across test datasets derived from  
919 a variety of altitudes not previously seen during training. This approach ensures a rigorous  
920 evaluation of the network’s adaptability and accuracy across diverse conditions.

921 The scope of these test campaigns is broad, focusing not only on aggregate performance



**Figure 9.** 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 O<sub>2</sub> and Si species. There are four hidden layers with 600, 500, 350, and 250 neurons respectively.

922 metrics across the entire dataset but also on detailed analyses for individual elemental species.  
 923 This includes a thorough examination of residuals to identify any systematic errors or biases in  
 924 predictions. Moreover, an essential component of our evaluation is the reconstruction of surface  
 925 maps. This process entails a visual comparison between the original, or "ground truth", maps  
 926 and the ones predicted by our algorithm. By doing so, we aim to provide a holistic view of  
 927 the algorithm's capability in reproducing detailed and accurate surface compositions, thereby  
 928 illustrating its potential for practical application in exospheric studies and planetary science.

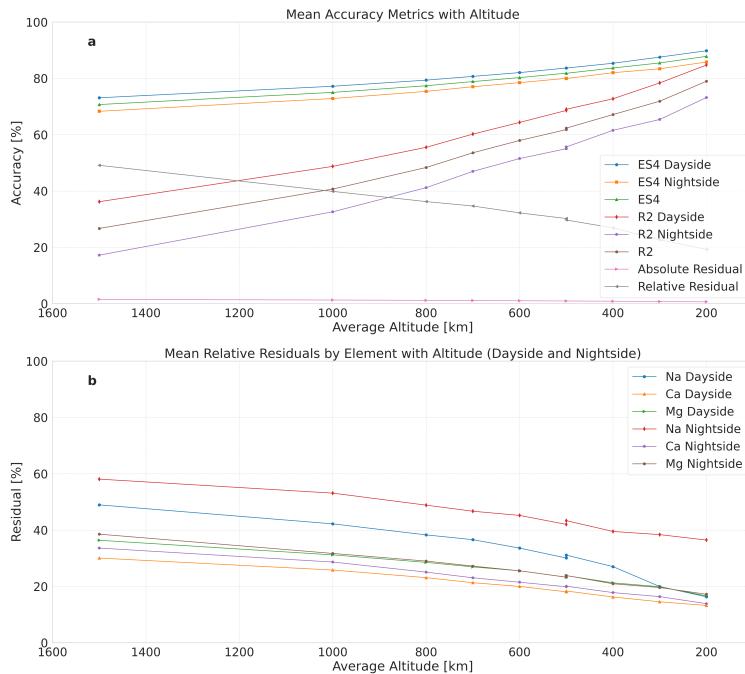
929 *4.2.1. Preliminary Test Campaign*

930 In our preliminary test campaign, we embarked on a performance evaluation using single-  
 931 simulation test datasets derived from 15 unique surface compositions, each leading to distinct

exospheres. This approach encompassed data from both the dayside and nightside, allowing for a robust examination of the MLP DNN algorithm's predictive accuracy and its capability in reconstructing surface elemental maps under varying conditions.

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

We utilized our suite of performance metrics, including the average ES4, R-squared, absolute, and relative residuals, to evaluate the predictions and reconstructions systematically. These metrics were plotted against the measurement altitudes to analyze the model's performance comprehensively, depicting them for the overall predicted output, individual elemental species, and separate analyses for dayside and nightside predictions (Figure 10).



**Figure 10.** Preliminary test campaign. Panel a shows the mean accuracy (dayside, nightside and overall) and residuals metrics of the MLP DNN predictions on the 15 test surfaces. In panel b are displayed the dayside and nightside mean relative residuals for the elements Na, Ca, and Mg.

The results of this preliminary testing suggested that predictions were generally more precise

947 for the dayside, a likely consequence of particle movements influenced by solar radiation pressure.  
948 This was especially true for volatile species, such as Sodium, on which we observed a notable  
949 discrepancy in predictive accuracy with respect to refractory ones, like Magnesium and Calcium.

950 The challenge of accurately predicting nightside compositions, coupled with the spin-orbit  
951 resonance of Mercury, hinted at the possibility to improve the results of our model's complete  
952 surface predictions and map reconstructions in a second test campaign. Given that in its  
953 following orbit around the Sun the nightside hemisphere of the planet is illuminated, this  
954 subsequent campaign aims to utilize dayside only measurements during two consecutive Mercury  
955 orbits, addressing the identified gaps and leveraging this insight gained from the preliminary  
956 campaign.

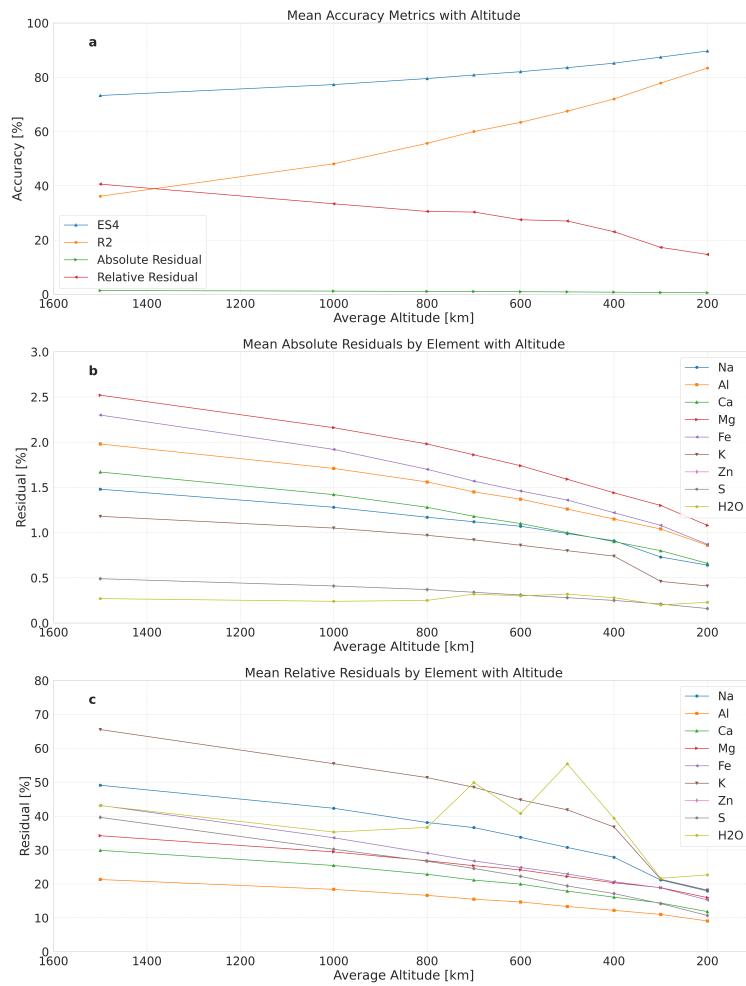
957 *4.2.2. Main Test Campaign*

958 The main test campaign was strategically designed to leverage the unique rotational and  
959 orbital dynamics of Mercury, particularly its 3:2 orbit-spin resonance, to observe the entire  
960 planetary surface under daylight conditions across two successive Mercury years. This approach  
961 utilized double-simulation compound predictions, focusing on the same 15 surface compositions  
962 from the preliminary campaign but observed at two consecutive perihelia. During these  
963 two periods, different halves of Mercury's surface were illuminated by the Sun, allowing for  
964 comprehensive daylight observation of the entire planet over the two simulations. For this  
965 campaign, predictions specifically targeted sunlit surface tiles, enabling an in-depth analysis  
966 of surface compositions that were previously obscured by darkness in the initial test phase.  
967 Measurements for this campaign were again taken at a range of altitudes from 200 to 1500 km.

968 A significant outcome of the combined odd-even orbit campaign was the improved accuracy  
969 in predicting volatile species' distributions, aligning more closely with the refractory species'  
970 predictions observed in the preliminary campaign. This enhancement in predictive accuracy for  
971 volatiles under daylight conditions underscores the importance of solar illumination in accurately  
972 assessing surface compositions.

973 The main campaign demonstrated a notable increase in overall prediction and map  
974 reconstruction accuracy, with the average ES4 metric reaching approximately 89.70% and the  
975 average R-squared metric reaching 83.41% at the lowest altitude of 200 km (Figure 11). This  
976 accuracy diminished at higher altitudes, attributed to the exosphere's dynamic nature and the

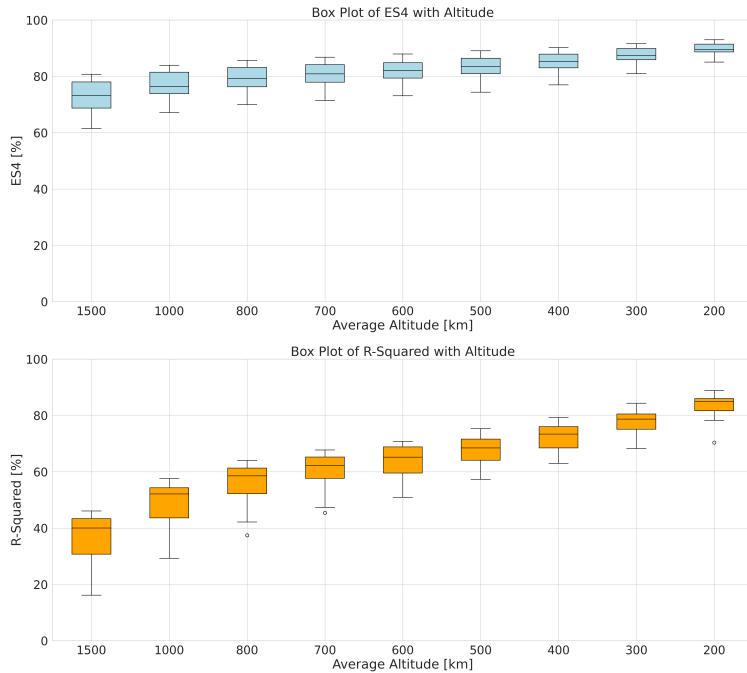
increased complexity in tracing back exospheric particles to their originating surface tiles. There was, however, a marked improvement compared to the preliminary findings, which highlights the efficacy of considering Mercury's solar exposure in enhancing predictive models' accuracy. By focusing solely on the dayside observations across two perihelia, the campaign effectively capitalized on optimized conditions for surface composition reconstruction.



**Figure 11.** Mean accuracy metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign. Panel **a** gives an overview of the overall metrics combining all elements with decrease in altitude. Panels **b** and **c** give respectively the absolute and relative residuals for all predicted elements averaged over the 15 predicted surfaces.

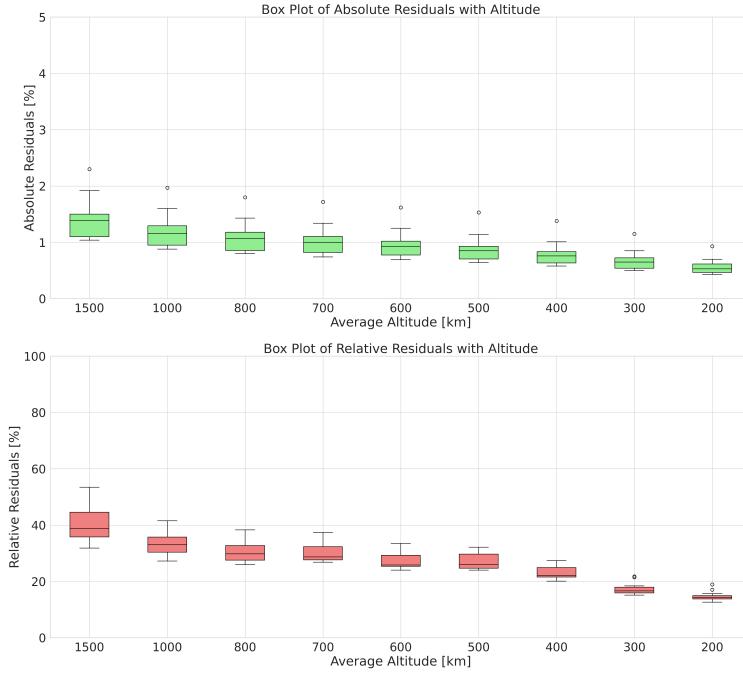
The box statistical plots shown on Figures 12 and 13 suggest also reduction in the range of prediction accuracies and residuals with decrease in altitude. While the difference between the worst and best predicted test sets at an altitude of 1500 km is close to 19.30% ES4 and 29.80% R-squared, the robustness in the prediction of different types of surfaces is increased and the

range is subsequently reduced to 7.8% ES4 and 18.5% R-squared at an altitude of 200 km. The best prediction (test set 15) at this lower altitude reaches the very respectable 92.98% ES4 and 88.83% R-squared.



**Figure 12.** Box plots of the accuracy metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign. [Possibly add this to Figure 11.](#)

Going deeper in the detailed statistics of the predictions by the MLP DNN per elemental species, we can observe its tendencies in the box plots of the absolute and relative predictions (Figures 14 to 22). Almost all elements have good prediction statistics at the lowest altitude, with the median relative residual of Aluminium particularly impressive at only 8.69% at 200 km (Figure 14). The algorithm has the most difficulties with the Water and Sodium elements with their median relative residual at 20.13% (Figure 17) and 17.86% (Figure 20) respectively at 200 km, which is nevertheless a good result. The robustness in prediction of the refractive elements is present throughout the altitudes even up to 1500 km, where the median for the relative residuals of Aluminium is at 21.18%, while those for Calcium and Magnesium are 30.76% (Figure 15), 33.01% (Figure 19), respectively. Some volatile elements, especially Potassium and Water are notably difficult to predict, with the median of the relative residual for Potassium hovering around 40% for most of the altitudes, and going below 20% only below altitudes of 300 km. While, the Water ice on the surface is predicted inconsistently with altitude, probably due to its

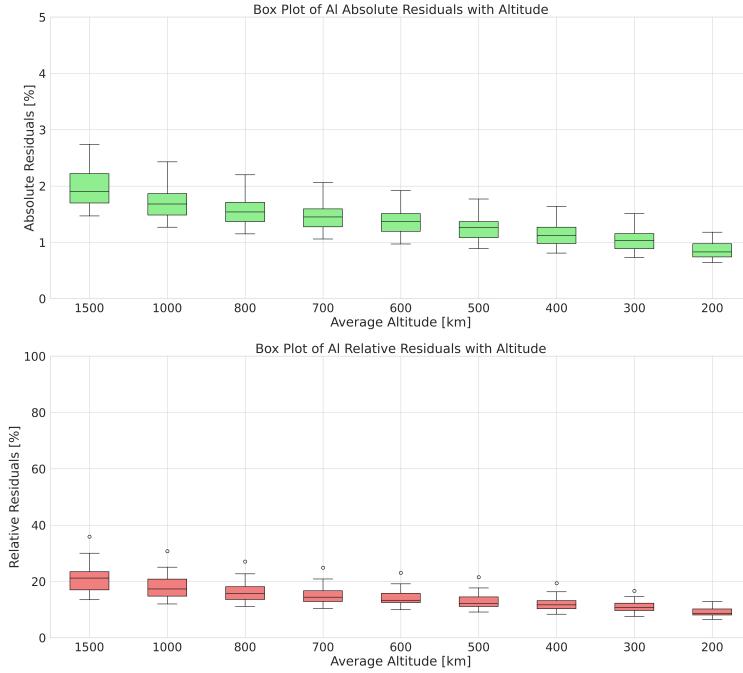


**Figure 13.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign. Possibly add this to Figure 11.

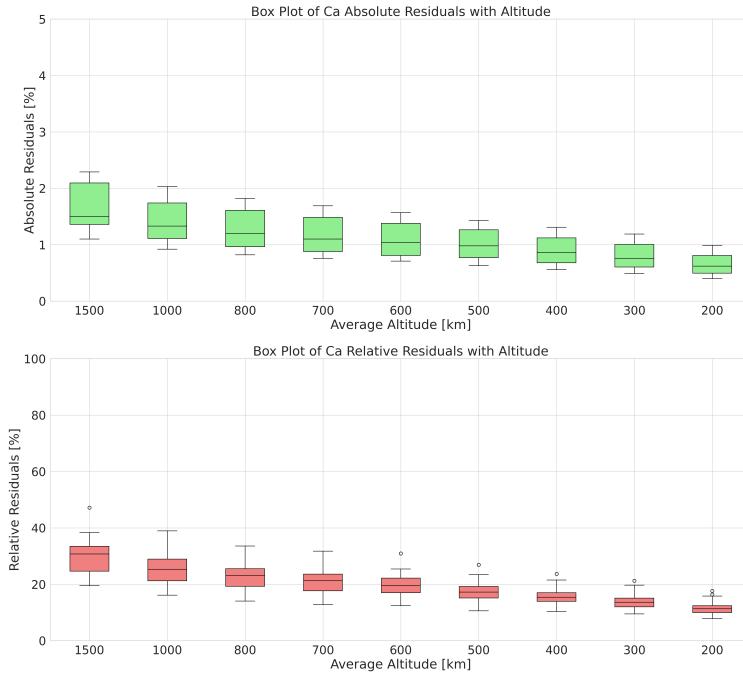
1002 low presence and small lifetimes in the exosphere in our present model. The prediction of the  
 1003 elements Sulfur and Zinc on the other hand is very robust throughout the test sets and altitudes.  
 1004 The median of their prediction errors is consistently below 20% up to an altitude of 600 km,  
 1005 and only at an altitude of 1500 km goes close to 40%. While the range of the prediction errors  
 1006 for the different types of surface is impressive through the altitudes.

1007 Additionally, our visual comparisons from the map reconstructions (Figures 23 to 31) and  
 1008 the accuracy metrics maps (Figure 32) highlighted the algorithm's strengths and weaknesses in  
 1009 predicting different elemental distributions.

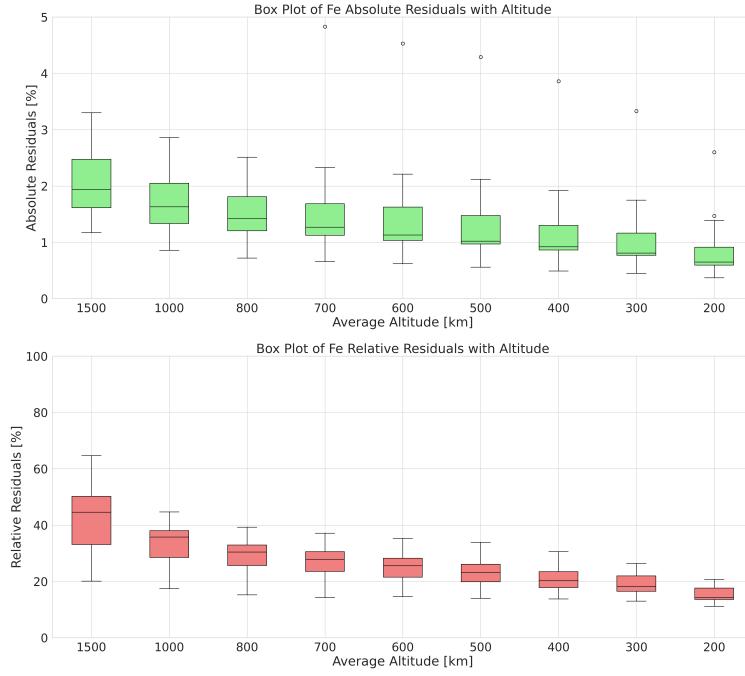
1010 The discrepancy in predictive accuracy between refractory species, such as Calcium (Figure  
 1011 24) and Magnesium (Figure 28), and volatile ones, like Sodium (Figure 29), is still present in  
 1012 the predictions of the particular test set 2 shown on the figures. However, there is a noticeable  
 1013 improvement by the daylight only predictions of the main testing campaign, compared to the  
 1014 one from the preliminary campaign, which included the night side prediction. This is shown  
 1015 on the map reconstructions of Sodium (test set 2), where the bottom-most panels in Figure  
 1016 29 show that increased errors on the night side (longitudes 0-90 and 270-360) and an average  
 1017 relative residual of 23.73% from the preliminary test, directly compared to the average of the



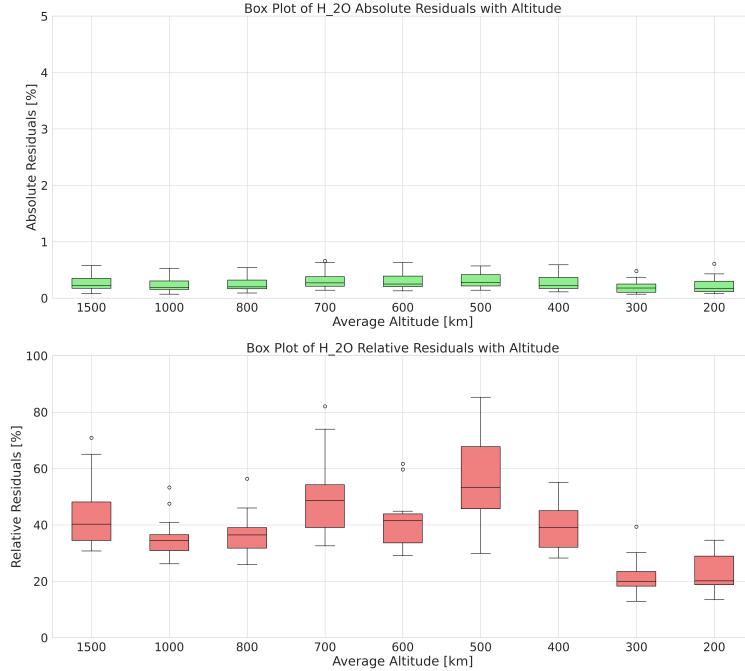
**Figure 14.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Aluminium.



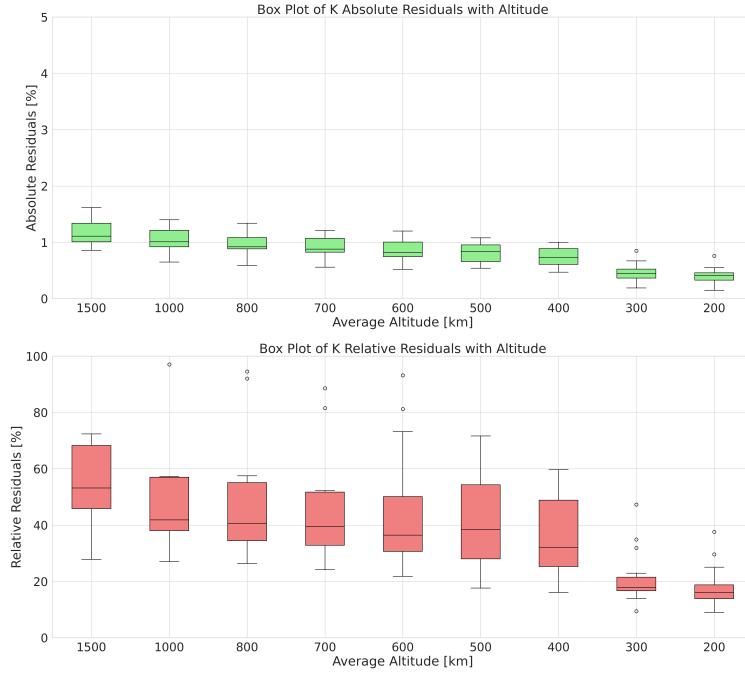
**Figure 15.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Calcium.



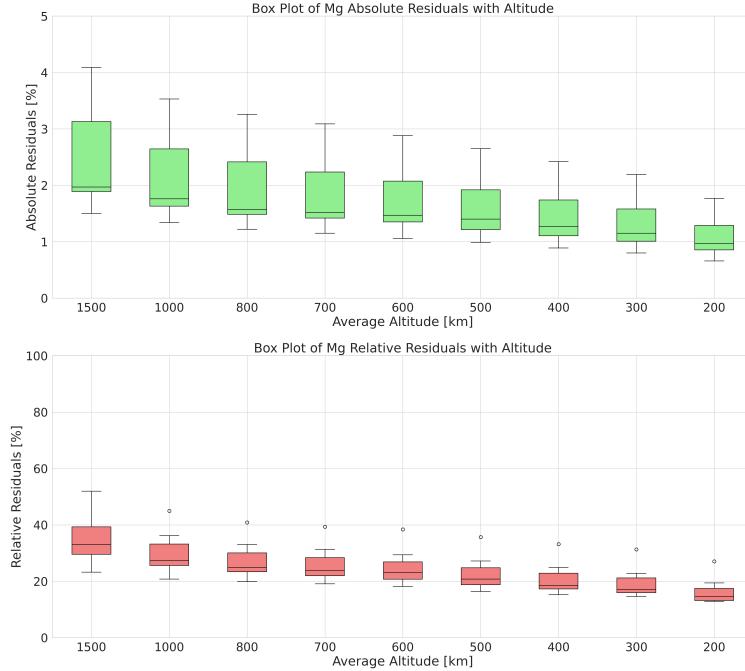
**Figure 16.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Iron.



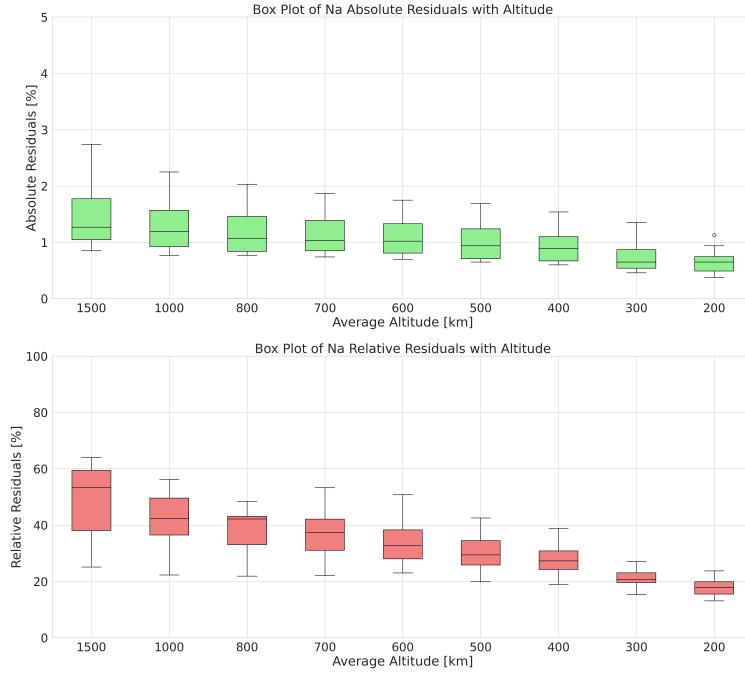
**Figure 17.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Water.



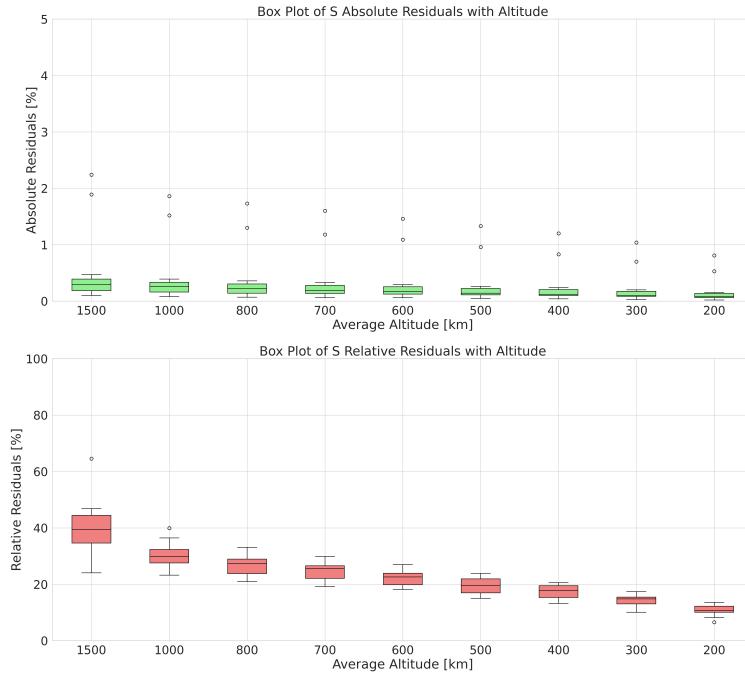
**Figure 18.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Potassium.



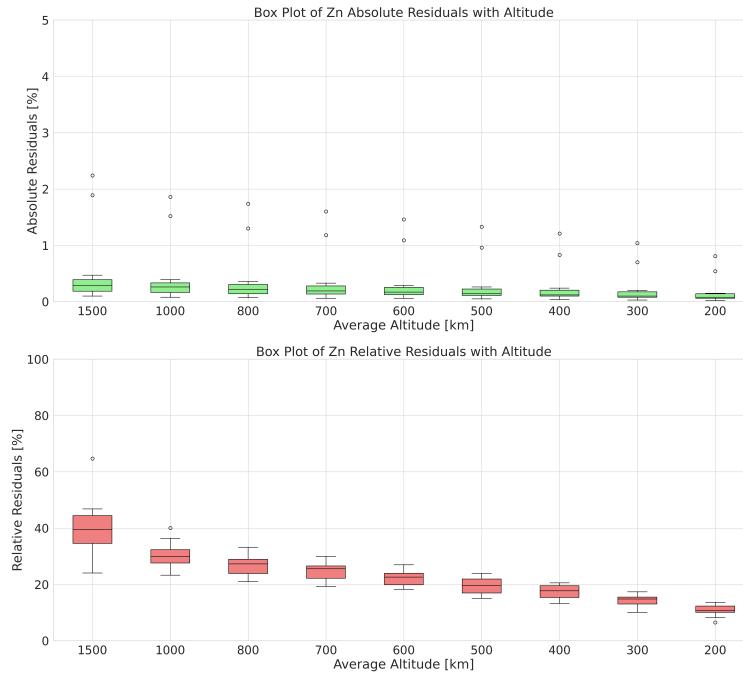
**Figure 19.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Magnesium.



**Figure 20.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Sodium.



**Figure 21.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Sulfur.

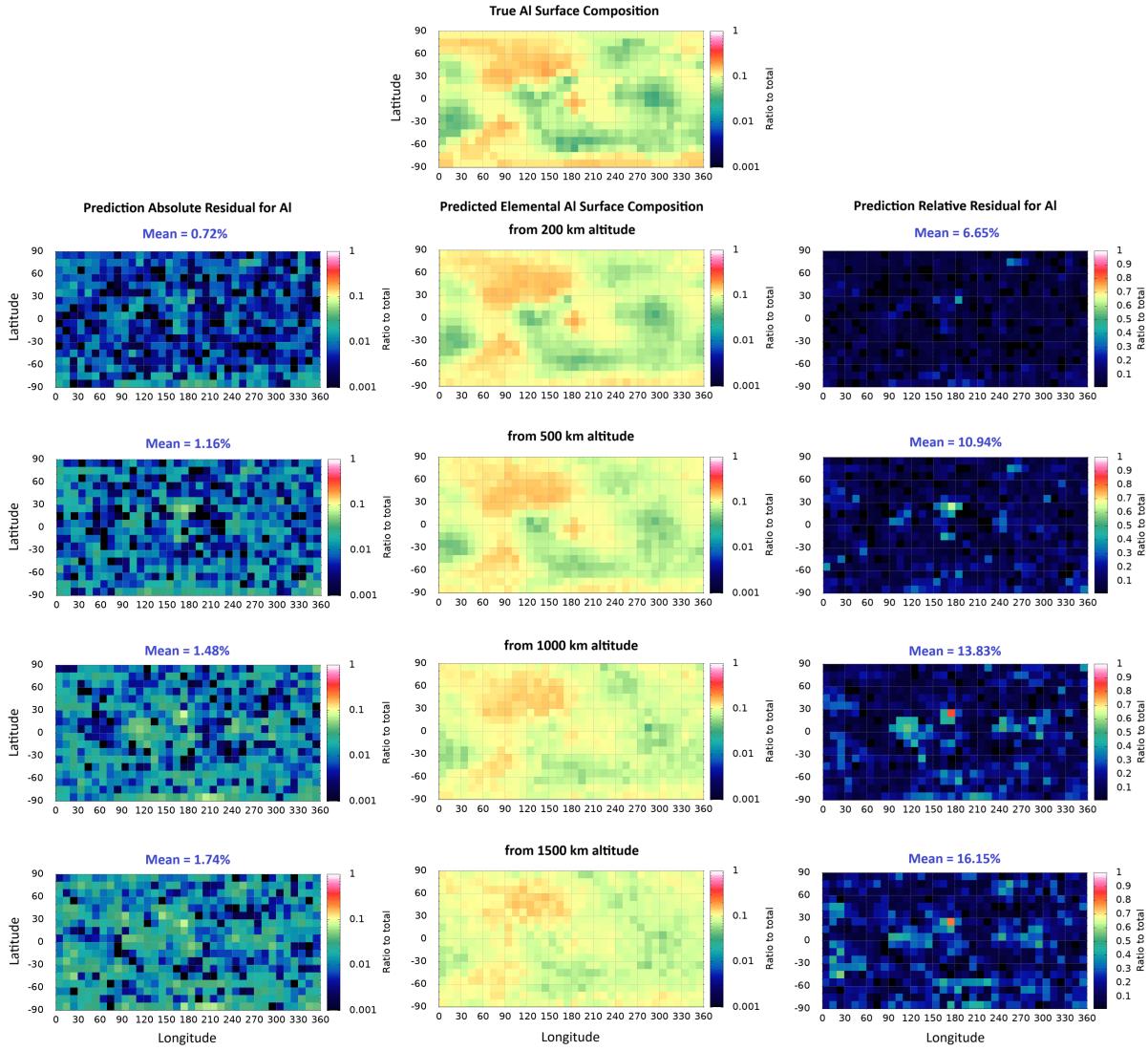


**Figure 22.** Box plots of the residuals metrics of the MLP DNN predictions on the 15 test surfaces of the main test campaign for the element Zinc.

1018 main campaign at 500 km (third row from top) where the average relative residual is reduced  
 1019 to 20.62%.

1020 From a closer examination of the reconstructed maps, we can clearly see that certain large  
 1021 scale patterns in the surface composition of all species are recognized from as far as 1500 km,  
 1022 even if details are missed by the MLP at that large distance. The algorithm starts to resolve with  
 1023 a relatively good accuracy at altitudes of 500-800 km. This is especially true for the refractory  
 1024 species (Ca, Mg, etc.), and going closer to the planet (down to 200 km) produces the most  
 1025 accurate predictions and reconstructions, even for volatiles (Na).

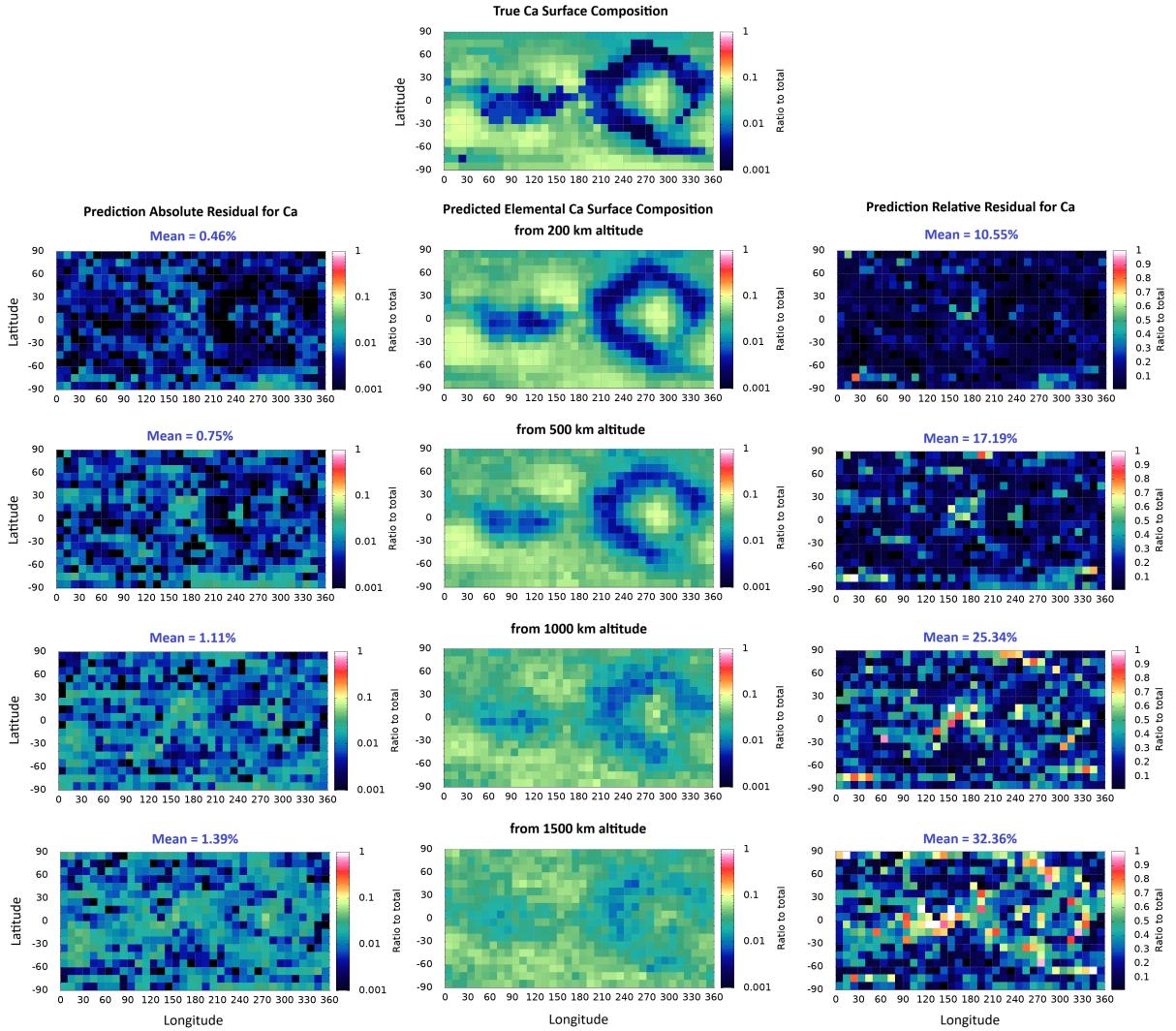
1026 The minor elements, such as water, suffer from jumps and anomalies in their prediction  
 1027 relative residuals, which is attributed to their low overall fraction in the premises of our model,  
 1028 which can result in its fraction being given less weight in the overall learning of the algorithm.  
 1029 The same is not observed for the sulfur and zinc predictions. Even though their fractions on the  
 1030 surface are low, they are coupled to each other in our model due to the only source for them  
 1031 being the sphalerite mineral. This means that the MLP was able to estimate the coupling of  
 1032 the predictions for these two fractions better, even if the input exospheric densities are widely  
 1033 different for the two elements.



**Figure 23.** 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.

## 1034 5. Discussion and Future Work

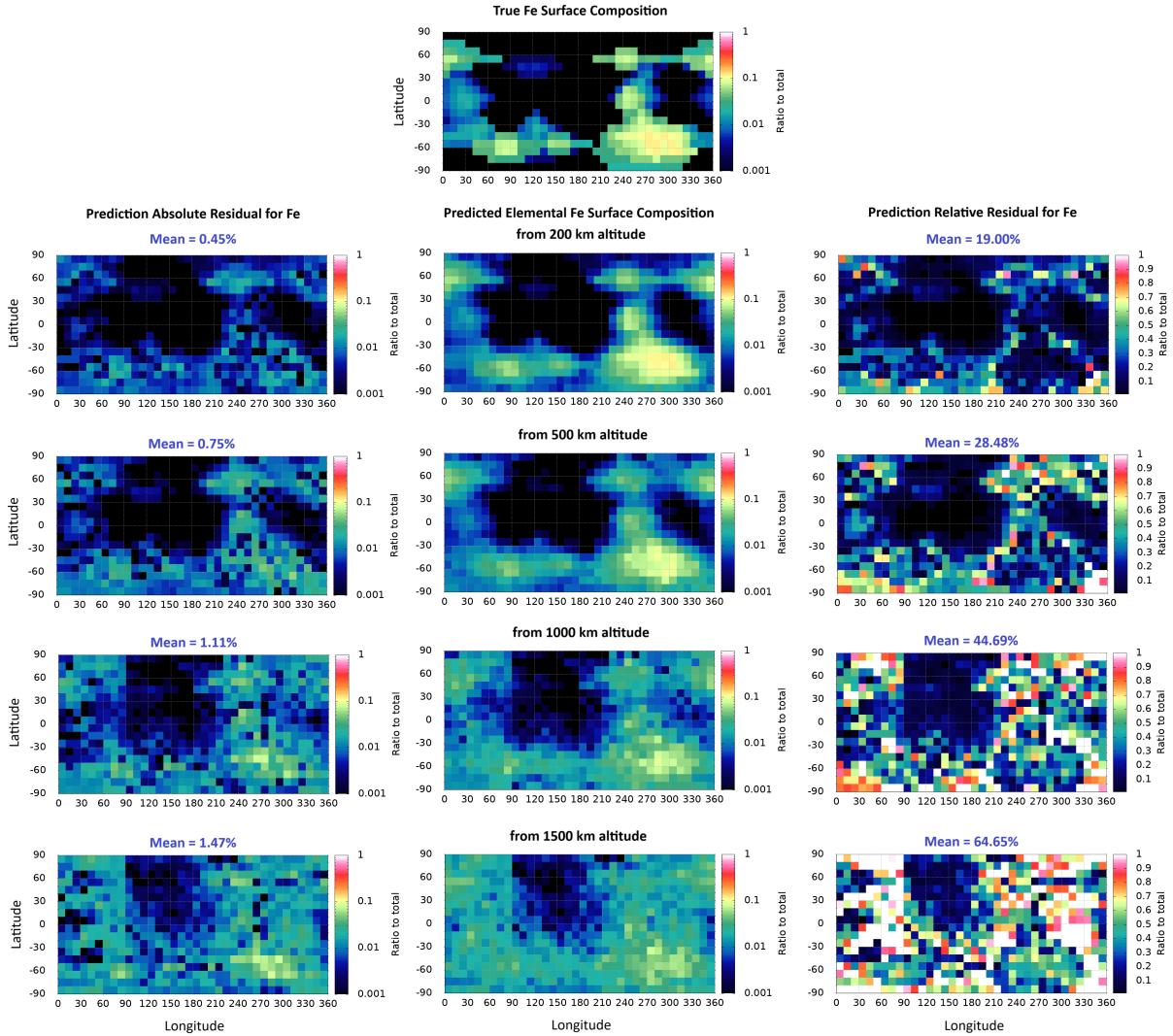
1035 This research presents a novel approach for deducing Mercury's surface composition using  
 1036 advanced deep neural networks to analyze exospheric density measurements. The algorithm  
 1037 is adept at predicting elemental compositions and reconstructing surface maps, crucial for  
 1038 understanding how neutral atoms transition from the planet's regolith into its exosphere. Key  
 1039 advancements include refining the simulated model's parameter space, developing a sophisticated



**Figure 24.** 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.

multilayer perceptron DNN architecture, applying Bayesian hyperparameter tuning for optimal configuration, and integrating domain-specific knowledge into the feature selection process.

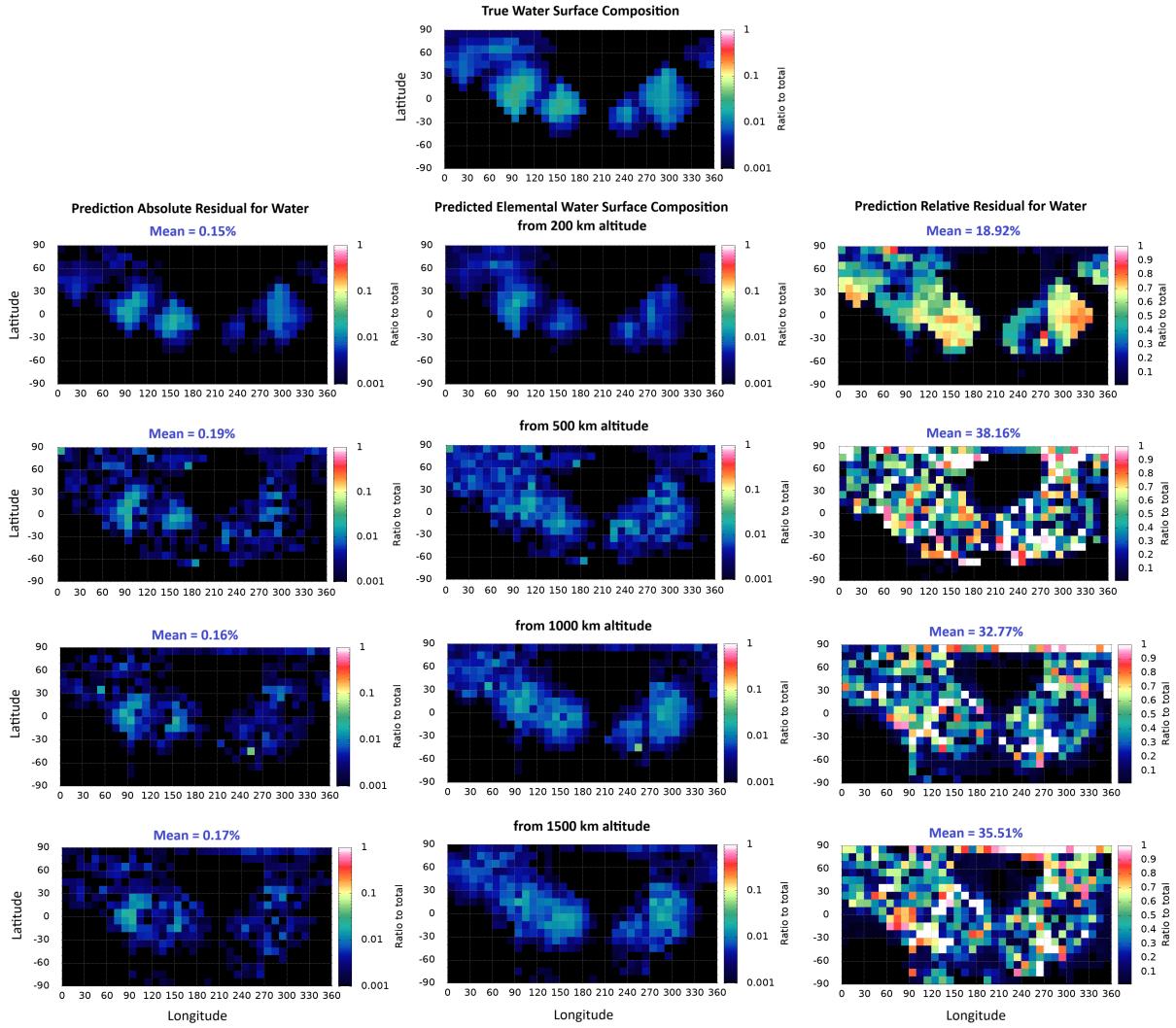
The principal accomplishment of this study is the creation of a DNN that accurately estimates surface-exosphere interactions on Mercury. This algorithm effectively represents a specific region of the physical parameter space, mapping complex relationships between exospheric and surface compositions. By yielding an estimator for exosphere generation processes, the study bridges



**Figure 25.** 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.

theoretical modeling with practical data-driven predictions. This has the potential to enhance our understanding of Mercury's exosphere and offers new avenues for research in planetary science and machine learning, suggesting potential for significant future discoveries.

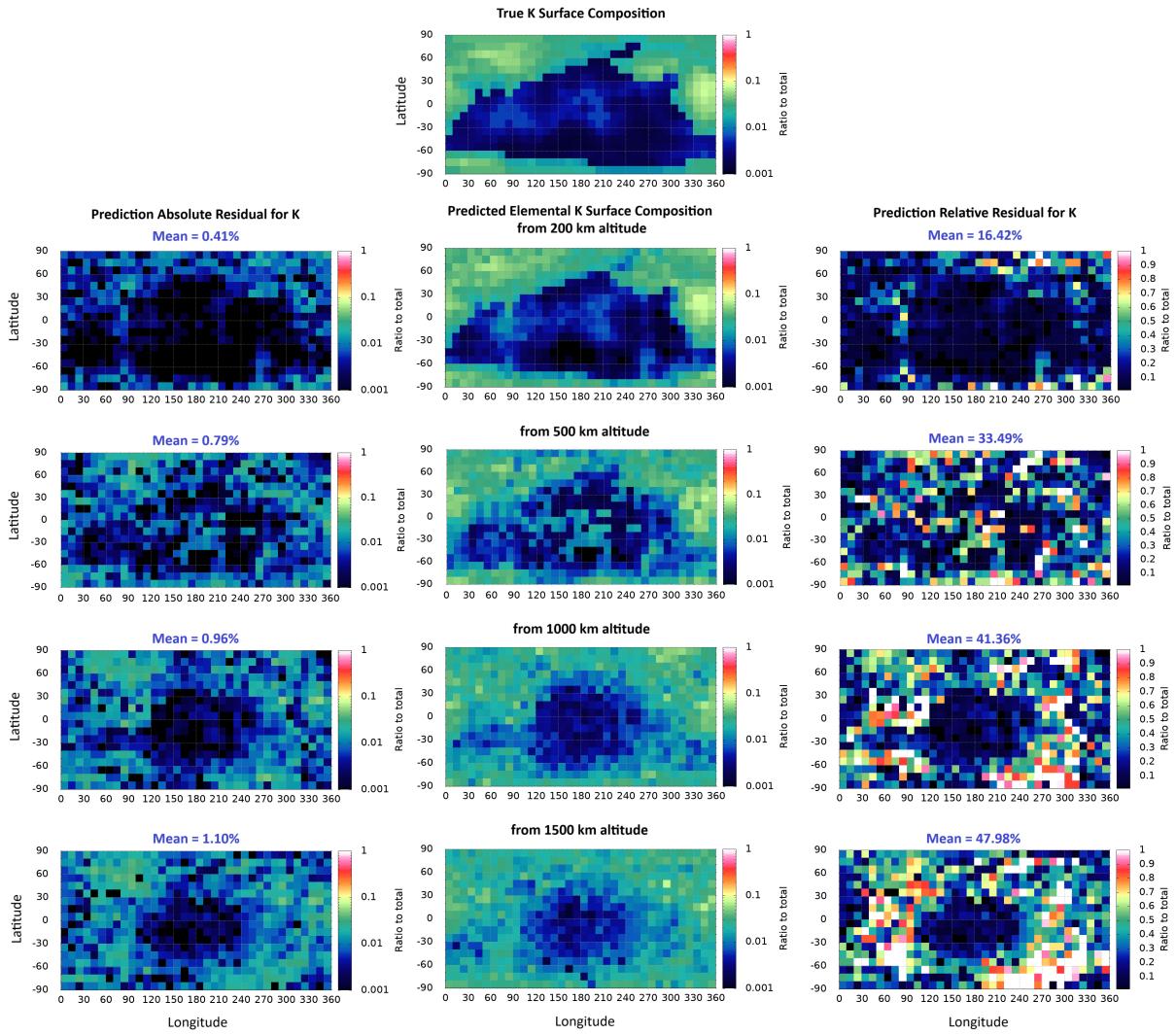
A preliminary test phase demonstrated the model's superior predictive performance on Mercury's dayside but highlighted less precise predictions for volatile elements like sodium (Na) and potassium (K) on the nightside. This led to a focused test campaign using only simulated



**Figure 26.** 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.

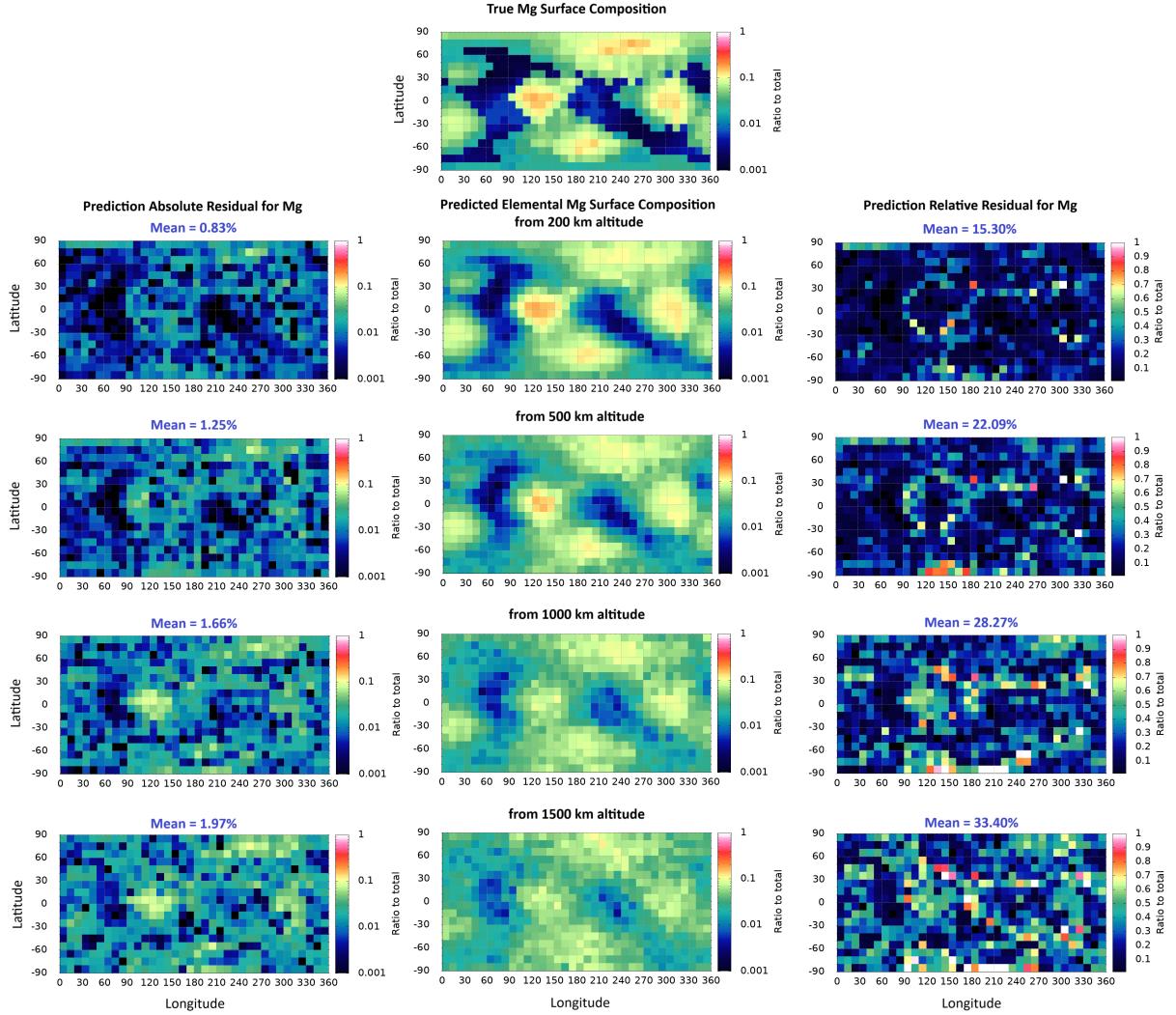
daylight observations from two consecutive Mercury years, leveraging the planet's spin-orbit resonance. This approach significantly improved predictions for volatile species and enabled a comprehensive reconstruction of Mercury's surface.

The developed algorithms for surface composition reconstruction achieved high fidelity in their predictions, with mean accuracy metrics reaching 89.70% (ES4) and 83.41% (R-squared) across 15 test sets at 200 km altitude. Additionally, both mean absolute and relative residuals of elemental predictions showed a robust trend of decreasing with altitude, reaching as low as



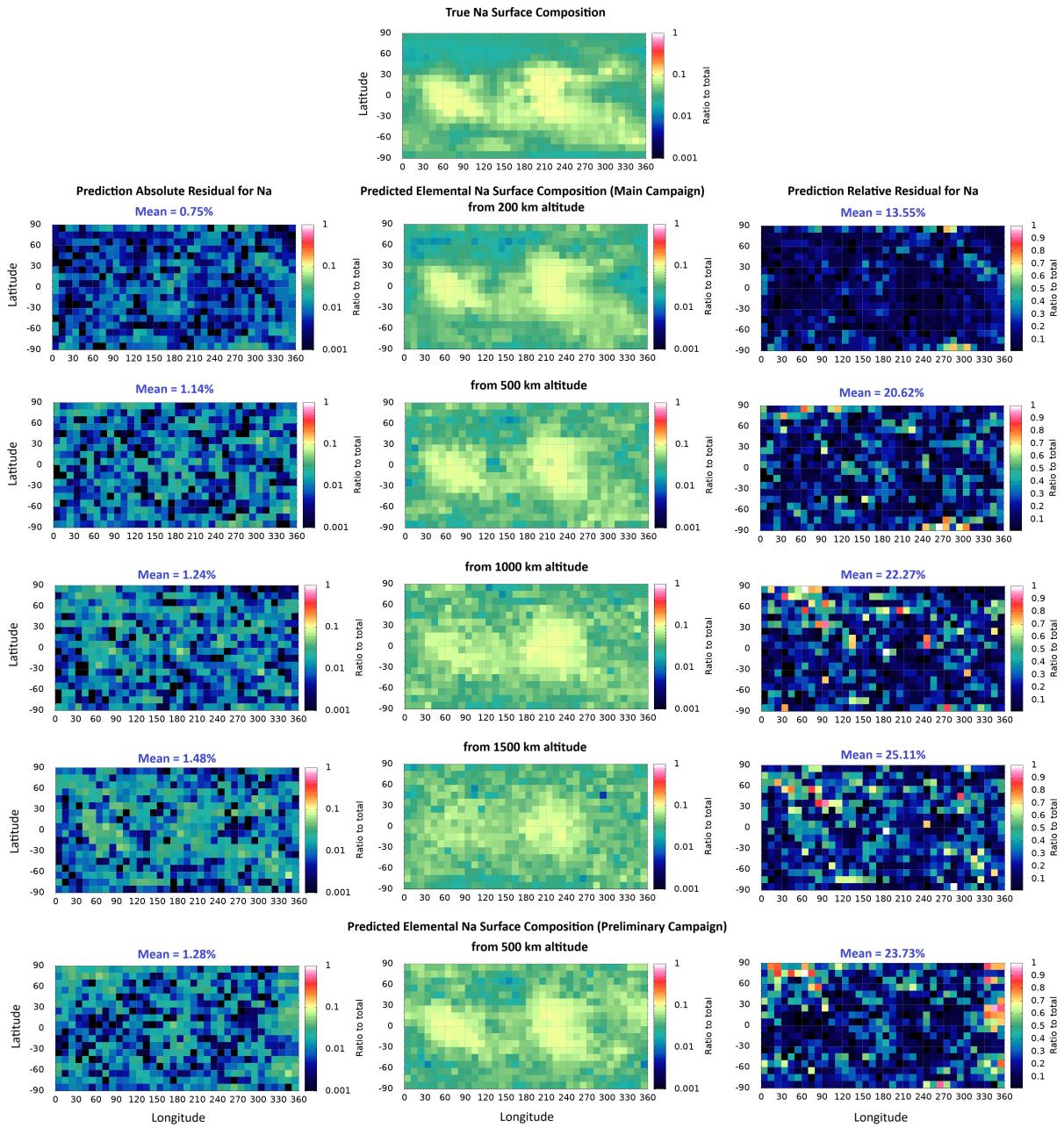
**Figure 27.** Main test campaign - MLP DNN map reconstructions of the same sample Potassium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "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.

1059 0.56% and 14.70% respectively. The effectiveness of these algorithms was particularly notable in  
 1060 reconstructing elemental composition maps from low-altitude measurements (200km to 700km).  
 1061 The extensive training and testing campaigns conducted in this study highlight significant  
 1062 potential for refining deep neural network algorithms used in surface reconstruction tasks.  
 1063 Acknowledging this potential, we have outlined a future development roadmap to advance these  
 1064 algorithms, particularly for application to measurements gathered in orbit around Mercury. This  
 1065 includes further exploration into feature engineering by applying more domain-specific knowledge

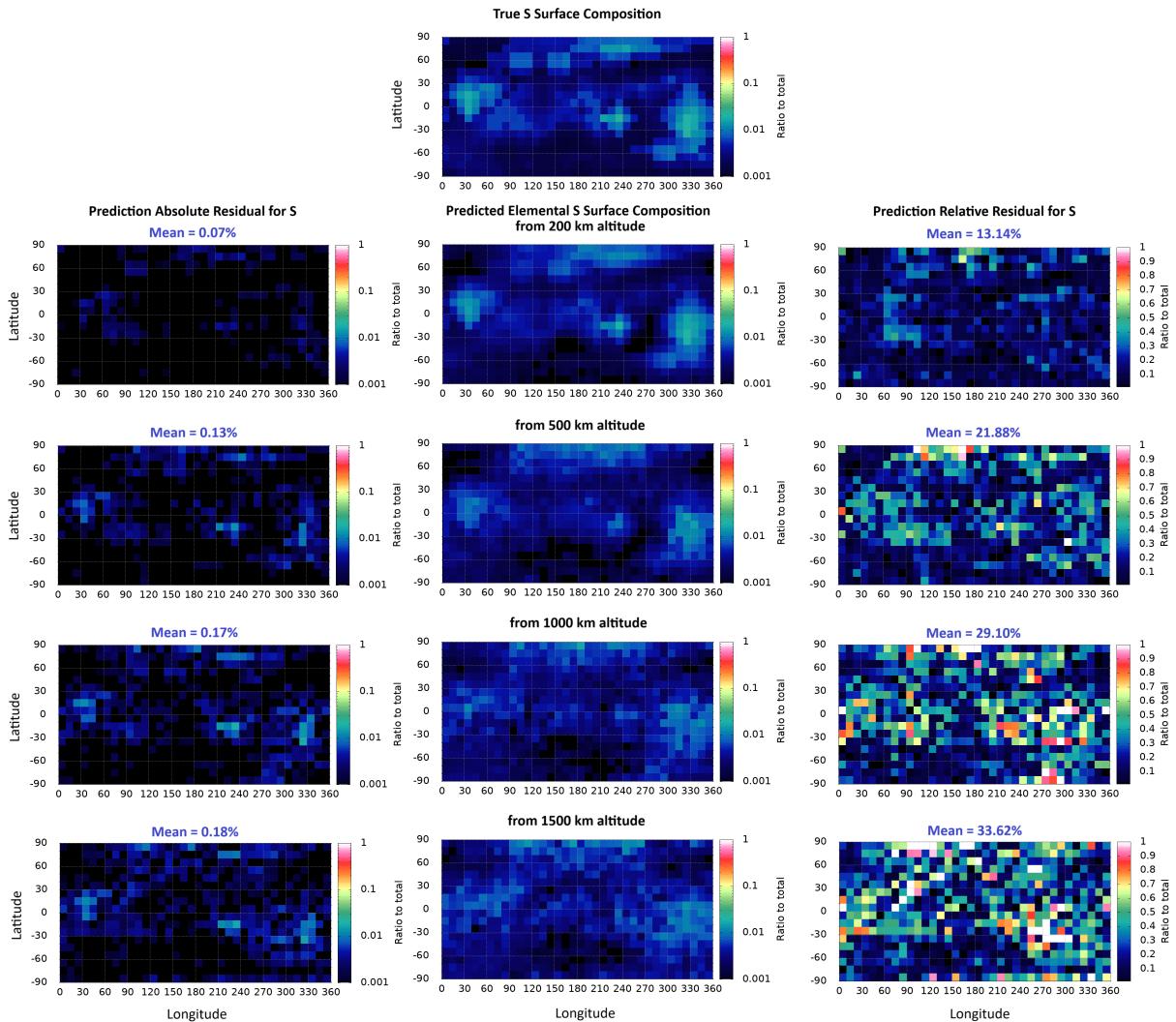


**Figure 28.** Main test campaign - MLP DNN map reconstructions of the same sample Magnesium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "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.

to optimize input parameters, providing a better representation of empirical data distributions.  
 Additionally, expanding and elaborating on the hyperparameter space is identified as a crucial next step. This will involve constructing a hyperparameter space that considers aspects such as network layer connectivity, optimization of loss functions, and the functions used within the hidden and output units. Exploring alternative DNN architectures also holds promise for enhancing the models' application, accuracy, and reliability. These developments aim to push the boundaries of what these algorithms can achieve in surface composition analysis.



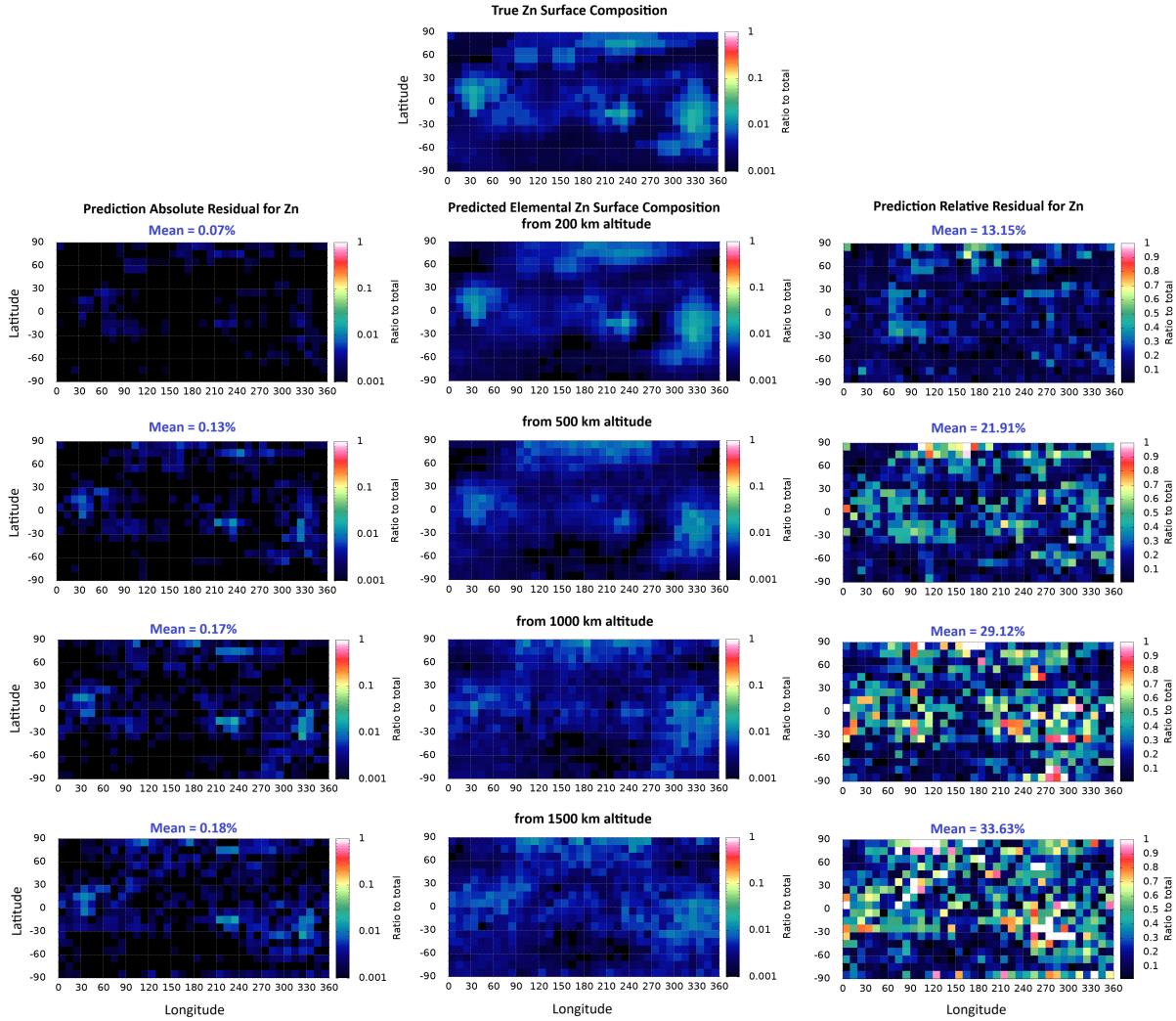
**Figure 29.** Main test campaign - MLP DNN map reconstructions of the same sample Sodium surface composition (test set number 2). Dayside only inputs of two simulated exospheres from consecutive Mercury perihelia at different altitude levels (200, 500, 1000, and 1500 km). The top-most map shows the "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.



**Figure 30.** 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.

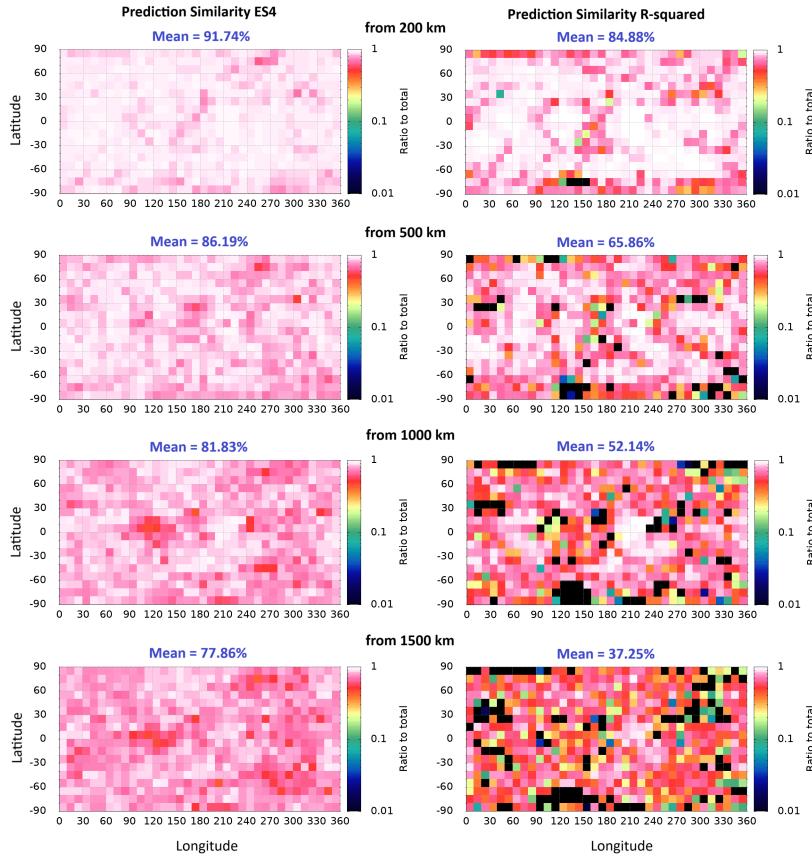
Another critical area of future research is testing our algorithms against data distributions from diverse regions within the physical processes parameter space. Such exploration is essential to assess the versatility and adaptability of the algorithms, ensuring they maintain high performance when applied to various models of exospheric production. Evaluating their efficacy in these contexts will reveal potential limitations and opportunities for improvements.

Moreover, expanding and refining the physical processes parameter space of our algorithm is crucial for future research. This endeavor will align the algorithm more closely with the complex



**Figure 31.** 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.

realities governing interactions between a planet's surface and its exosphere. A key part of this exploration includes delving into previously compressed dimensions within the parameter space, aiming to more accurately reflect the intricate dynamics of real-world processes. Unraveling these dimensions, previously simplified in our model, is expected to provide deeper insights into planetary science mechanisms and enhance the estimator's ability to capture the full spectrum of surface-exosphere interactions.



**Figure 32.** 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. *This figure needs to be polished a little bit - panel letters.*

1086     The contrast between using simulated data and incorporating real observational data  
 1087     into our algorithm development merits further exploration. While simulations provide a  
 1088     controlled environment for testing various scenarios, they do not capture the full complexity  
 1089     and unpredictability of actual exospheric data. This could lead to significant deviations  
 1090     in the parameter space from those assumed in our simulations, as real processes and their  
 1091     interdependent variables may change over time and are not fully represented in simulations.

1092     Shifting our focus from simulated to real physical processes is a bold and potentially  
 1093     transformative step. Developing an estimator capable of effectively processing and analyzing  
 1094     real-world data from Mercury’s exosphere would significantly advance our understanding of  
 1095     planetary surfaces and their interactions with their environments. This progress would not only

## 5 DISCUSSION AND FUTURE WORK

---

1096 deepen our theoretical knowledge but also offer practical insights into the formation, dynamics,  
1097 and evolution of planetary exospheres.

1098 This study is performed in anticipation of the upcoming BepiColombo ESA/JAXA mission.  
1099 The mission will deploy two spacecraft—the Mercury Planetary Orbiter (MPO) and the Mercury  
1100 Magnetospheric Orbiter (MMO)—equipped with a suite of instruments aimed at understanding  
1101 Mercury’s surface, exosphere, and magnetosphere (Benkhoff et al., 2010; Milillo et al., 2010,  
1102 2020). Particularly, our study targets future utilization of measurements from the SERENA  
1103 (Search for Exospheric Refilling and Emitted Natural Abundances) suite on the MPO, which  
1104 includes sensors like STROFIO (STart from a ROTating Field mass spectrOMeter), MIPA  
1105 (Miniature Ion Precipitation Analyser), PICAM (Planetary Ion CAMera), and ELENA (Emitted  
1106 Low Energy Neutral Atoms) to analyze exospheric gas composition, plasma fluxes, and ion  
1107 precipitation (Orsini et al., 2010, 2020; Milillo and Wurz, 2014). At the same time, notable  
1108 instruments aboard BepiColombo, which may provide images of the surface, include MIXS,  
1109 MGNS, MERTIS and SIMBIO-SYS (Benkhoff et al., 2010).

1110 The application of these methods to the observational data from BepiColombo’s suite of  
1111 instruments offers a promising path to refine models of exosphere generation. By comparing  
1112 predicted surface compositions with actual measurements, we can more accurately constrain  
1113 our models, enhancing our understanding of planetary processes.

1114 In conclusion, this research establishes a solid foundation for advancing our understanding of  
1115 planetary surface-exosphere interactions, particularly around Mercury. By utilizing exospheric  
1116 measurements as inputs to deep neural networks, we’ve taken a significant step forward,  
1117 enhancing the capabilities of estimators and broadening our understanding of planetary science.  
1118 Integrating this method into the SERENA ground system represents another notable advance  
1119 in space exploration. With sophisticated AI algorithms, the BepiColombo mission’s potential to  
1120 uncover insights into Mercury’s exosphere dynamics is greatly enhanced. Moreover, the ongoing  
1121 development and refinement of deep neural networks in this study promise to revolutionize  
1122 our approach to studying planetary bodies within our Solar System, providing new tools for  
1123 understanding the complex processes that govern the environments of celestial objects.

*1124* **Acknowledgments**

*1125* This work was supported by the Italian Space Agency (ASI) - SERENA contract 2018-8-HH.O

*1126* "Scientific participation in the mission BepiColombo SERENA - Phase E1". *Is this still the*

*1127* *right contract?*

1128 **Appendix A. Detailed Datasets**

1129 *Appendix A.1. Mean Mineral and Elemental Fractions in the Datasets*

1130 *Add table for the mean fractions for 10-300 training datasets, validation dataset, all 15 test*  
1131 *datasets.*

1132 *Appendix A.2. Consolidated Datasets*

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 A1.** 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 A2.** 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.

## APPENDIX B DETAILED TRAINING CAMPAIGN AND FINALIZATION OF THE DNN ARCHITECTURE

---

<sup>1133</sup> **Appendix B. Detailed Training Campaign and Finalization of the DNN**

<sup>1134</sup> **Architecture**

<sup>1135</sup> *Appendix B.1. Starting Point*

<sup>1136</sup> Our journey in the big training campaign endeavour commenced with a baseline training dataset,  
<sup>1137</sup> which, being the simplest, was least representative of the true data generation distribution.  
<sup>1138</sup> This dataset comprised 10 subsets with a total of 6,480 data points from the F00 feature set  
<sup>1139</sup> (only exospheric density measurements). The initial architecture of the MLP DNN was modeled  
<sup>1140</sup> closely after the structure proposed in the preliminary study of Kazakov et al. (2020), featuring  
<sup>1141</sup> a four-layered network with 400, 200, 200, and 100 neurons, respectively. Notably, our study  
<sup>1142</sup> expanded the input layer to accommodate a greater number of elements - 11 total input elements.

<sup>1143</sup> The output layer of the network employs softmax units, designed to predict the surface  
<sup>1144</sup> composition of the same 11 elements provided as inputs. Initially, the minibatch size was set to  
<sup>1145</sup> 1,024 examples. The regularization L-2 coefficient and the learning rate were chosen as  $1.0 \times 10^{-6}$   
<sup>1146</sup> and  $0.5 \times 10^{-4}$ , respectively, to balance the trade-off between learning efficiency and the risk of  
<sup>1147</sup> overfitting.

<sup>1148</sup> *Appendix B.2. Eliminating Skewed Predictions*

<sup>1149</sup> The initial analysis of predicting fractionated surface elemental composition revealed that the  
<sup>1150</sup> accuracy metrics were significantly skewed by the prevalence of certain abundant elements,  
<sup>1151</sup> notably oxygen ( $O_2$ ) and silicon (Si), which are omnipresent in most of the minerals in our model.  
<sup>1152</sup> This skewness, stemming from the algorithm's propensity to more accurately predict these two  
<sup>1153</sup> elements, was addressed by excluding them from the prediction vector and adjusting it to ensure  
<sup>1154</sup> a normalized sum of 1. Consequently, the refined model focuses on predicting the normalized  
<sup>1155</sup> proportions of the remaining nine elements, with a subsequent denormalization process applied  
<sup>1156</sup> for the map reconstruction purposes. This strategic exclusion of the most abundant elements  
<sup>1157</sup> led to a marked enhancement of approximately 4% in the predictive  $R^2$  accuracy for the other  
<sup>1158</sup> nine elements. It is important to note, however, that the input vector maintained its original  
<sup>1159</sup> composition of 11 elements.

<sup>1160</sup> This decision to modify the output layer by removing two elements was driven by a clear  
<sup>1161</sup> rationale: the omnipresent elements, though significant, held less interest for the objectives  
<sup>1162</sup> of our study compared to the other elements. This approach underscores our commitment to

1163 optimizing the model's performance where it matters most, despite recognizing that alternative  
1164 configurations of the output layer might exist.

1165 *Appendix B.3. Training Set Size and Learning Curve Examination*

1166 Exploring the behavior of the initial MLP DNN involved examining its performance in relation  
1167 to the expansion of the training dataset size and the extension of training duration. The aim was  
1168 to demonstrate the algorithm's nominal operation during both training and inference phases by  
1169 analyzing its learning curves. This included assessing training and generalization accuracies  
1170 across a training dataset and the hold-out validation dataset, respectively. Additionally,  
1171 identifying the optimal training duration for inference was crucial to mitigate the risk of  
1172 overfitting, in line with the guidance provided by Bengio (2015).

1173 Our investigation spanned training sets ranging from 10 to 200 unaugmented data subsets. We  
1174 observed a clear positive relationship between increasing the dataset size and the enhancement  
1175 of generalization accuracy. As anticipated, this expansion led to a decrease in training accuracy,  
1176 a phenomenon depicted in Figure ??.

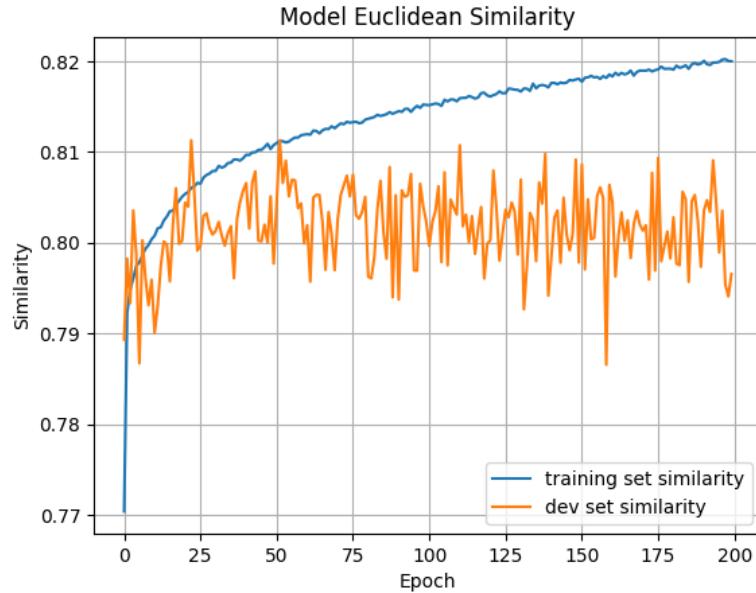
1177 Add figure that shows the increasing generalization accuracy with increasing training set size.

1178 In parallel, the algorithm's behavior was monitored in terms of its optimization process over  
1179 multiple iterations (epochs) across the entire training dataset, employing stochastic gradient  
1180 descent to converge to the minimum of the loss function. Analysis of the learning curves  
1181 revealed a maximum in predictive performance on the validation dataset after 40 epochs. This  
1182 was in contrast to the outcomes observed at 200 epochs of SGD, despite the training dataset's  
1183 distribution increasingly aligning with each additional training iteration.

1184 The learning curve depicted in Figure B1 also hints at the potential for further enhancements  
1185 in training predictions, given the rapid ascent observed towards the training's culmination.  
1186 However, to ensure robust inference capabilities, it's imperative to diminish the variance. This  
1187 could potentially be achieved by incorporating a greater number of training examples and/or  
1188 intensifying the regularization measures.

1189 *Appendix B.4. Feature Sets Examination*

1190 The evaluation of feature sets played a pivotal role in optimizing the performance of the MLP  
1191 DNN, particularly through the training of the network with various input features across the  
1192 unaugmented 200-subset training dataset. The assortment of the examined feature sets is



**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.

1193 detailed in Table B1. The iterative process of enhancing input features resulted in substantial  
 1194 improvements in prediction accuracy for several modified feature sets with ES4 going up by  
 1195 about 2%, while  $R^2$  increasing with as much as 5% from F00 to F11. This underscores the  
 1196 critical importance of a well-curated and comprehensive feature set in the development of  
 1197 neural networks capable of tackling intricate tasks, such as predicting the surface elemental  
 1198 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.

## APPENDIX B DETAILED TRAINING CAMPAIGN AND FINALIZATION OF THE DNN

### Appendix B.5 Hyperparameter Optimization and DNN Structural Components Finalization

After rigorous testing and evaluation, the feature set that emerged as superior, offering the most consistent and highest accuracy, was F11. This feature set encompasses: (1) logarithmic transformations of elemental species exospheric densities, which provide a normalized scale for comparing densities of various elements; (2) the logarithm of the altitude at which measurements were taken, introducing a scale that accommodates the wide range of altitudes without skewing the data; (3) Sun incidence angle, accounting for the variation in solar energy impacting the elemental composition; (4) the presence of H+ ions arriving through open field lines, a feature indicating solar wind interaction with the planetary surface; (5) Cosine of solar time longitude, offering a representation of the position in solar time longitude; and (6) Sine of latitude, providing a function to capture latitudinal variations.

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

#### Appendix B.5. Hyperparameter Optimization and DNN Structural Components Finalization

There was actually an initial hyperparameter search performed at 60 unaugmented datasets producing 5-6 local minima in the evaluation on the validation set, therefore 5-6 distinct MLP DNNs showing promising results. How to explain this, though?

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

During this campaign, the selection of loss functions emerged as a pivotal consideration, with

1229 our experiments revealing substantial variations in model performance across different functions.

1230 The discerning application of loss functions, particularly the adoption of the KL-divergence for

1231 evaluating probability-like outputs, marked a jump in performance.

1232 The culmination of our hyperparameter optimization efforts led to the finalization of the MLP

1233 DNN architecture, characterized by a four-layered structure with 600, 500, 350, and 250 neural

1234 units respectively (Figure 9). An adjustment was made to the regularization coefficient, setting

1235 it to the found higher value of  $1.0 \times 10^{-5}$  to enhance model generalization. Concurrently, the

1236 learning rate was optimized to  $0.5 \times 10^{-4}$ , balancing the trade-off between learning speed and

1237 stability. Training was conducted in mini-batches of 512 examples, a size determined through

1238 our optimization exercises to be close to optimal. This meticulously optimized structure and

1239 parameter set represent the culmination of our comprehensive campaign to refine the MLP DNN,

1240 ensuring it stands as a robust model for our advanced predictive task.

1241 *Appendix B.6. Augmented Data Study*

1242 In the concluding phase of our training campaign, we embarked on a strategic initiative to

1243 enhance the representability of the empirical distribution, thereby aligning it more closely with

1244 the true data-generating distribution—a target that remains inherently elusive due to limited

1245 direct access. This endeavor was pursued through the deliberate augmentation of our training

1246 datasets, an approach that involved the integration of additional examples derived from the same

1247 exospheric observations that constituted our initial datasets. However, these new inclusions were

1248 distinct in their representation of varying altitudes, thereby enriching the diversity and depth

1249 of our training data.

1250 The initial expansion of our dataset to encompass 200 augmented subsets had already

1251 demonstrated significant promise in enhancing the model’s performance. Motivated by these

1252 preliminary successes, we ambitively expanded our dataset even further to include a total of

1253 300 augmented subsets, culminating in an impressive 1,360,800 examples. This substantial

1254 augmentation effort was driven by the rationale that incorporating measurements from varying

1255 altitudes would not only bolster the dataset’s comprehensiveness but also empower our model

1256 to predict with greater accuracy across a diverse range of altitude-specific inputs.

1257 The outcome of this labor were unmistakably positive, with the augmented datasets markedly

1258 improving the robustness and accuracy of our MLP DNN, increasing further the validation set’s

1259 ES4 to 84.0% (+1.5%) and its  $R^2$  to 63.5% (+3.5%). The strategic inclusion of altitude-varied  
1260 examples was particularly impactful, enabling the algorithm to achieve enhanced predictive  
1261 precision for inputs across different altitudes.

1262 *Appendix B.7. Implications and Results of the Training*

1263 As our meticulous exploration of the hyperparameter space culminated in identifying a region  
1264 that, while not conclusively the ultimate minimum, demonstrates unparalleled accuracy in  
1265 inferences on the hold-out validation dataset, we arrived at several pivotal implications and  
1266 results from our training campaign. This journey through hyperparameter optimization has  
1267 yielded a collection of finely tuned multilayer perceptron deep neural networks, each reflecting  
1268 a nuanced understanding of the underlying data-generating processes.

1269 Firstly, one of the outcomes of this campaign is the demonstration of the algorithm's efficiency,  
1270 achieving optimal training within 40 complete epochs. This not only highlights the effectiveness  
1271 of our chosen architecture but also underscores the potential for accuracy improvements with  
1272 the expansion of the training dataset. Such findings affirm the architectural decisions made in  
1273 designing our MLP DNN for the task at hand.

1274 Secondly, our investigation revealed the critical role of specific features in guiding the  
1275 algorithm toward more precise predictions of exospheric measurements and surface composition.  
1276 The identification of these key features underscores the importance of thoughtful feature selection  
1277 in enhancing model performance.

1278 Thirdly, the exploration led to the refinement of the MLP's internal structure, significantly  
1279 bolstered by experiments with various loss functions and output layers, alongside the application  
1280 of Bayesian hyperparameter optimization. While acknowledging that the realm of possible  
1281 architectural enhancements remains vast, the current configuration stands as a testament to the  
1282 robustness and efficacy of our model.

1283 Lastly, the strategic augmentation of our dataset with additional exospheric observations  
1284 has unequivocally improved the algorithm's predictive capabilities. This expansion not only  
1285 enriches the model's training environment but also enhances its ability to generalize across a  
1286 broader spectrum of the empirical distribution, thereby moving closer to the elusive true data-  
1287 generating distribution.

1288 The combined efforts of hyperparameter exploration, architectural fine-tuning, and dataset

1289 augmentation have significantly propelled our model's performance. Through this comprehensive  
1290 training campaign, we have not only achieved a high degree of accuracy in our predictions  
1291 but also laid a solid foundation for future research to build upon, promising even greater  
1292 advancements in our understanding and representation of complex data-generating processes.  
1293 Through this concerted effort, we have significantly advanced the model's capacity to generalize  
1294 from the empirical distribution to the true underlying data-generating distribution.

1295 Add table of trainings?

1296 How many total training runs?

1297 How much total time and average time?

1298 Add some other training graphs and information here.

- 1299 **References**
- 1300 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis,  
1301 A. and Dean, J., Devin, M., et al. (2015). Tensorflow: Large-scale machine learning on  
1302 heterogeneous systems. *Software available from tensorflow.org*.
- 1303 Bengio, Y. (2015). Early inference in energy-based models approximates back-propagation.  
1304 Technical Report arXiv:1510.02777, Université de Montréal.
- 1305 Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M., Ferri,  
1306 P., Middleton, H. R., and Ziethe, R. (2010). Bepicolombo - comprehensive exploration of  
1307 mercury: Mission overview and science goals. *Planetary and Space Science*, 58(1):2–20.
- 1308 Berezhnoy, A. (2018). Chemistry of impact events on mercury. *Icarus*, 300:210–222.
- 1309 Bergstra, J., Bardenet, R., Bengio, Y., and Balazs, K. (2011). Algorithms for hyper-parameter  
1310 optimization. In *Proceedings of the 24th International Conference on Neural Information  
1311 Processing Systems*, pages 2546–2554.
- 1312 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 1313 Cassidy, T. A., Merkel, A. W., Burger, M. H., Sarantos, M., Killen, R. M., McClintock,  
1314 W. E., and Vervack, R. J. (2015). Mercury’s seasonal sodium exosphere: Messenger orbital  
1315 observations. *Icarus*, 248:547–559.
- 1316 Chollet, F. et al. (2015). Keras. *Github. GitHub repository*.
- 1317 Cintala, M. J. (1992). Impact-induced thermal effects in the lunar and mercurian regoliths. *J.  
1318 Geophys. Res.*, 97:947–973.
- 1319 Ciresan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep big simple  
1320 neural nets for handwritten digit recognition. *Neural Computation*, 22:1–14.
- 1321 Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-Interscience.
- 1322 Domingue, D. L., Koehn, P. L., Killen, R. M., et al. (2007). Mercury’s atmosphere: A surface-  
1323 bounded exosphere. *Space Science Reviews*, 131(1-4):161–186.
- 1324 Gamborino, D., Vorburger, A., and Wurz, P. (2019). Mercury’s subsolar sodium exosphere:  
1325 An ab initio calculation to interpret mascs/uvvs observations from messenger. *Annales  
1326 Geophysicae*, 37:455–470.
- 1327 Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In

---

**REFERENCES**

---

**REFERENCES**

- 1328     *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*,  
1329     volume 15, pages 315–323.
- 1330     Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge,  
1331     MA.
- 1332     Grava, C., Killen, R. M., Benna, M., Berezhnoy, A. A., Halekas, J. S., Leblanc, F., Nishino,  
1333     M. N., Plainaki, C., Raines, J. M., Sarantos, M., Teolis, B. D., Tucker, O. J., Vervack, R. J.,  
1334     and Vorburger, A. (2021). Volatiles and refractories in surface-bounded exospheres in the  
1335     inner solar system. *Space Science Reviews*, 217(5):61.
- 1336     Head, T. et al. (2018). scikit-optimize/scikit-optimize: v0.5.2. Zenodo.
- 1337     Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*,  
1338     11(10):428–434.
- 1339     James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical  
1340     Learning: with Applications in R*. Springer.
- 1341     Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory  
1342     and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.
- 1343     Johnson, R., Leblanc, F., Yakshinskiy, B., and Madey, T. (2002). Energy distributions for  
1344     desorption of sodium and potassium from ice: the na/k ratio at europa. *Icarus*, 156:136–  
1345     142.
- 1346     Kazakov, A. N. et al. (2020). Deep neural networks for analysis of mercury’s planetary exosphere.  
1347     *Journal of Physics: Conference Series*, 1548:012–014.
- 1348     Killen, R. (2016). Pathways for energization of ca in mercury’s exosphere. *Icarus*, 268:32–36.
- 1349     Killen, R. M. and Burger, M. H. (2019). Understanding mercury’s exosphere: Models derived  
1350     from messenger observations. In *Mercury: The View After MESSENGER*. Cambridge  
1351     University Press, London.
- 1352     Killen, R. M., Cremonese, G., Lammer, H., et al. (2007). Processes that promote and deplete  
1353     the exosphere of mercury. *Space Science Reviews*, 132(2-4):433–509.
- 1354     Killen, R. M., Potter, A. E., Reiff, P., Sarantos, M., Jackson, B. V., Hick, P., and Giles, B.  
1355     (2001). Evidence for space weather at mercury. *J. Geophys. Res.*, 106:20509–20525.
- 1356     Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. In *International  
1357     Conference on Learning Representations*.

---

*REFERENCES*

---

*REFERENCES*

- 1358 Leblanc, F., Sarantos, M., Domingue, D., Milillo, A., Savin, D., Prem, P., Benkhoff, J., Zender,  
1359 J., Galli, A., Murakami, G., Sasaki, S., Thompson, M., and Raines, J. (2023). How does  
1360 the thermal environment affect the exosphere/surface interface at mercury? *The Planetary  
1361 Science Journal*, 4:227.
- 1362 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.
- 1363 Mangano, V. et al. (2015). Themis na exosphere observations of mercury and their correlation  
1364 with in-situ magnetic field measurements by messenger. *Planetary and Space Science*, 115:102–  
1365 109.
- 1366 Merkel, A. W., Vervack, R. J. J., Killen, R. M., Cassidy, T. A., McClintonck, W. E., Nittler,  
1367 L. R., and Burger, M. H. (2018). Evidence connecting mercury’s magnesium exosphere to its  
1368 magnesium-rich surface terrane. *Geophysical Research Letters*, 45:6709–6797.
- 1369 Milillo, A., Fujimoto, M., Kallio, E., Kameda, S., Leblanc, F., Narita, Y., Cremonese, G., Laakso,  
1370 H., Laurenza, M., Massetti, S., McKenna-Lawlor, S., Mura, A., Nakamura, R., Omura, Y.,  
1371 Rothery, D., Seki, K., Storini, M., Wurz, P., Baumjohann, W., and Sprague, A. (2010).  
1372 The bepicolombo mission: An outstanding tool for investigating the hermean environment.  
1373 *Planetary and Space Science*, 58:40–60.
- 1374 Milillo, A., Fujimoto, M., Murakami, G., et al. (2020). Investigating mercury’s environment  
1375 with the two-spacecraft bepicolombo mission. *Space Science Reviews*, 216. Article no. 93.
- 1376 Milillo, A., Sarantos, M., Grava, C., Janches, D., Lammer, H., Leblanc, F., Schorghofer, N.,  
1377 Wurz, P., Teolis, B., and Murakami, G. (2023). Future directions for the investigation of  
1378 surface-bounded exospheres in the inner solar system. *Space Science Reviews*, 219.
- 1379 Milillo, A. and Wurz, P. (2014). *SERENA Science Performance Report*.
- 1380 Milillo, A., Wurz, P., Orsini, S., Delcourt, D., Kallio, E., Killen, R., Lammer, H., Massetti,  
1381 S., Mura, A., Barabash, S., Cremonese, G., Daglis, I., De Angelis, E., Di Lellis, A., Livi, S.,  
1382 Mangano, V., and Torkar, K. (2005). Surface-exosphere-magnetosphere system of mercury.  
1383 *Space Science Reviews*, 117:397–443.
- 1384 Minsky, M. and Papert, S. A. (2017). *Perceptrons: An Introduction to Computational Geometry*.  
1385 MIT Press, Cambridge, MA.
- 1386 Mura, A., Milillo, A., Orsini, S., and Massetti, S. (2007). Numerical and analytical model of

---

*REFERENCES*

---

*REFERENCES*

- 1387 mercury's exosphere: Dependence on surface and external conditions. *Planetary and Space*  
1388 *Science*, 55:1569–1583.
- 1389 Mura, A., Wurz, P., Lichtenegger, H., Schleicher, H., Lammer, H., Delcourt, D., Milillo, A.,  
1390 Orsini, S., Massetti, S., and Khodachenko, M. (2009). The sodium exosphere of mercury:  
1391 Comparison between observations during mercury's transit and model results. *Icarus*, 200:1–  
1392 11.
- 1393 Orsini, S., Livi, S., Lichtenegger, H., et al. (2020). Serena: Particle instrument suite for sun-  
1394 mercury interaction insights on-board bepicolombo. *Space Science Reviews*. In Press.
- 1395 Orsini, S., Livi, S., Torkar, K., Barabash, S., Milillo, A., Wurz, P., Di Lellis, A. M., and Kallio,  
1396 E. (2010). Serena: A suite of four instruments (elena, strofio, picam and mipa) on board  
1397 bepicolombo-mpo for particle detection in the hermean environment. *Planetary and Space*  
1398 *Science*, 58:166–181.
- 1399 Plainaki, C., Mura, A., Milillo, A., Orsini, S., Livi, S., Mangano, V., Massetti, S., Rispoli, R.,  
1400 and De Angelis, E. (2017). Investigation of the possible effects of comet encke's meteoroid  
1401 stream on the ca exosphere of mercury. *Journal of Geophysical Research: Planets*, 122:1217–  
1402 1226.
- 1403 Pokorný, P., Sarantos, M., and Janches, D. (2017). Reconciling the dawn–dusk asymmetry  
1404 in mercury's exosphere with the micrometeoroid impact directionality. *The Astrophysical*  
1405 *Journal Letters*, 842(2):L17.
- 1406 Rothery, D. A., Massironi, M., Alemanno, G., et al. (2020). Rationale for bepicolombo studies  
1407 of mercury's surface and composition. *Space Science Reviews*, 216:66.
- 1408 Rumelhart, D., Hinton, G., and Williams, R. (1986a). Learning representations by back-  
1409 propagating errors. *Nature*, 323:533–536.
- 1410 Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning internal representations  
1411 by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed*  
1412 *Processing*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge.
- 1413 Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, 3rd edition*.  
1414 Prentice Hall Press, USA.
- 1415 Sarantos, M., Slavin, J., Benna, M., Boardsen, S., Killen, R., Schriver, D., and Trávníček, P.

*REFERENCES*

---

*REFERENCES*

- 1416      (2009). Sodium-ion pickup observed above the magnetopause during messenger's first mercury  
1417      flyby: Constraints on neutral exospheric models. *Geophysical Research Letters*, 36.
- 1418      Schaible, M. J. et al. (2017). Solar wind sputtering rates of small bodies and ion mass  
1419      spectrometry detection of secondary ions. *Journal of Geophysical Research: Planets*, 122:1968–  
1420      1983.
- 1421      Wurz, P. and Lammer, H. (2003). Monte-carlo simulation of mercury's exosphere. *Icarus*,  
1422      164:1–13.
- 1423      Wurz, P., Whitby, J. A., Rohner, U., Martin-Fernandez, J. A., Lammer, H., and Kolb, C. (2010).  
1424      Self-consistent modelling of mercury's exosphere by sputtering, micrometeorite impact and  
1425      photon-stimulated desorption. *Planetary and Space Science*, 58:1599–1616.