

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

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

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

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

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

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

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

¹⁶ E-mail: adrian.kazakov@inaf.it

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

orbital run through Mercury's exosphere, as inputs and predicts the chemical elements of the surface regolith below. It serves as an estimator for the surface-exosphere interaction and the processes leading to exosphere formation, which, in our simulated setup, include micrometeoroid impact vaporization (MIV), ion sputtering (SP), photon-stimulated desorption (PSD), and thermal desorption (TD). Extensive training and testing campaigns demonstrate the 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, are manifest in the composition and dynamics of 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; Killen and Burger, 2019; Gamborino et al., 2019; Wurz et al., 2010). Four predominant processes release atoms and molecules from the surface into the exosphere: micrometeoroid impact vaporization (MIV), sputtering after solar wind and heavy

57 ion impacts (SP), thermal desorption (TD), and photon-stimulated desorption (PSD). MIV
58 and SP are particularly indicative of the regolith composition below, as they involve higher
59 energy transfers capable of dislodging neutral species from their minerals. Conversely, TD
60 and PSD, being less energetic, tend to release atoms and molecules that are weakly bonded
61 to minerals, such as volatile elements, most of which eventually fall back and are reabsorbed
62 by the surface (Killen et al., 2007; Mura et al., 2009; Leblanc et al., 2003; Gamborino et al.,
63 2019). Once in the exosphere, the released particles undergo further transformations due to
64 interactions with radiation pressure, photons, and charged particles. Such interactions can
65 modify the charge, chemical state, and movement of these exospheric constituents. However, in
66 a first approximation in the sparse exosphere, the atomic and molecular abundances resulting
67 from these actors could be traced back to the planet, connecting the surface properties, like
68 composition, mineralogy, and physical state to the different processes and the dynamics of
69 matter around the planet (Milillo et al., 2020; Rothery et al., 2020). There has already been
70 direct evidence that this is the case for the Magnesium exosphere, which is directly related to the
71 Magnesium-rich surface below, as shown by Merkel et al. (2018). [Possibly cite an implication](#).

72 To gain a deeper understanding of Mercury's exosphere, scientists use sophisticated modeling
73 techniques to simulate the various active processes and their effects on the planetary surface,
74 thereby attempting to replicate the generation of the exosphere. This extensively applied method
75 compares the results of simulations to those measured from space (Sarantos et al., 2009; Cassidy
76 et al., 2015; Plainaki et al., 2017) or from Earth (Mura et al., 2007; Mangano et al., 2015; Wurz
77 and Lammer, 2003). The models utilize mathematical relations about the physical processes
78 in action to produce the exosphere, drawing upon our knowledge from observations and/or
79 laboratory experiments. However, the inherent complexity of these interactions, which includes
80 electromagnetic, chemical, mechanical, thermal, and other effects, adds significant challenges.
81 Some of these effects have not been precisely evaluated for each release process, leading to a
82 broad range of simulated results with considerable uncertainty, depending on the assumptions
83 made at the outset.

84 This uncertainty is further compounded by the multidimensional parameter space each model
85 relies on. Each dimension in this space corresponds to a variable, and collections of variables
86 represent distinct processes active in the interaction. The lack of precise or complete data results
87 in a wide range of possible values for each parameter, thereby expanding the uncertainty within

88 the overall parameter space and often yielding a spectrum of potential outcomes rather than
89 definitive predictions. Additionally, the validity of the chosen parameters and the representation
90 of all relevant physical processes pose significant challenges. To address these, a multifaceted
91 approach is required, involving the refinement of models through improved measurements,
92 continual reassessment of the model structure, and advanced statistical methods to better
93 understand and quantify uncertainties.

94 In parallel, machine learning algorithms, particularly deep neural networks (DNNs), offer
95 a novel approach to capture the highly nonlinear relationships between the variables. These
96 algorithms can resolve, to some extent, the data generation mechanisms (Russell and Norvig,
97 2009; LeCun et al., 2015; Goodfellow et al., 2016), providing a tool to explore in depth the
98 relationships between the components of Mercury’s environment. This work will demonstrate
99 how DNNs, especially multilayer perceptrons (MLPs), can be employed within the data analysis
100 of Mercury’s exosphere to reconstruct the elemental surface map underneath. Suitable for
101 nonlinear regression tasks, DNNs scale effectively with increasing training data and input
102 parameters (Minsky and Papert, 2017; Hinton, 2007; Ciresan et al., 2010), offering a promising
103 direction for tackling the complexities inherent in modeling Mercury’s exosphere.

104 Building upon the foundational work of Kazakov et al. (2020), this study extends, refines
105 and advances further the application of deep neural networks in predicting Mercury’s mineral
106 composition from exospheric measurements. The initial method, presented in Kazakov et al.
107 (2020), employed a highly simplified model focusing solely on the generation of the exosphere
108 through micrometeoroid impact vaporization, presumed to be the predominant release process
109 on the planet’s night side. This approach, while innovative, was based on assumptions and
110 models that were overly simplistic, limiting the applicability of the results for broader analysis.
111 Despite these constraints, the basic DNNs employed in their study demonstrated a moderately
112 accurate performance in mineralogy reconstruction with the limited dataset available.

113 In contrast, the current research represents a more advanced continuation of this method,
114 involving a thorough investigation of improved algorithms applied to data generated under more
115 complex and plausible assumptions for the exospheric models. We have expanded the algorithm’s
116 application to include the prediction of elemental composition, utilizing data from models that
117 simulate all major processes – MIV, SP, TD, and PSD – along with an exosphere influenced by
118 solar radiation pressure and photolysis. The MLPs underwent extensive optimization, exploring

119 a wide range of their hyperparameters. The prediction task was more precisely defined to
120 predict elemental composition from datasets enhanced with additional features and augmented
121 with more examples.

122 Building upon the simpler models of the previous work, this study refines the methodological
123 approach, thereby improving the predictive accuracy and reliability of the DNNs employed.
124 Notably, the DNNs in this study serve as an effective representation of a specific region of the
125 physical parameter space, learning to map the complex relationships between inputs (exospheric
126 composition) and outputs (surface composition). This learning process yields an estimator
127 for the processes governing the exosphere's generation, aiming to bridge the divide between
128 theoretical modeling and practical, data-driven predictions. Enhancing our understanding of
129 these intricate interactions, the development of this method holds potential not only for refining
130 the models of exospheric generation but also for providing valuable insights into the interaction
131 between Mercury's surface and its exosphere. This, in turn, not only deepens our understanding
132 of planetary exospheres but also unlocks a multitude of opportunities for further exploration
133 and discovery in both the realms of planetary science and machine learning, paving the way for
134 innovative research directions.

135 This study is performed in anticipation of the upcoming BepiColombo ESA/JAXA mission,
136 which, with its two spacecraft - the Mercury Planetary Orbiter (MPO) and the Mercury
137 Magnetospheric Orbiter (MMO) - presents a unique opportunity to study the planet from two
138 different vantage points with a variety of instruments and sensors (Milillo et al., 2020). The
139 investigation of the highly complex interaction between the outside environment and Mercury's
140 surface, exosphere, and magnetosphere is one of the main objectives of the BepiColombo mission
141 to the innermost planet. This mission could reveal not only insights into Mercury itself but also a
142 broader perspective on the forces shaping our Solar System (Benkhoff et al., 2010; Milillo et al.,
143 2010). The development of the deep neural networks and the subsequent scientific analysis
144 are envisioned as a part of the Ground System of the suite of neutral and charged particle
145 detectors SERENA (Search for Exospheric Refilling and Emitted Natural Abundances) on-board
146 the MPO. In particular, the study is directed towards the future utilization of measurements
147 from the SERENA sensors: the mass spectrometer STROFIO (STart from a ROTating Field
148 mass spectrometer), which will measure the exospheric gas composition; MIPA (Miniature Ion
149 Precipitation Analyser) and PICAM (Planetary Ion CAMera), which will detect plasma fluxes

from the Solar Wind and magnetosphere toward the surface; and ELENA (Emitted Low Energy Neutral Atoms), providing the ion precipitation map onto the surface through back-scattered neutral detection (Orsini et al., 2010, 2020; Milillo and Wurz, 2014). At the same time, notable instruments aboard BepiColombo, which may provide images of the surface, include **MIXS**, **MGNS**, **MERTIS** and **SIMBIO-SYS** (Benkhoff et al., 2010).

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

2. Method

2.1. Prediction Task and General Characteristics of the Method

In this study, we extend the algorithm initially proposed by Kazakov et al. (2020), applying it to supervised multivariate regression of exospheric data at Mercury using a multilayer perceptron deep neural network. The objective of this enhanced DNN is to infer the regolith source material, believed to be a primary contributor to Mercury's exosphere. Specifically, the DNN predicts elemental surface composition fractions from exospheric density measurements, governed by the equation:

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

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

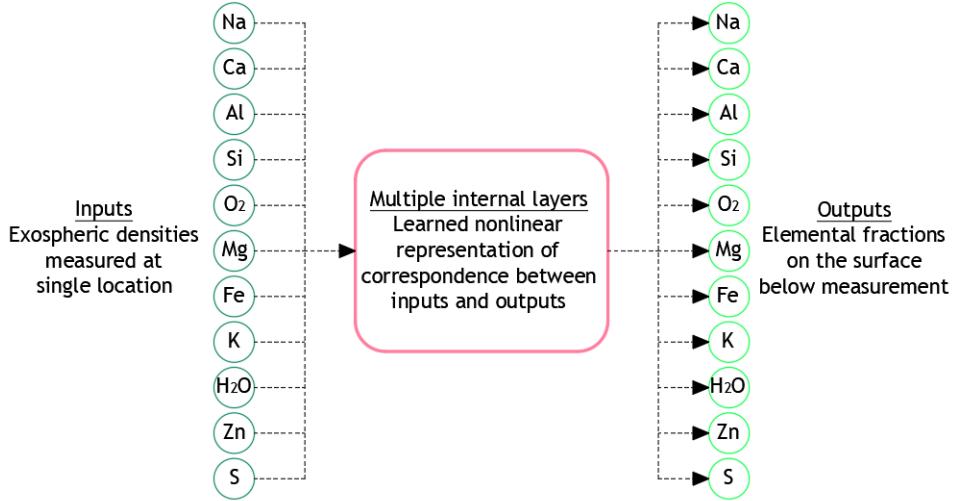


Figure 1. DNN prediction task schematics. The input to the neural network on the left are the exospheric densities at a single location in the exosphere. The output is the relative surface elemental composition as fractions summing up to 1 at a surface area just below the exospheric measurement. The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

Significant methodological advancements in this study include:

- Extending the algorithm's application to elemental composition prediction.
- Developing DNNs capable of operating in a multi-process environment, considering all four primary active processes (MIV, SP, TD, PSD) for neutral species release.
- Implementing a more sophisticated and realistic data production model with refined assumptions for exosphere generation processes.
- Enhancing feature definition to improve algorithmic inference, accommodating measurements at varying altitudes and illumination conditions.
- Training the algorithms with augmented data sets.
- Employing hyperparameter tuning to optimize the DNN design parameters.

- Conducting a comprehensive exploration of the DNN hyperparameter space.
- Investigating the formation of the physical processes parameter space.

The MLP, a class of feedforward artificial neural network, is particularly effective in multivariate regression tasks. It comprises an input layer, several hidden layers, and an output layer, with each neuron in one layer fully connected to every neuron in the subsequent layer. The MLP's strength in regression tasks stems from its ability to model complex, nonlinear functions through its layered structure and nonlinear activation functions (Minsky and Papert, 2017). This architecture enables MLPs to discern complex patterns in multi-dimensional data, a typical scenario in multivariate regression. Enhanced by backpropagation (Rumelhart et al., 1986a) and various optimization algorithms (Kingma and Ba, 2014), MLPs excel in training network weights, ensuring precise and robust predictive models. Their versatility makes them well-suited for modeling complex patterns in high-dimensional data (LeCun et al., 2015).

The training of these networks involves multiple phases, including data preprocessing, model building, and iterative training with backpropagation, performed in our case using the Keras programming framework with the TensorFlow backend (Abadi et al., 2015). Leveraging Python-based algorithms, our approach benefits from the flexibility and computational power of the Keras framework, enabling the effective implementation of complex DNN architectures and ensuring the reproducibility of our findings (Chollet et al., 2015). Data preprocessing is a crucial step where the raw data is cleaned, normalized, and transformed to be suitable for neural network processing. This is followed by the model building phase, where the network architecture is defined and configured by an iterative search of the space of the hyperparameters that define the network. The iterative training process then employs backpropagation, adjusting the network weights (internal DNN parameters) based on error minimization (Rumelhart et al., 1986b). These steps collectively contribute to the robust and efficient development and refinement of our deep learning models. [Maybe move this paragraph to the end of this subsection, or to section 4.1.](#)

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

- Capturing the intricate relationships between inputs (exospheric densities) and outputs

224 (surface elemental fractions).

- 225 • Reverse engineering the exospheric generation processes, encompassing process yields,
226 chemistry, energy, and distribution characteristics.
- 227 • Predicting the elemental composition of surface tiles.
- 228 • Reconstructing a comprehensive elemental fraction map of Mercury's surface.
- 229 • Formulating an estimator for surface processes contributing to the exosphere.

230 Figure 2 illustrates the Multilayer Perceptron Deep Neural Network basic structure chosen
231 for this task.

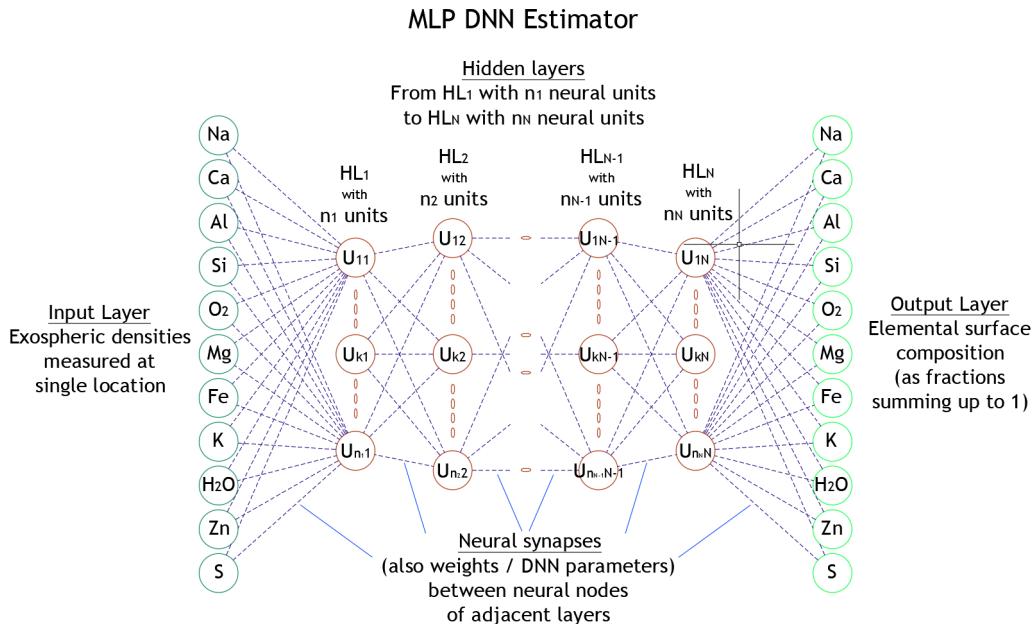


Figure 2. MLP DNN basic structure schematics. Exospheric densities form the input layer on the left. The output layer is formed from the relative surface elemental composition as fractions summing up to 1. There are N number of hidden layers with n_1 to n_N number of neural units (neurons). This structure 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} .

232 *2.2. Deep Neural Network Architecture*

233 The architecture of a multilayer perceptron plays a pivotal role in defining its capability
234 to process and learn from complex datasets, as well as transform input data into meaningful
235 outputs through a series of computational steps. Each layer within this architecture serves a

236 distinct purpose, working in harmony to decipher the intricate patterns hidden within the data.
237 From the input layer, which receives the raw data, to the hidden layers that perform nonlinear
238 transformations, and finally to the output layer that produces the prediction in its proper form,
239 every component of the MLP is crucial.

240 In addition to these layers, the effectiveness of an MLP is also greatly influenced by the
241 loss function, which quantifies the difference between the predicted outputs and the actual
242 values, guiding the network towards accuracy. The regularizer plays a critical role in preventing
243 overfitting, ensuring that the model generalizes well to new, unseen data. Furthermore, the
244 choice of an optimization algorithm is crucial for efficiently adjusting the network's weights
245 and biases to minimize the loss function. Hyperparameter tuning is another essential aspect,
246 involving the adjustment of parameters such as learning rate, number of layers, and number of
247 neurons per layer, to optimize the network's performance.

248 The full architecture of the MLP, including these additional components and their interplay,
249 is detailed in Figure 3. This comprehensive view provides a deeper understanding of how the
250 MLP functions as a whole. The harmonious operation of these components underpins the MLP's
251 proficiency in handling complex regression tasks. Our study leverages this capability to unravel
252 the nuances of Mercury's exosphere. In the following paragraphs, we delve into the specifics of
253 each key element of the MLP architecture, beginning with the input layer, to unravel how they
254 collectively contribute to the network's predictive power.

255 *Input Layer*

256 The input layer, serving as the entry point of the MLP, plays a crucial role in introducing
257 the raw data into the network. In our study, this involves the processing of exospheric density
258 measurements from Mercury. Each neuron in the input layer corresponds to a unique feature of
259 the dataset. For example, distinct elemental density measurements are represented by separate
260 neurons.

261 Prior to entering the MLP, the data is subjected to normalization or standardization. This
262 preprocessing step is critical for ensuring that all features contribute equally to the learning
263 algorithm, thereby preventing biases towards features with larger scales or higher variability
264 (Goodfellow et al., 2016). Specifically, standardization is applied by adjusting each feature to
265 have a zero mean and unit variance. Mathematically, this is represented as:

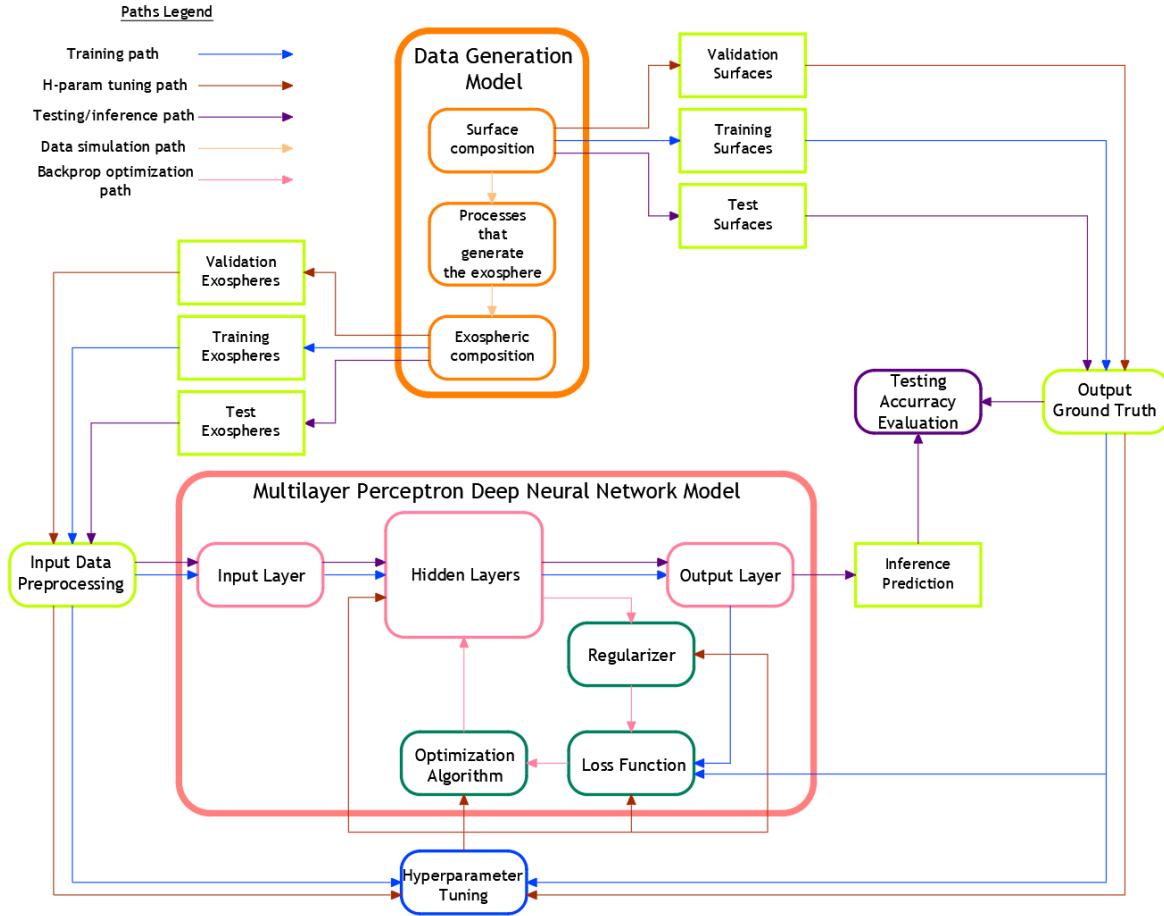


Figure 3. MLP DNN architecture overview. The data generation model produces both the inputs and the outputs for training, validation and testing the algorithm. This data is passed through the MLP DNN model in the training, hyperparameter tuning and testing phases, respectively. The backpropagation optimization 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.

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

where \mathbf{x} is the standardized vector of input features, \mathbf{x}_{orig} is the original vector of input features, $\boldsymbol{\mu}$ is the vector of means of the feature values, and $\boldsymbol{\sigma}$ is the standard deviations vector. Such standardization enhances the efficiency and stability of the network's learning process. Moreover, the scope of the input features can be expanded to incorporate additional relevant information, which can further refine the model's predictions. In the context of our exospheric

271 data analysis, features such as the altitude of measurement and the sun-illumination angle
272 (or its equivalent in solar local time) can be included. These additional parameters allow the
273 MLP to consider not only the density measurements but also the contextual conditions under
274 which these measurements were taken. Incorporating such contextual features can be pivotal
275 in understanding the complex interactions within Mercury's exosphere, thereby enriching the
276 model's predictive accuracy. They are discussed in detail in Section 3.

277 In the MLP, the input layer does not actively compute but rather functions as a conduit,
278 distributing the processed data to the subsequent hidden layers. The design of this layer, which
279 is fully connected to the first hidden layer, ensures comprehensive consideration of every aspect
280 of the input data, setting the stage for the intricate processing that follows in the network.

281 *Hidden Layers*

282 The hidden layers form the core of the MLP architecture, where the actual processing and
283 learning occur (Minsky and Papert, 2017; Hinton, 2007). Positioned between the input and
284 output layers, these layers perform the crucial task of transforming input data into a form that
285 the output layer can use for making predictions. Each hidden layer is composed of a set of neural
286 units - neurons - and each neuron in these layers is fully connected to all neurons in the preceding
287 and succeeding layers, creating a dense network of synaptic connections. The structure of these
288 layers is represented mathematically by combination matrices or weight matrices, which, along
289 with the activation function applied at each neuron, helps form an estimation of the relationships
290 among the processes acting between the input layer and the output layer.

291 In the context of our study on Mercury's exospheric data, the hidden layers enable the MLP to
292 discern and represent the intricate relationships between various exospheric parameters. For our
293 specific application, the architecture is designed to sufficiently capture the nuances of Mercury's
294 exospheric composition and the underlying processes that govern it, involving multiple hidden
295 layers with a substantial number of neurons.

296 A key component of these hidden layers is the activation function. The most commonly used
297 activation function in modern MLPs is the Rectified Linear Unit (ReLU) MLPs (Glorot et al.,
298 2011). ReLU introduces necessary nonlinearity into the network essential for capturing complex
299 patterns while maintaining effective gradient propagation, which is critical for preventing the
300 vanishing gradient problem in deep networks. The ReLU function is defined as $a(\mathbf{z}) = \max(0, \mathbf{z})$,

301 where \mathbf{z} is the input to the activation function.

302 The transformation within each hidden layer can be described by the equation:

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

303 where \mathbf{W}^T represents the weight matrix, \mathbf{x} is the input vector to the hidden layer (input
 304 features or activations from a previous hidden layer), and \mathbf{b} is the bias vector of the affine
 305 transformation. This equation encapsulates the affine transformation followed by the application
 306 of the ReLU activation function, enabling the network to learn and represent complex nonlinear
 307 relationships.

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

313 Output Layer

314 The output layer is the final layer in an MLP and plays a critical role in determining the
 315 format and nature of the final predictions made by the network. In the context of our study,
 316 where the goal is to predict the elemental composition fractions of Mercury's surface, the output
 317 layer is designed to match this specific objective.

318 The number of neurons in the output layer corresponds to the number of outputs the model
 319 is expected to produce. For our multivariate regression task in predicting elemental composition
 320 fractions, there is one neuron in the output layer for each element being predicted. For example,
 321 if we are predicting the fractions of 11 different elements, the output layer would consist of 11
 322 neurons.

323 Each neuron in the output layer represents a specific element, and its output value corresponds
 324 to the predicted fraction of that element. The activation function used in the output layer is
 325 crucial and depends on the nature of the prediction task. For regression tasks like ours, where
 326 the output is a set of continuous values that sum to 1 (representing fractions), the softmax
 327 function is often used (Joachims, 2002). The softmax function converts the raw output of the
 328 network into a probability distribution, ensuring that the predicted fractions are non-negative

329 and sum up to one, aligning perfectly with the physical reality of our task.

330 The formula for the softmax function is as follows:

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

331 where \hat{y}_i is the predicted fraction for the i -th element, and z represents the raw output values
332 from the final hidden layer.

333 The output layer is the culmination of all the preceding layers' transformations and learning.

334 The input data, after being processed through the hidden layers, culminates in a set of predictions
335 that emerge from this layer. In our study, these predictions represent a critical understanding of
336 Mercury's exospheric composition, derived from the intricate patterns learned by the network.

337 Loss Function

338 The loss function plays a pivotal role in guiding the optimization process. It quantifies the
339 difference between the predicted outputs of the network and the actual target values of supervised
340 learning, providing a measure of the model's accuracy. In our study, where the MLP is used for
341 predicting the elemental composition fractions of Mercury's surface, different loss functions can
342 be employed based on their suitability for the task.

343 Kullback–Leibler Divergence: Among the various loss functions experimented with, the
344 Kullback–Leibler (KL) divergence, or KL loss, has shown to be particularly effective in our most
345 successful MLP tests. KL divergence is a measure from the field of information theory, gauging
346 how one probability distribution diverges from a second, reference probability distribution (Cover
347 and Thomas, 2006). For our regression task, it is especially useful because it directly measures
348 how the predicted probability distribution (the output from the MLP) diverges from the actual
349 distribution (the true elemental composition). The formula for KL divergence is:

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

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

352 Mean Absolute Error (MAE) and Mean Squared Error (MSE): In addition to the KL
353 divergence, Mean Absolute Error and Mean Squared Error were also evaluated. MAE is

354 calculated as the average of the absolute differences between the predicted values and actual
 355 values, offering a straightforward interpretation of average error magnitude. MSE, on the
 356 other hand, calculates the average of the squares of the errors (James et al., 2013). MSE is
 357 particularly sensitive to outliers and tends to heavily penalize larger errors, making it suitable
 358 for applications where avoiding large errors is crucial. However, for our task of predicting a
 359 probability distribution, MSE and MAE might not be as effective as KL divergence in capturing
 360 the probabilistic nature of the output.

361 The choice of KL divergence aligns well with the probabilistic nature of our regression task,
 362 ensuring that the model is not just accurate in terms of raw prediction values but also in terms
 363 of the overall distribution of predictions. This is particularly relevant in cases where the relative
 364 proportions of different elements are more critical than their absolute quantities.

365 *Regularizer*

366 Regularization is an essential technique in neural network training, involving the addition of
 367 constraints or penalties to the loss function. This process is critical for reducing the variance of
 368 the model when making inferences on unseen data, thereby enhancing its generalization ability.
 369 In our study, we implement L2 regularization, also known as weight decay, on the weights of
 370 each hidden layer (Bishop, 2006). This specific form of regularization constrains the magnitude
 371 of the weights, preventing them from becoming excessively large. Such a constraint is crucial as
 372 it helps to avoid overfitting the model to the specific dataset used for training.

373 The concept of L2 regularization can be mathematically expressed as:

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

374 where $J(\mathbf{W}, \mathbf{b}, \mathbf{x}, \mathbf{y})$ represents the original loss function, λ is the regularization coefficient,
 375 and θ denotes the vector of all weight parameters (unfolded from the \mathbf{W} matrices). The addition
 376 of this regularization term (penalty) to the loss function ensures that the model not only fits the
 377 training data well but also maintains the flexibility to perform accurately on new, unseen data.

378 The choice of λ , the regularization coefficient, is critical. If λ is too large, it can lead to
 379 underfitting, where the model is overly simplified and fails to capture the underlying trends
 380 in the data. Conversely, a very small λ might not effectively prevent overfitting. Therefore,
 381 selecting an appropriate λ value is often achieved through validation and empirical testing.

382 In a multivariate regression task such as ours, where the model needs to understand complex
 383 relationships between various features in the surface-exosphere interaction at Mercury, L2
 384 regularization helps in maintaining a balance between MLP model complexity and its ability
 385 to generalize. Additionally, L2 regularization can also aid in dealing with multicollinearity in
 386 the data, where independent variables are highly correlated. By penalizing the weights, L2
 387 regularization helps in reducing the impact of these correlations on the model's performance.

388 *Optimization (Learning) Algorithm*

389 In the training of our multilayer perceptron for predicting Mercury's surface composition,
 390 the Adam optimization algorithm plays a critical role (Kingma and Ba, 2014). Adam, short for
 391 Adaptive Moment Estimation, is a refinement of the classic stochastic gradient descent approach,
 392 widely recognized for its effectiveness in handling large-scale data and intricate models.

393 The fundamental mechanism of Adam involves updating the weights of the combination
 394 matrices for each hidden layer to minimize the total error as indicated by the loss function.
 395 This is achieved through backpropagation optimization (Rumelhart et al., 1986a,b), where the
 396 weights are adjusted following their gradients with respect to the loss function:

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

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

402 Adam distinguishes itself through its adaptive learning rate mechanism. Unlike conventional
 403 approaches where a uniform learning rate is applied across all weights, **Adam adjusts the learning**
404 rate individually for each weight. This adaptability is particularly advantageous for datasets
 405 characterized by noisy or sparse gradients.

406 Furthermore, Adam maintains two moving averages for each weight: one for the gradients
 407 and another for the square of the gradients. This dual tracking facilitates a more intelligent
 408 adaptation of the learning rate for each parameter, leading to enhanced efficiency and stability
 409 in training.

410 In our MLP model, Adam's proficiency in managing sparse gradients and dynamically
411 adjusting the learning rate is invaluable, considering the complexity of the task. The optimizer's
412 parameters, including the initial learning rate, beta values for the moment estimates, and epsilon
413 (a small number to avoid division by zero), are generally selected based on empirical evidence
414 and fine-tuned through experimentation.

415 For complex multivariate regression tasks like predicting the elemental composition fractions,
416 Adam proves to be an excellent choice due to its effective handling of large parameter spaces
417 and its robustness to variations in the data's scale or distribution. The optimizer's capability to
418 swiftly converge to a solution and adeptly handle non-stationary objectives, as often encountered
419 in dynamic datasets, significantly enhances the model's performance and reliability.

420 *Hyperparameter Tuning*

421 While the core architecture of the neural network lays the foundation for predicting Mercury's
422 surface composition, the model's overall efficacy and robustness are provided through the
423 meticulous tuning of its hyperparameters. These adjustments are crucial, not only for optimizing
424 the model's performance but also for ensuring its adaptability and reliability across various
425 scenarios. Hyperparameters in our MLP model encompass several key elements:

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

438 To fine-tune these hyperparameters, we employed a Bayesian optimization strategy using the
439 Gaussian Process (GP) approach, as outlined in Bergstra et al. (2011). The tuning process was

440 facilitated by the scikit-optimize library (Head et al., 2018), which utilizes a prior probability
441 distribution function to identify the hyperparameter configuration that minimizes the total loss
442 on a cross-validation dataset.

443 By systematically adjusting parameters like the number of layers and neurons, the learning
444 rate, and the regularization coefficients, we can substantially enhance the model’s learning
445 capabilities and overall performance. Techniques like grid search or randomized search are
446 commonly used in this context, but our approach with Bayesian optimization offers a more
447 efficient and effective means of navigating the hyperparameter space. This process ensures that
448 our model achieves an optimal balance between learning efficiency and complexity, tailoring it
449 precisely to the unique characteristics of the analysis of Mercury’s surface-exosphere interaction.

450 **3. Exospheric Data**

451 With the BepiColombo mission’s nominal operations set to commence in 2026, the anticipated
452 data from the SERENA instrument remains unavailable for our current analysis. Consequently,
453 to apply our methods for reverse engineering a model of the exosphere in order to reconstruct
454 Mercury’s surface composition, we have undertaken the task of generating our own datasets
455 through simulations based on this model.

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

465 *3.1. Data Generation Model*

466 A variety of environmental models have been used historically to simulate the exosphere,
467 especially data obtained during the measurements of orbiting spacecraft (Sarantos et al., 2009;
468 Cassidy et al., 2015; Plainaki et al., 2017; Mura et al., 2007; Mangano et al., 2015; Wurz and
469 Lammer, 2003). Our approach inversely reconstructs the processes that create the exosphere,

470 using a specific model to train our deep neural networks for predicting surface compositions.

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

482 Our method is designed to be adaptable, not limited to a single model but capable
483 of reconstructing various exosphere generation models and predicting surface compositions
484 accordingly. The models successfully trained on our generated datasets can be termed 'base
485 nodes.' These nodes define the constraints of the representation of the real exospheric generation
486 processes within the physical parameter space.

487 Our study primarily revolves around one such 'base node' model and its parameter space.
488 In this context, the data generation model, utilizing an exospheric simulation by Mura et al.
489 (2007), assumes an important role in simulating conditions within Mercury's exosphere. While
490 the model itself is not the central focus of our investigation, it is crucial for generating the
491 datasets that underpin our analytical algorithms. This model comprehensively describes the
492 physical processes shaping the exosphere, allowing us to define specific regions in the parameter
493 space. These regions govern the data distribution our multilayer perceptron deep neural network
494 aims to estimate.

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

501 physical processes.

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

506 *3.1.1. Surface and Regolith*

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

513 Firstly, the surface in our model is represented as a grid comprised of 36×18 surface tiles in a
514 modified Mercator projection. Each tile measures $10^\circ \times 10^\circ$, which, at the equator, translates to
515 approximately $425 \text{ km} \times 425 \text{ km}$. This averaging of composition inevitably reduces the complexity
516 of the parameter space, as it results in the loss of finer details in the spatial relations of the
517 spread of the different species - elemental and mineral - on the planetary surface. In addition
518 to the surface grid sizes in longitude and latitude, other relationships are captured implicitly in
519 the represented subset of parameter space.

520 The next subset of parameters for our simulation of Mercury's surface and regolith is derived
521 from its mineralogical composition. We select a specific set of minerals, as detailed in Table 1,
522 believed to be present on Mercury's surface (Wurz et al., 2010). These minerals are assumed
523 to exist in varying proportions, contributing to the regolith's overall mineral composition. This
524 selection process inherently defines another parameter subspace within our model, capturing
525 some of the complex coexistence relationships among different minerals. These relationships
526 delineate zones characterized by dominant primary minerals and their secondary counterparts,
527 enforcing the presence of some of the minerals on the surface. Additionally, constraints on the
528 minimal fractions of specific minerals and the presence of water ice further refine this parameter
529 subspace, offering an implicit subset of complexity to our model. From the included minerals,
530 hedenbergite (a mineral deposited primarily from meteorites), sphalerite (a mineral resulting

from volcanic activity), and water ice are considered as rare minerals, and their presence is reduced by 80% in the random surface generation, compared to the other six minerals. Furthermore, in the randomized creation of the surface, they are not allowed to be distributed everywhere on the surface. The overall mean fractions of the minerals in our datasets are reported in Table 1, while a more detailed description for the different types of datasets is shown in the Appendix.

Mineralogy and Relationships (Baseline Model)				
Mineral Name	Chemical Formula	Decomposed to	Rarity	Mean Fraction
Anorthite	$\text{CaAl}_2\text{Si}_2\text{O}_8$	$\text{Ca}, 2\text{Al}, 2\text{Si}, 4\text{O}_2$	Can be primary	?
Albite	$\text{NaAlSi}_3\text{O}_8$	$\text{Na}, \text{Al}, 3\text{Si}, 4\text{O}_2$	Can be primary	?
Orthoclase	KAlSi_3O_8	$\text{K}, \text{Al}, 3\text{Si}, 4\text{O}_2$	Can be primary	?
Enstatite	$\text{Mg}_2\text{Si}_2\text{O}_6$	$2\text{Mg}, 2\text{Si}, 3\text{O}_2$	Can be primary	?
Diopside	$\text{MgCaSi}_2\text{O}_6$	$\text{Mg}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	Can be primary	?
Ferrosilite	$\text{Fe}_2\text{Si}_2\text{O}_6$	$2\text{Fe}, 2\text{Si}, 3\text{O}_2$	Can be primary	?
Hedenbergite	$\text{FeCaSi}_2\text{O}_6$	$\text{Fe}, \text{Ca}, 2\text{Si}, 3\text{O}_2$	Rare	?
Sphalerite	ZnS	Zn, S	Rare	?
Water Ice	H_2O	H_2O	Rare	?

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.

Once the mineralogical composition of each surface tile within the datasets is established, the next step involves determining the decomposition process of these minerals into their constituent elemental species (atoms or molecules). In our approach, we adopt the classical additive method for breaking down the regolith minerals into their elemental components (Wurz et al., 2010). This implies an assumption that the surface, on average, encompasses a complete pool of atoms and molecules derived from these minerals, which are then subjected to external environmental forces. This approach, as an approximation, considers the full fraction of volatile species (such as Na, K, H₂O, S, and O₂) as being readily available for thermal and photon-stimulated desorption 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

549 our algorithms.

Elements and Relationships (Baseline Model)				
Element Name	Designation	From Mineral	Mean Elemental Fraction	Elemental Relations
Aluminium	Al	Anorthite, Albite, Orthoclase	?	?
Calcium	Ca	Anorthite, Diopside, Hedenbergite	?	?
Iron	Fe	Ferrosilite, Hedenbergite	?	?
Sodium	Na	Albite	?	?
Oxygen	O ₂	All, except Sphalerite, Water Ice	?	?
Sulfur	S	Sphalerite	?	?
Water Vapor	H ₂ O	Water Ice	?	?
Zinc	Zn	Sphalerite	?	?
Silicium	Si	All, except Sphalerite, Water Ice	?	?
Potassium	K	Orthoclase	?	?
Magnesium	Mg	Enstatite, Diopside	?	?

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).

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

558 *3.1.2. Environmental Conditions*

559 [References are missing...](#)

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

565 Mercury’s proximity to the Sun significantly influences its interaction with the surrounding
 566 environment. Its notably eccentric orbit causes substantial variation in distance from the

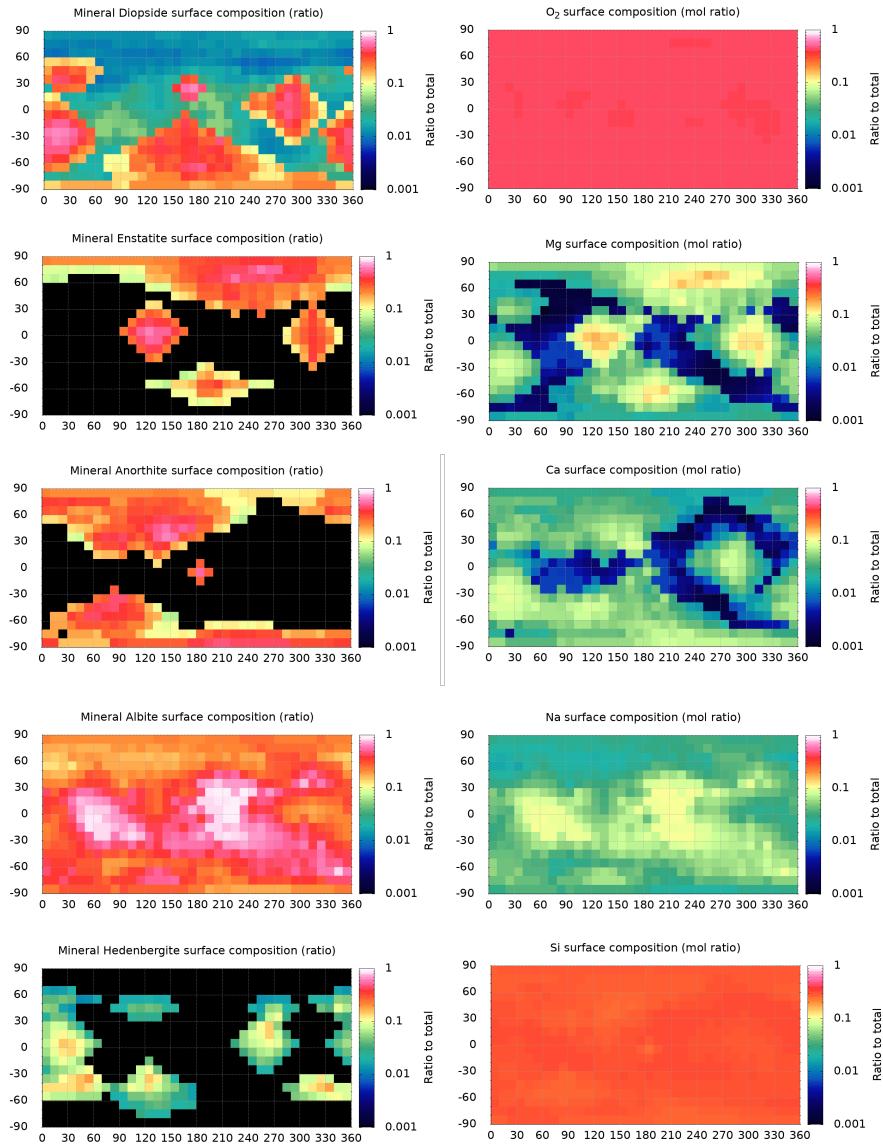


Figure 4. From mineralogy to elemental composition. The figure depicts the resulting elemental maps for Calcium, Magnesium and Sodium from Albite, Anorthite, Enstatite, Diopside, and Hedenbergite. Shown are also the resulting Silicium and Oxygen maps for abundance comparison. **Improve figure by removing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

567 Sun between perihelion (closest approach) and aphelion (furthest distance). In our model, we
 568 specifically focus on conditions at perihelion, where Mercury is about 0.31 astronomical units
 569 (AU) or 46 million kilometers from the Sun. This close distance markedly impacts the intensity
 570 of the solar influence, thereby affecting the ranges of the effects and processes that contribute to
 571 generating the exosphere. One such influence is the equivalent photon flux, which is the photon
 572 flux at Earth's orbit adjusted to account for Mercury's closer position to the Sun by a factor of

573 $1/r^2$, where r is the distance to the Sun in AU. A photon flux at Earth of $3.0 \times 10^{15} \text{ cm}^{-2}\text{s}^{-1}$
574 is considered.

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

579 Furthermore, the dust environment around Mercury is considered for particles smaller than
580 100 μm 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
581 Mercury's vicinity in agreement with the modal impact velocity reported by Cintala (1992),
582 compared to a planet velocity at perihelion of 59 km/s. However, our model does not differentiate
583 between the origins of these dust particles—whether they come from the Main Belt Asteroids,
584 Jupiter Family Comets, Oort Cloud Comets, or Halley Type Comets—nor does it consider
585 the full ranges and exact distributions of particle sizes and velocities. Additionally, no large
586 meteorite impacts or increases of fluxes due to particularly dense cometary streams, such as
587 from comet Encke (cite someone), are considered. Grain size distribution influence of the dust
588 particles is also not represented in our physical parameter space.

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

597 In extending our discussion on environmental conditions affecting Mercury, it's important to
598 clarify the limitations of our model, especially concerning magnetic field effects. Our study
599 does not encompass the full spectrum of influences that the interplanetary magnetic field
600 and Mercury's own magnetic field might have on the planet's surface and exosphere. This
601 omission includes a variety of magnetic field characteristics such as Mercury's dipole moment, the
602 thickness and strength of the Harris sheet, dipole shift, and other pertinent electric and magnetic
603 field parameters. Nor does it engage with complex magnetic field models that could offer a more

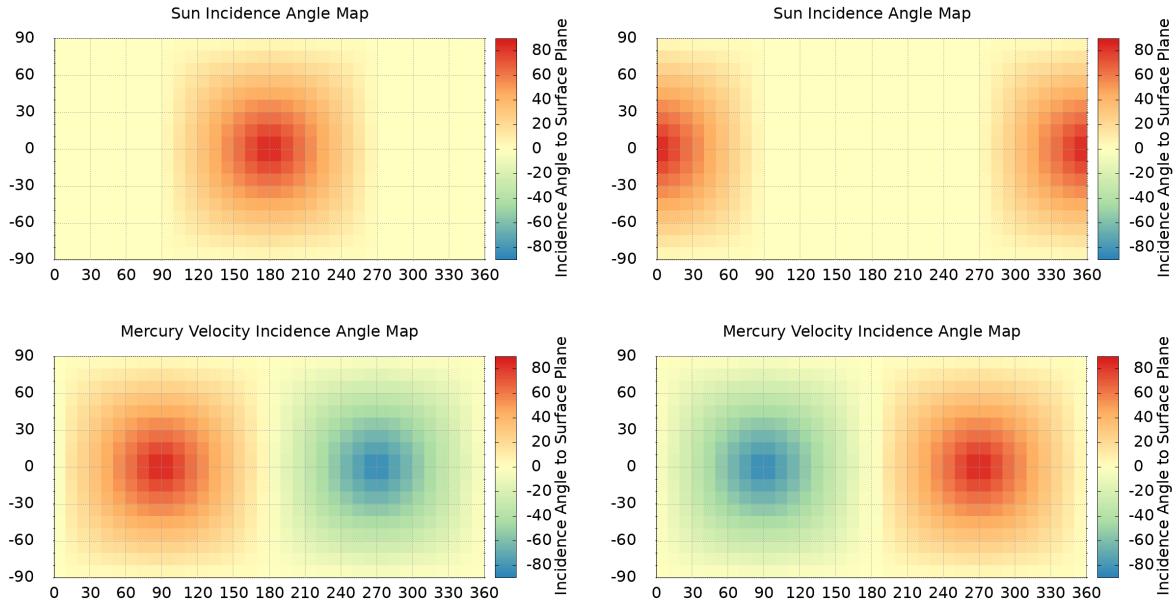


Figure 5. Maps of incidence angles due to planet orientation at two consecutive perihelia. [Add more explanation...](#) Possibly improve/update this figure.

604 detailed understanding of these interactions. This simplification is considered satisfactory as the
 605 algorithm will not be taught to learn a shift of the magnetic dipole dependence and dynamical
 606 effects at this stage.

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

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

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

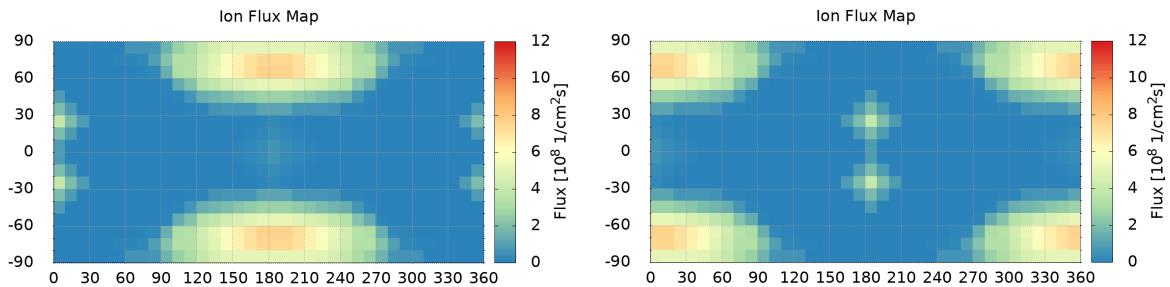


Figure 6. Maps of ion fluxes at the surface at two consecutive perihelia. Add more explanation... Possibly improve/update this figure.

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

Environmental Conditions					
Group	Condition	Units	Value	Sources Affected	Processes Affected
Star Activity	Solar wind velocity	km/s	450	Proton flux	SP
	Solar wind density	1/cm ³	60	Proton flux	SP
	Photon flux at Earth	1/cm ² s	3.0×10^{15}	Equivalent photon flux	PSD
Comets and Asteroids	Dust particle size	μm	0-100	Micrometeorite flux	MIV
	Mean flux of dust particles	g/cm ² s	1.0×10^{-16}	Micrometeorite flux	MIV
	Mean velocity of dust particles	km/s	20	Micrometeorite flux	MIV
Magnetic fields	Cusp footprints size	Map		Ion precipitation zone	SP
	Cusp footprints location	Map		Ion precipitation zone	SP
	Cusp footprints ion flux distribution	Map		Ion flux	SP
	Cusp footprints area coefficient	-	0.4	Ion flux	SP
Planet	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.

3.1.3. Sources and Processes for Generation of the Exosphere

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

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

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

Micrometeorite fluxes and impact vaporization

In our model, micrometeorites are represented as an incoming flux of interplanetary dust that expels matter from the surface regolith in relation to the incoming flux and velocity, as described by Cintala (1992). The distribution of this flux onto Mercury's surface is influenced by the velocity of Mercury and its projection onto the surface area where the flux is calculated, as illustrated in Figure 5. We employ a simple relationship between the angle of incidence and the modification of the mean flux onto the surface, defined as:

$$\Delta\phi_{MIV} = \frac{V_{mm} \cos \beta_{Surf}}{V_M}, \quad (8)$$

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

This approach yields outflows of surface matter ranging from 3.5×10^{-16} to 7×10^{-16} $\text{gcm}^{-2}\text{s}^{-1}$. These values are conservatively estimated to be about a factor of 2 smaller than those suggested by Cintala (1992) and two orders of magnitude smaller than those proposed by Pokorný et al. (2017), fitting within the parameter space interest to modelers without overly emphasizing this omnipresent process. Additionally, this assumption poses a more challenging

scenario for the algorithm due to the inherent representability of the surface composition by the exosphere generated due to the MIV. The vaporized species include larger molecules such as CaO, NaOH, NaO, and others, resulting from the complex chemistry within the impact-produced cloud (Killen, 2016; Bereznay, 2018). However, for our initial DNN analysis iteration, we assume these species have very short photolysis lifetimes, quickly breaking down into their constituent elements without further energization.

The model adopts Maxwellian distributions for the velocities of the excited particles, with 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	1.0×10^{-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.

667 *Ion precipitation and Ion sputtering (SP)*

668 The ion sputtering is initiated by a flux of bombarding ions, predominantly comprising solar
669 wind protons, which efficiently ejects atoms/molecules from the surface regolith (Wurz et al.,
670 2010; Killen et al., 2007). The ion flux's impact is localized to areas where the open magnetic
671 field lines intersect the surface.

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

$$\phi = C \rho_{sw} v_{sw}, \quad (9)$$

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

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

$$\frac{d\Phi_n}{dE_e} = Y c \int_{E_{\min}}^{E_{\max}} \frac{d\Phi_I}{dE_i} f_S(E_e, E_i) dE_i, \quad (10)$$

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

$$f_S(E_e, T_m) = c_n \frac{E_e}{(E_e + E_b)^3} \times \left[1 - \left(\frac{E_e + E_b}{T_m} \right)^{1/2} \right], \quad (11)$$

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

$$T_m = E_i \frac{4m_1 m_2}{(m_1 + m_2)^2}, \quad (12)$$

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

For this investigation, we assume a uniform yield efficiency of 0.1 for all species, a figure that is relatively high compared to existing literature Schaible et al. (2017). However, we apply a conservative reduction in yield to account for the regolith's porosity (a uniform factor of 0.35) and the microshadows within it (a uniform factor of 0.40). This adjustment modestly diminishes the sputtering effect, a deliberate choice to complicate the prediction of surface composition by DNN algorithms in high-latitude regions receiving solar wind precipitation. The angular distribution 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 ₂	H ₂ O	
Mean ion flux	Source	1/cm ² s						1.08 x 10 ⁹						
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.

Temperature map and Thermal desorption (TD)

Thermal desorption becomes notably efficient at temperatures above 400 K, a threshold where the vibration of loosely bound atoms and molecules increases significantly (Mura et al., 2007). In our model, TD primarily affects a specific population of molecules—namely Na, K, H₂O, S—which are loosely bound to Mercury's surface.

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

$$T_s(\phi, \theta) = T_{\min} + (T_{\max} - T_{\min})(\cos\phi\cos\theta)^{1/4}, \quad (13)$$

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

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

$$\Phi_n = \nu N c e^{-\frac{U_d}{k_B T}}, \quad (14)$$

where ν denotes the vibrational frequency of the species, N the surface density of the regolith, c the fractional presence of the species within the regolith, U_d the species' binding energy, k_B the Boltzmann constant, and T the temperature at which desorption occurs. For the efflux of particles from the surface, a Maxwellian-Boltzmann distribution is assumed, reflecting the statistical nature of the thermal motion contributing to the desorption process.

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

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

711 Photon flux and Photon-stimulated desorption (PSD)

712 Photon-stimulated desorption is initiated by the interaction of incoming photons with the
 713 surface, each photon possessing the capability to eject atoms or molecules from a population
 714 of loosely bound volatile species. The efficiency of this process is contingent upon the cross-

section for photon impact (Wurz and Lammer, 2003; Killen et al., 2001; Wurz et al., 2010). At perihelion, the incident photon flux is quantified as $3.1 \times 10^{16} \text{ cm}^{-2}\text{s}^{-1}$.

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

$$\Phi_n = N_c \int \Phi_\gamma(E) \sigma(E) dE, \quad (15)$$

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

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

$$\Phi_n(\phi, \lambda)^* = \Phi_n \cos(\phi) \cos(\lambda), \quad (16)$$

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

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

$$f(E) = \beta(1 + \beta) \frac{EU^\beta}{(E + U)^{2+\beta}}, \quad (17)$$

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

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

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

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

731 *3.1.4. Dynamics of the Exosphere*

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

737 *Photo-ionization*

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

743 In our model, once ionization occurs, the recombination of the produced ion with an electron
 744 is not considered. This assumption is based on the low density of particles in the exosphere,
 745 which makes such recombination events exceedingly rare. Furthermore, the trajectory of these
 746 ions around the planet is not tracked post-ionization. As a result, this ionization process is
 747 depicted as a net loss to the exosphere.

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

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

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

751 *Solar Radiation Pressure*

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

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

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

763 *Other Simulation Parameters*

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

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

771 • **The number of particles** included in the simulation significantly affects the accuracy
772 and complexity of the model. A higher number of particles allows for a more detailed and
773 nuanced simulation of the exosphere, capturing better the statistical behavior of particle
774 populations. However, the need for more computational resources and time increases with
775 the number of particles. Balancing detailed simulations with computational efficiency is
776 essential. For our purposes, considering the coarse resolution and the high number of
777 simulations required, we have chosen a particle count of 50,000 for our simulation runs.

778 • **The exospheric grid** is a virtual representation of space around the planet, divided into
779 discrete cells, serving as the framework for tracking particle positions and movements.
780 This grid enables the simulation to map particle density, velocity, and direction across the
781 exosphere's different regions. The grid's resolution—its fineness or coarseness—directly
782 affects the simulation's precision. A finer grid captures more detailed spatial variations in
783 particle behavior but demands greater computational resources. For our study, we have
784 extended the surface 2D grid in 100 km altitude steps around the planet to a final altitude
785 of 5000 km.

786 • **The Coriolis force**, resulting from the planet's rotation, affects the motion of particles in
787 the exosphere. It causes the path of moving particles to curve relative to the planet's surface,
788 influencing their trajectories based on motion direction and latitude. While the Coriolis
789 force doesn't change particle speed, it redirects their paths, creating complex circulation
790 patterns within the exosphere. In our study, we have omitted the Coriolis force effect,
791 thereby simplifying the parameter space.

792 • **Magnetic field parameters** are used for modeling the behavior of charged particles within
793 the exosphere. The planet's magnetic field interacts with these particles, affecting their
794 trajectories through Lorentz force. However, our model does not account for recombination
795 between ions and electrons or track ions' paths around the planet, thus omitting magnetic
796 field effects from our simulations.

797 *Combining the Singular Exospheres*

798 To generate Mercury's exosphere, we ultimately employ a simplified additive approach. This
799 method entails conducting separate simulations for each of the four primary surface release
800 processes, reproduced for each distinct species. Following these individual simulations, we
801 aggregate the outcomes to compose the overall exosphere. This process essentially involves
802 adding together the resulting singular exospheres generated for each species, without considering
803 interactions between the various processes. For instance, we do not account for potential
804 competition among processes for a finite pool of particles at the surface. Similarly, the exospheres
805 for different species are treated as non-interacting entities. An example for the resulting Sodium
806 exosphere is shown on Figure 7.

807 Add some more sentences to complete this section...

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

809 *3.2. Generation of the Datasets*

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

814 *3.2.1. Training, Validation and Test Datasets*

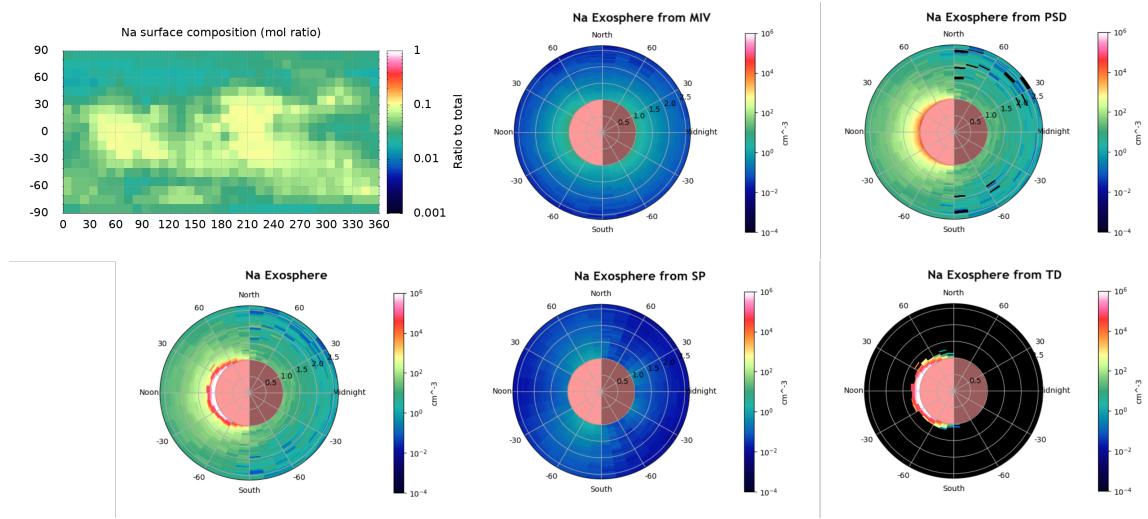


Figure 7. Resulting exosphere from the example Sodium surface composition. The figure depicts the generated singular Sodium exospheres from the MIV, SP, TD and PSD processes, as well as their aggregate. **Improve figure by removing unnecessary duplication of coordinates, bars, etc.**

815 We produce three distinct types of datasets: training, hold-out validation, and testing
 816 datasets. This subdivision is a fundamental practice in machine learning, ensuring that the
 817 algorithm is trained on a diverse set of data, validated for accuracy, and finally tested for
 818 generalization to unseen data.

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

828 During our training we perform a hold-out validation using a single validation dataset to
 829 evaluate the trained network throughout each training epoch and identify the positions in the
 830 learning curve where the algorithm starts to train to represent the training set data distribution
 831 too well and its generalization ability to predict is reduced. Additionally, the validation set is
 832 used during the hyperparameter tuning, where again the algorithm is evaluated on its ability to

833 generalize.

834 On the other hand, the test datasets consist of examples (data points) that are never seen
835 during training, and the MLP DNN algorithm's learning phase has not been influenced by them.
836 This is important to make sure that the algorithm's performance evaluation is not influenced
837 by improvement in its accuracy due to evaluation on previously seen examples.

838 *3.2.2. Datasets Generation Procedure*

839 The creation of our datasets adheres to the systematic procedure described below: (1) The
840 surface elemental composition is randomly generated based on varied mineralogies, establishing
841 the 'ground truth' for each dataset. (2) We simulate environmental effects to craft a static
842 representation of the exosphere at a specific moment, taking into account the necessary physical
843 and chemical processes. (3) Specific points within the exosphere are defined. (4) Exospheric
844 measurements are taken at these points, and these measurements are compiled into data subsets,
845 or observations, each representing a comprehensive snapshot of exospheric conditions from a
846 particular simulated exosphere. (5) Relevant features are selected or engineered from these
847 observations to serve as inputs for the algorithms, ensuring that each data point is informatively
848 rich. (6) For the training datasets, multiple observations/subsets representing different
849 exospheres are aggregated, enhancing the dataset's complexity and variability. Conversely, the
850 hold-out validation and testing datasets are each derived from individual observations/subsets
851 to maintain their distinctiveness and integrity for unbiased evaluation.

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

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

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

862 *3.2.3. Positioning in the Exosphere*

We strategically define various positions within the exosphere to place virtual sensors, aiming to provide measurements the exosphere's constituents in a comprehensive way. The selection of these sensor locations is guided by a prioritization of the collection of diverse and relevant data. This involves considering several factors such as altitude, latitude, and varying environmental conditions to not only gather comprehensive data but also to facilitate the reconstruction of the complete surface maps from the Deep Neural Network (DNN) predictions, which are based on these measurements.

To achieve a balance between capturing spatial variability and enabling the reconstruction of accurate surface maps, we position points in the exosphere directly above each surface grid tile's center. This placement strategy results in a total of 648 measurement data points for each data subset, ensuring thorough coverage and data collection across the exosphere.

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

To enhance the comprehensiveness of our training datasets, we employ data augmentation techniques to some of them, introducing measurements from a range of altitudes beyond the baseline. This **strategy** not only increases the volume of our dataset but also infuses it with a richer variety of data examples. By incorporating measurements taken both above and below the initially chosen altitude, we aim to encapsulate a broad range of exospheric conditions within our dataset. This augmentation allows our algorithm to cultivate a more detailed understanding of how exospheric measurements correlate with surface elemental compositions across different altitudes. Consequently, this enriched dataset facilitates a dual advantage for the learning process: it significantly expands the training dataset, offering a broader base for the algorithm to learn from, and it introduces a variety of measurement conditions that, despite their differences, link back to consistent ground truths. Through this method, we enable the algorithm to discern and learn from the nuanced variations in exospheric behaviors attributable to changes in altitude, thus sharpening its predictive accuracy and robustness. In our training we used one baseline **training dataset type with examples only at altitudes of 500 km**, and one augmented training

894 dataset type with measurements formed at altitudes of 200, 500, 800, 1100, 1400, 1700, and
895 2000 km.

896 For the validation and testing phases of our DNN algorithms, however, we adopt a different
897 strategy. The hold-out validation dataset is consistently set at the MPO’s periherm altitude
898 of 500 km to maintain consistency in evaluation conditions. On the other hand, to ensure
899 a thorough assessment of the algorithm’s performance, we generate multiple test datasets at
900 different altitudes, ranging from 200 km to 1500 km. This varied altitude approach for the
901 test datasets allows for a comprehensive analysis and evaluation across a broader spectrum
902 of exospheric conditions, enabling us to more accurately gauge the algorithm’s efficacy and
903 robustness in predicting the surface elemental composition.

904 *3.2.4. Measurement and Observations Creation*

905 In this section, we outline the methodology for generating measurements and observations
906 within our simulated environment, focusing on the detection and quantification of particle
907 densities across various elemental species in the exosphere. Our approach employs a
908 hypothetical, idealized virtual sensor designed to simulate “pristine” measurements. This
909 conceptual sensor operates without considering the complexities and potential inaccuracies
910 introduced by real-world sensor characteristics such as detector noise and sensitivity limitations.

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

918 *3.2.5. Feature Selection and Engineering*

919 The effectiveness of deep neural network algorithms in modeling complex relationships within
920 data is significantly influenced by the selection and engineering of input features. These features,
921 which describe each data point within all three types of datasets, are pivotal in guiding the
922 algorithm towards an accurate and meaningful representation of the underlying data structure.
923 Carefully chosen or engineered features can enhance the algorithm’s ability to discern patterns

924 and relationships, thereby improving its overall performance.

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

- 929 • **Altitude of Measurement:** This feature is critical for capturing altitude-specific
930 dynamics, enabling the algorithm to identify how the distribution of neutral species changes
931 with altitude relative to their source points on the surface.
- 932 • **Logarithm of Exospheric Density:** By applying the base 10 logarithm to the exospheric
933 density of each species, we introduce a constraint that aids the algorithm in exploring
934 nonlinear relationships, acknowledging the exponential decrease in density with altitude.
- 935 • **Local Solar Time Dependency:** Represented as $\cos(\phi - 180)$ ([check value](#)), this feature
936 helps differentiate between exospheric populations on the dayside, nightside, and the
937 transitional terminator regions, enhancing the model's spatial awareness.
- 938 • **Latitude Dependency:** Using $\sin(\gamma)$ ([check value](#)), this feature allows the algorithm to
939 account for latitude-specific phenomena, such as ion sputtering, which varies across different
940 latitudinal zones.
- 941 • **Logarithm of Altitude:** This engineered feature is intended to highlight non-linear
942 altitude effects on the measured parameters, providing another layer of depth to the altitude-
943 related analysis.
- 944 • **Sun Incidence Angle:** A linear feature that guides the algorithm in understanding the
945 impact of solar radiation on exospheric generation, based on its incidence angle relative to
946 the surface.
- 947 • **Proton Flux Virtual Data:** Integrating virtual measurements akin to those from the
948 MIPA ion detector of the SERENA suite, this feature hints at sputtering effects induced
949 by proton precipitation on the surface, offering a proxy for understanding underlying ion
950 induced physical processes.

951 Each data point in our datasets is defined by combinations of these features, forming distinct
952 feature sets that illuminate to the algorithm various aspects of the exosphere's behavior. The
953 compilation of these feature sets is crucial for unraveling the capabilities of neural networks in

predicting surface compositions and contributing insights into the mechanisms governing particle release into the exosphere. The assortment and implications of these feature sets are detailed in Table 9.

Feature Set Name	Feature Sets Explored						
	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 9. Feature sets used in training and testing of the MLP DNN algorithm.

3.2.6. Datasets Consolidation

Note that there may be duplications of information with the section on positioning in the exosphere - should be fixed.

In this section, we delve into the methodology behind the consolidation of data subsets to assemble the final datasets used for training, validation, and testing within our study. The process of dataset consolidation is critical for ensuring that our DNN algorithms are exposed to a broad and representative spectrum of exospheric conditions, facilitating robust training and accurate evaluation.

For validation and test dataset purposes, employing single data subsets allows for straightforward prediction and reconstruction of surface maps corresponding to individual surface-exosphere simulations. For the hold-out validation dataset, we adopt a more focused approach by selecting a single baseline subset based on measurements at a 500 km altitude. This subset is chosen for its relevance to the anticipated operational conditions of the BepiColombo mission and serves as a standard benchmark for preliminary evaluation of the algorithm's performance.

The test datasets, on the other hand, are designed to thoroughly assess the algorithm's predictive capabilities under varied conditions. We generate single baseline subsets from 15 distinct surface-exosphere simulations, providing a diversified testing ground. Additionally, to

4 RESULTS

explore the algorithms' responsiveness to altitude variations, we create multiple test sets from the same 15 surface-exosphere pairs, but with measurements taken at constant altitudes for each altitude-varied subset ranging from 200 km to 1,500 km.

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 10. 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.

However, the training phase demands a more complex strategy in order to approximate the relationships between the employed physical processes accurately. This in turn requires to produce a training dataset that captures a larger part of the data distribution. To address this, we aggregate multiple data subsets or observations to form the training dataset. This aggregation is crucial for introducing the necessary diversity and comprehensiveness into the training data, ensuring that the algorithm encounters a wide range of exospheric phenomena and learns to generalize across different scenarios. The training datasets are constructed from a variety of subsets, with the number of included subsets ranging from 10 baseline observations (yielding a total of 6,480 data points) to 300 augmented observations (resulting in a staggering 1,360,800 data points). It should be noted that utilizing each training dataset, results in an algorithm trained to represent a different empirical data distribution derived from this particular dataset.

4. Results

In this section, we detail the findings of our investigation, which are divided into two distinct phases to provide a comprehensive understanding of our study's outcomes. The first phase

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 11. 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.

992 focuses on the configuration and optimization of the deep neural network, specifically through the
 993 process of hyperparameter tuning. This entails a systematic exploration of the hyperparameter
 994 space during the training phase to identify the optimal settings that enhance the DNN's ability to
 995 model the data accurately. The second phase evaluates the performance of the DNN, now finely
 996 tuned with the optimal hyperparameter configuration, in interpreting and making predictions
 997 on data derived from various process parameter spaces during the testing phase.

998 Our approach anticipates the dual avenues of future research in this domain. Firstly, it lays
 999 the groundwork for further refining the DNN's architecture and tuning process, aiming at an
 1000 even more precise representation and understanding of the selected physical processes parameter
 1001 space. Secondly, it sets the stage for applying the optimized DNN structure to different regions
 1002 within the processes parameter space, testing its robustness and adaptability in estimating
 1003 data distributions from novel and varied conditions. This bifurcated strategy underscores our
 1004 commitment to advancing the field by not only enhancing algorithmic performance in current
 1005 scenarios but also ensuring their applicability and efficacy in diverse scenarios.

1006 4.1. Performance Metrics

1007 Evaluating the performance of machine learning models, particularly in complex tasks such
 1008 as predicting the composition of the exosphere with a multilayer perceptron, necessitates the
 1009 use of precise and insightful metrics. These metrics not only offer a quantitative assessment of

1010 the model’s predictive accuracy but also shed light on areas for potential improvement, thereby
 1011 contributing to the refinement of the model’s predictive capabilities.

1012 In our study, we employ a dual approach to performance evaluation, utilizing both custom
 1013 and standard metrics to gain a comprehensive understanding of our deep neural network’s
 1014 effectiveness. The centerpiece of our evaluation strategy is the Euclidean similarity 4 (ES4)
 1015 metric, which synthesizes elements of Euclidean distance and cosine similarity to measure the
 1016 accuracy of predictions. This metric is particularly valuable as it provides a nuanced perspective
 1017 on prediction accuracy by considering both the magnitude and direction of the predicted and
 1018 actual values in a multidimensional space. It shows a better resemblance to a percentile
 1019 accuracy when measured on the analyzed data distribution than other traditional metrics used
 1020 in regression tasks, such as the R^2 . The ES4 metric is mathematically represented as:

$$\text{ES4} = \left(1 - \frac{\sqrt{\sum_i (\hat{\mathbf{y}}_i - \mathbf{y}_i)^2}}{\sqrt{\sum_i \mathbf{y}_i^2}} \right) \times \left(\frac{\hat{\mathbf{y}}_i \cdot \mathbf{y}_i}{\|\hat{\mathbf{y}}_i\| \|\mathbf{y}_i\|} \right), \quad (18)$$

1021 where $\hat{\mathbf{y}}_i$ denotes the predicted surface composition and \mathbf{y}_i represents the actual surface
 1022 composition.

1023 Moreover, we incorporate the R-squared (R^2) metric into our evaluation framework. The
 1024 R^2 metric, commonly used in regression analysis, quantifies the proportion of the variance in
 1025 the dependent variable that is predictable from the independent variable(s). In the context of
 1026 multivariate regression, R^2 is defined as:

$$R^2 = 1 - \frac{\sum_i (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_i (\mathbf{y}_i - \bar{\mathbf{y}})^2}, \quad (19)$$

1027 where $\bar{\mathbf{y}}$ is the mean of the actual values. This metric is particularly useful for assessing
 1028 the model’s ability to capture the variance in the data, offering insights into how well the
 1029 model’s predictions approximate the actual data distribution compared to a naive model that
 1030 only predicts the mean.

1031 Additionally, we evaluate the model using absolute and relative residuals, which provide
 1032 further granularity in understanding the model’s performance. These residuals help identify the
 1033 absolute and relative differences between predicted and actual values, offering a direct measure
 1034 of prediction error.

1035 By combining these metrics, we achieve a multidimensional evaluation of our DNN’s

1036 performance, encompassing both the accuracy of individual predictions and the model's overall
1037 ability to capture the complexity of the data. This comprehensive assessment ensures that
1038 we not only validate the model's outputs but also deepen our understanding of its predictive
1039 behavior across a variety of scenarios.

1040 *4.2. Training Phase*

1041 An extensive training campaign was undertaken to explore both the empirical distribution
1042 represented in the training datasets, and the hyperparameter space of the neural network
1043 architecture. This effort aimed to develop an accurate estimator that demonstrates optimal
1044 generalization capabilities by closely approaching the true data generating distribution.
1045 Ultimately, we sought to refine and finalize the components of the multilayer perceptron deep
1046 neural network. This endeavor is significant, as mapping the trajectory through hyperparameter
1047 space holds considerable promise for future research in this domain.

1048 Our journey commenced with a baseline training dataset, which, being the simplest, was least
1049 representative of the true data generation distribution. This dataset comprised 10 subsets with
1050 a total of 6,480 data points from the F00 feature set (only exospheric density measurements).
1051 The initial architecture of the MLP DNN was modeled closely after the structure proposed in
1052 the study by Kazakov et al. (2020), featuring a four-layered network with 400, 200, 200, and 100
1053 neurons, respectively. Notably, our study expanded the input layer to accommodate a greater
1054 number of elements - 11 total input elements.

1055 The output layer of the network employs softmax units, designed to predict the surface
1056 composition of the same 11 elements provided as inputs. Initially, the minibatch size was set to
1057 1,024 examples. The regularization L-2 coefficient and the learning rate were chosen as 1.0×10^{-6}
1058 and 0.5×10^{-4} , respectively, to balance the trade-off between learning efficiency and the risk of
1059 overfitting.

1060 *Eliminating Skewed Predictions*

1061 The initial analysis of predicting fractionated surface elemental composition revealed that
1062 the accuracy metrics were significantly skewed by the prevalence of certain abundant elements,
1063 notably oxygen (O_2) and silicon (Si), which are omnipresent in most of the minerals in our model.
1064 This skewness, stemming from the algorithm's propensity to more accurately predict these two
1065 elements, was addressed by excluding them from the prediction vector and adjusting it to ensure

1066 a normalized sum of 1. Consequently, the refined model focuses on predicting the normalized
1067 proportions of the remaining nine elements, with a subsequent denormalization process applied
1068 for the map reconstruction purposes. This strategic exclusion of the most abundant elements
1069 led to a marked enhancement of approximately 4% in the predictive R^2 accuracy for the other
1070 nine elements. It is important to note, however, that the input vector maintained its original
1071 composition of 11 elements.

1072 This decision to modify the output layer by removing two elements was driven by a clear
1073 rationale: the omnipresent elements, though significant, held less interest for the objectives
1074 of our study compared to the other elements. This approach underscores our commitment to
1075 optimizing the model's performance where it matters most, despite recognizing that alternative
1076 configurations of the output layer might exist.

1077 *Training Set Size and Learning Curve Examination*

1078 Exploring the behavior of the initial MLP DNN involved examining its performance
1079 in relation to the expansion of the training dataset size and the extension of training
1080 duration. The aim was to demonstrate the algorithm's nominal operation during both
1081 training and inference phases by analyzing its learning curves. This included assessing training
1082 and generalization accuracies across a training dataset and the hold-out validation dataset,
1083 respectively. Additionally, identifying the optimal training duration for inference was crucial to
1084 mitigate the risk of overfitting, in line with the guidance provided by Bengio (2015).

1085 Our investigation spanned training sets ranging from 10 to 200 unaugmented data subsets. We
1086 observed a clear positive relationship between increasing the dataset size and the enhancement
1087 of generalization accuracy. As anticipated, this expansion led to a decrease in training accuracy,
1088 a phenomenon depicted in Figure ??.

1089 Add figure that shows the increasing generalization accuracy with increasing training set size.

1090 In parallel, the algorithm's behavior was monitored in terms of its optimization process over
1091 multiple iterations (epochs) across the entire training dataset, employing stochastic gradient
1092 descent to converge to the minimum of the loss function. Analysis of the learning curves
1093 revealed a zenith in predictive performance on the validation dataset after 40 epochs. This
1094 was in contrast to the outcomes observed at 200 epochs of SGD, despite the training dataset's
1095 distribution increasingly aligning with each additional training iteration.

1096 The learning curve depicted in Figure 8 also hints at the potential for further enhancements
 1097 in training predictions, given the rapid ascent observed towards the training's culmination.
 1098 However, to ensure robust inference capabilities, it's imperative to diminish the variance. This
 1099 could potentially be achieved by incorporating a greater number of training examples and/or
 1100 intensifying the regularization measures.

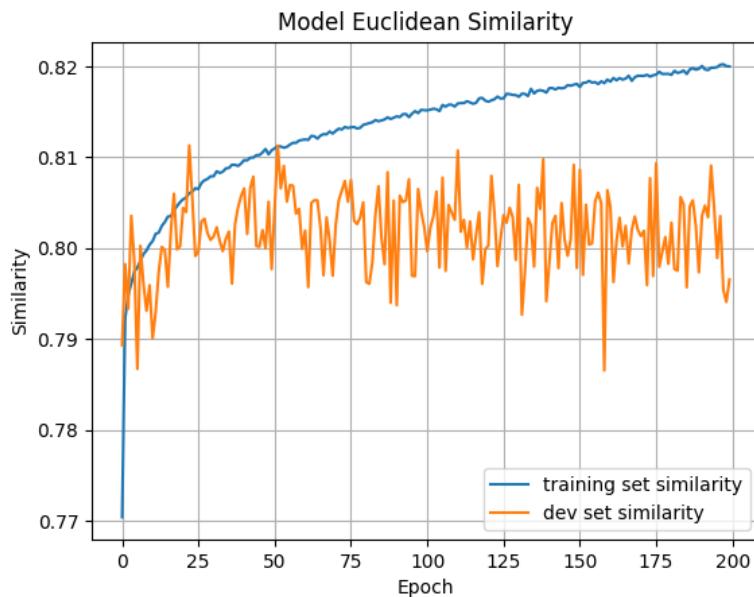


Figure 8. Learning curve for the MLP DNN training. The blue and orange curves show the evolution of the average prediction similarity of the full training dataset (200 subsets, 129,600 data points) and the development hold-out validation dataset (1 subset, 648 data points) respectively.

1101 *Feature Sets Examination*

1102 The evaluation of feature sets played a pivotal role in optimizing the performance of the
 1103 MLP DNN, particularly through the training of the network with various input features across
 1104 the unaugmented 200-subset training dataset. The iterative process of enhancing input features
 1105 resulted in substantial improvements in prediction accuracy for several modified feature sets
 1106 with ES4 going up by about 2%, while R^2 increasing with as much as 5% from F00 to F11.
 1107 This underscores the critical importance of a well-curated and comprehensive feature set in the
 1108 development of neural networks capable of tackling intricate tasks, such as predicting the surface
 1109 elemental compositions.

1110 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.

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

1141 The discerning application of loss functions, particularly the adoption of the KL-divergence for
 1142 evaluating probability-like outputs, marked a jump in performance.

1143 The culmination of our hyperparameter optimization efforts led to the finalization of the
 1144 MLP DNN architecture, characterized by a four-layered structure with 600, 500, 350, and 250
 1145 neural units respectively (Figure ??). An adjustment was made to the regularization coefficient,
 1146 setting it to the found higher value of 1.0×10^{-5} to enhance model generalization. Concurrently,
 1147 the learning rate was optimized to 0.5×10^{-4} , balancing the trade-off between learning speed and
 1148 stability. Training was conducted in mini-batches of 512 examples, a size determined through
 1149 our optimization exercises to be close to optimal. This meticulously optimized structure and
 1150 parameter set represent the culmination of our comprehensive campaign to refine the MLP DNN,
 1151 ensuring it stands as a robust model for our advanced predictive task.

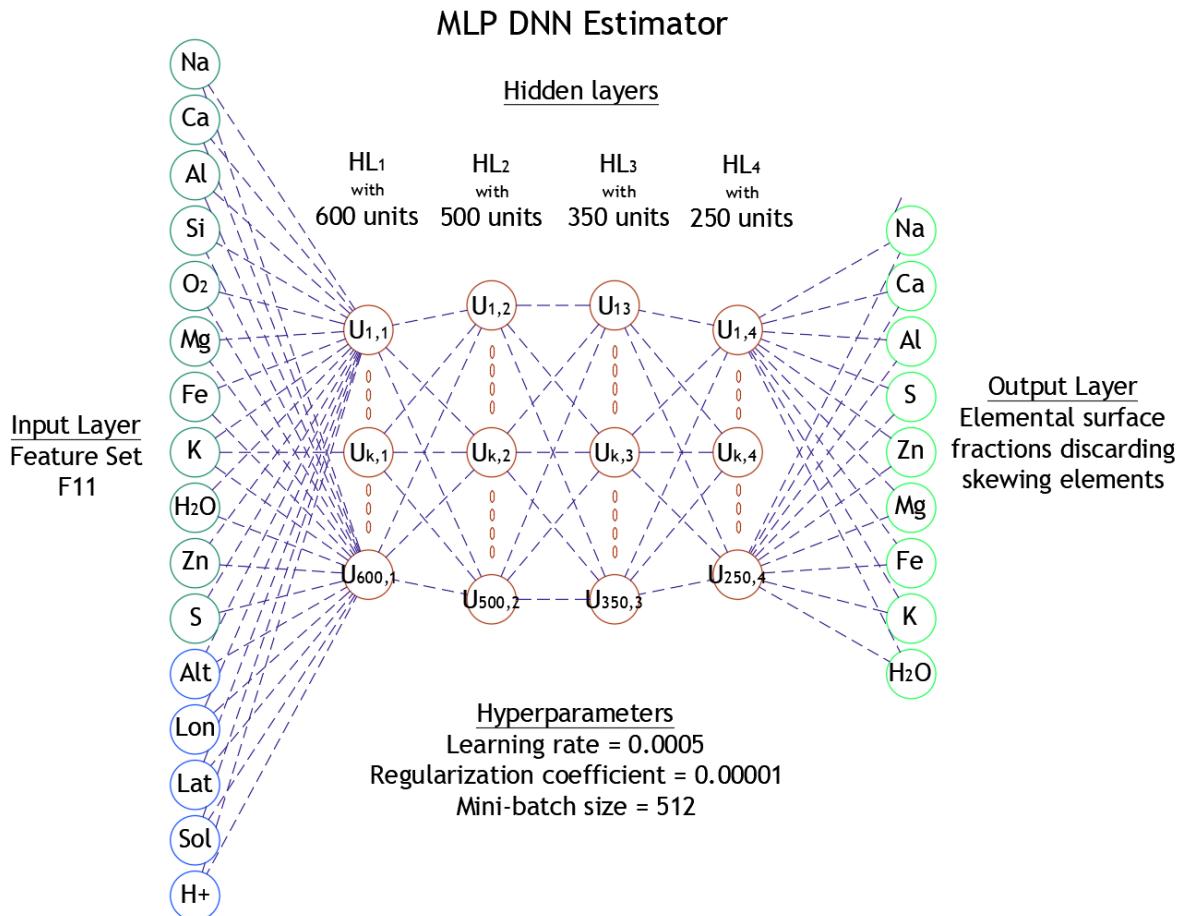


Figure 9. Finalized multilayer perceptron deep neural network. The input layer consists of the features collected in feature set F11, the output layer is adjusted to remove the skewing predictions towards better estimating the O₂ and Si species. There are four hidden layers with 600, 500, 350, and 250 neurons respectively.

1152 *Augmented Data Study*

1153 In the concluding phase of our training campaign, we embarked on a strategic initiative to
1154 enhance the representability of the empirical distribution, thereby aligning it more closely with
1155 the true data-generating distribution—a target that remains inherently elusive due to limited
1156 direct access. This endeavor was pursued through the deliberate augmentation of our training
1157 datasets, an approach that involved the integration of additional examples derived from the same
1158 exospheric observations that constituted our initial datasets. However, these new inclusions were
1159 distinct in their representation of varying altitudes, thereby enriching the diversity and depth
1160 of our training data.

1161 The initial expansion of our dataset to encompass 200 augmented subsets had already
1162 demonstrated significant promise in enhancing the model’s performance. Motivated by these
1163 preliminary successes, we ambitively expanded our dataset even further to include a total of
1164 300 augmented subsets, culminating in an impressive 1,360,800 examples. This substantial
1165 augmentation effort was driven by the rationale that incorporating measurements from varying
1166 altitudes would not only bolster the dataset’s comprehensiveness but also empower our model
1167 to predict with greater accuracy across a diverse range of altitude-specific inputs.

1168 The fruits of this labor were unmistakably positive, with the augmented datasets markedly
1169 improving the robustness and accuracy of our MLP DNN, increasing further the validation set’s
1170 ES4 to 84.0% (+1.5%) and its R² to 63.5% (+3.5%). The strategic inclusion of altitude-varied
1171 examples was particularly impactful, enabling the algorithm to achieve enhanced predictive
1172 precision for inputs across different altitudes.

1173 *Implications and Results of the Training*

1174 As our meticulous exploration of the hyperparameter space culminated in identifying a region
1175 that, while not conclusively the ultimate minimum, demonstrates unparalleled accuracy in
1176 inferences on the hold-out validation dataset, we arrived at several pivotal implications and
1177 results from our training campaign. This journey through hyperparameter optimization has
1178 yielded a collection of finely tuned multilayer perceptron deep neural networks, each reflecting
1179 a nuanced understanding of the underlying data-generating processes.

1180 Firstly, one of the outcomes of this campaign is the demonstration of the algorithm’s efficiency,
1181 achieving optimal training within 40 complete epochs. This not only highlights the effectiveness

1182 of our chosen architecture but also underscores the potential for accuracy improvements with
1183 the expansion of the training dataset. Such findings affirm the architectural decisions made in
1184 designing our MLP DNN for the task at hand.

1185 Secondly, our investigation revealed the critical role of specific features in guiding the
1186 algorithm toward more precise predictions of exospheric measurements and surface composition.
1187 The identification of these key features underscores the importance of thoughtful feature selection
1188 in enhancing model performance.

1189 Thirdly, the exploration led to the refinement of the MLP’s internal structure, significantly
1190 bolstered by experiments with various loss functions and output layers, alongside the application
1191 of Bayesian hyperparameter optimization. While acknowledging that the realm of possible
1192 architectural enhancements remains vast, the current configuration stands as a testament to the
1193 robustness and efficacy of our model.

1194 Lastly, the strategic augmentation of our dataset with additional exospheric observations
1195 has unequivocally improved the algorithm’s predictive capabilities. This expansion not only
1196 enriches the model’s training environment but also enhances its ability to generalize across a
1197 broader spectrum of the empirical distribution, thereby moving closer to the elusive true data-
1198 generating distribution.

1199 The combined efforts of hyperparameter exploration, architectural fine-tuning, and dataset
1200 augmentation have significantly propelled our model’s performance. Through this comprehensive
1201 training campaign, we have not only achieved a high degree of accuracy in our predictions
1202 but also laid a solid foundation for future research to build upon, promising even greater
1203 advancements in our understanding and representation of complex data-generating processes.
1204 Through this concerted effort, we have significantly advanced the model’s capacity to generalize
1205 from the empirical distribution to the true underlying data-generating distribution.

1206 *4.3. Testing Phase*

1207 The ultimate evaluation of our multilayer perceptron deep neural network algorithm’s
1208 performance hinges on its ability to accurately predict surface compositions and reconstruct
1209 elemental surface maps using datasets it has not previously encountered. This phase is crucial
1210 as it tests the algorithm’s generalization capabilities beyond the conditions it was trained under.

1211 Our research incorporates two distinct test campaigns, designed to assess the MLP network’s

predictive prowess. These campaigns were structured to apply the final network, fine-tuned with an extensive training set comprising 300 augmented subsets, across test datasets derived from a variety of altitudes not previously seen during training. This approach ensures a rigorous evaluation of the network's adaptability and accuracy across diverse conditions.

The scope of these test campaigns is broad, focusing not only on aggregate performance metrics across the entire dataset but also on detailed analyses for individual elemental species. This includes a thorough examination of residuals to identify any systematic errors or biases in predictions. Moreover, an essential component of our evaluation is the reconstruction of surface maps. This process entails a visual comparison between the original, or "ground truth", maps and the ones predicted by our algorithm. By doing so, we aim to provide a holistic view of the algorithm's capability in reproducing detailed and accurate surface compositions, thereby illustrating its potential for practical application in exospheric studies and planetary science.

4.3.1. Primary Test Campaign

In our primary test campaign, we embarked on an extensive evaluation using single-simulation test datasets derived from 15 unique surface compositions, each leading to distinct exospheres. This comprehensive 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 meticulously 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.a).

Additionally, our visual comparisons from the map reconstructions (Figures 11, 12, 13, and

4.3 Testing Phase

4 RESULTS

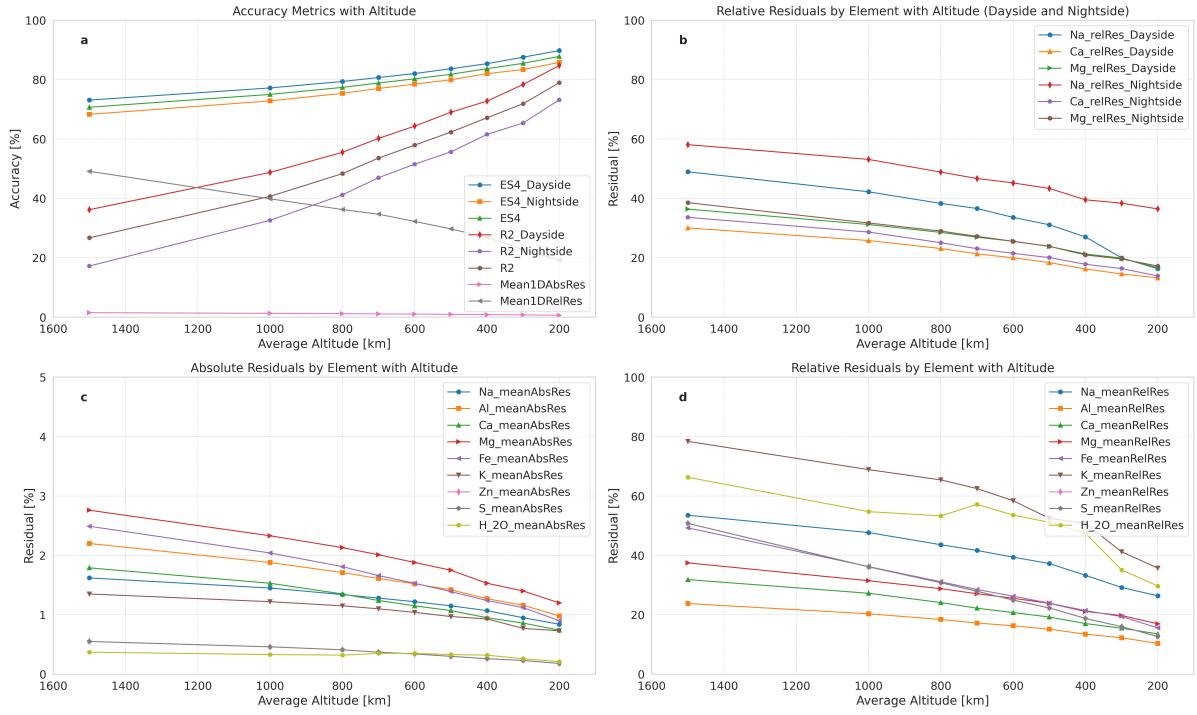


Figure 10. Mean accuracy metrics of the MLP DNN predictions on the 15 test surfaces of the primary test campaign. [This figure needs to be described better.](#)

1242 14) particularly for elements like Magnesium (Mg), Calcium (Ca), and Sodium (Na), highlighted
 1243 the algorithm's strengths and weaknesses in predicting different elemental distributions. We
 1244 observed a notable discrepancy in predictive accuracy between refractory species, such as
 1245 Magnesium and Calcium, and volatile ones, like Sodium. The results further suggested
 1246 that predictions were generally more precise for the dayside, a likely consequence of particle
 1247 movements influenced by solar radiation pressure.

1248 A key finding from our campaign was the impact of altitude on prediction accuracy. Lower
 1249 altitude measurements (200 km) yielded the most accurate reconstructions, with an ES4 reaching
 1250 up to 87.81%, and R-squared reaching up to 78.97%. This accuracy diminished at higher
 1251 altitudes, attributed to the exosphere's dynamic nature and the increased complexity in tracing
 1252 back exospheric particles to their originating surface tiles.

1253 The challenge of accurately predicting night side compositions, coupled with the decreasing
 1254 accuracy with altitude, underscored the need for a secondary test campaign. This
 1255 subsequent campaign aims to focus on reconstructing comprehensive surface maps from dayside
 1256 measurements during consecutive Mercury orbits, addressing the identified gaps and leveraging

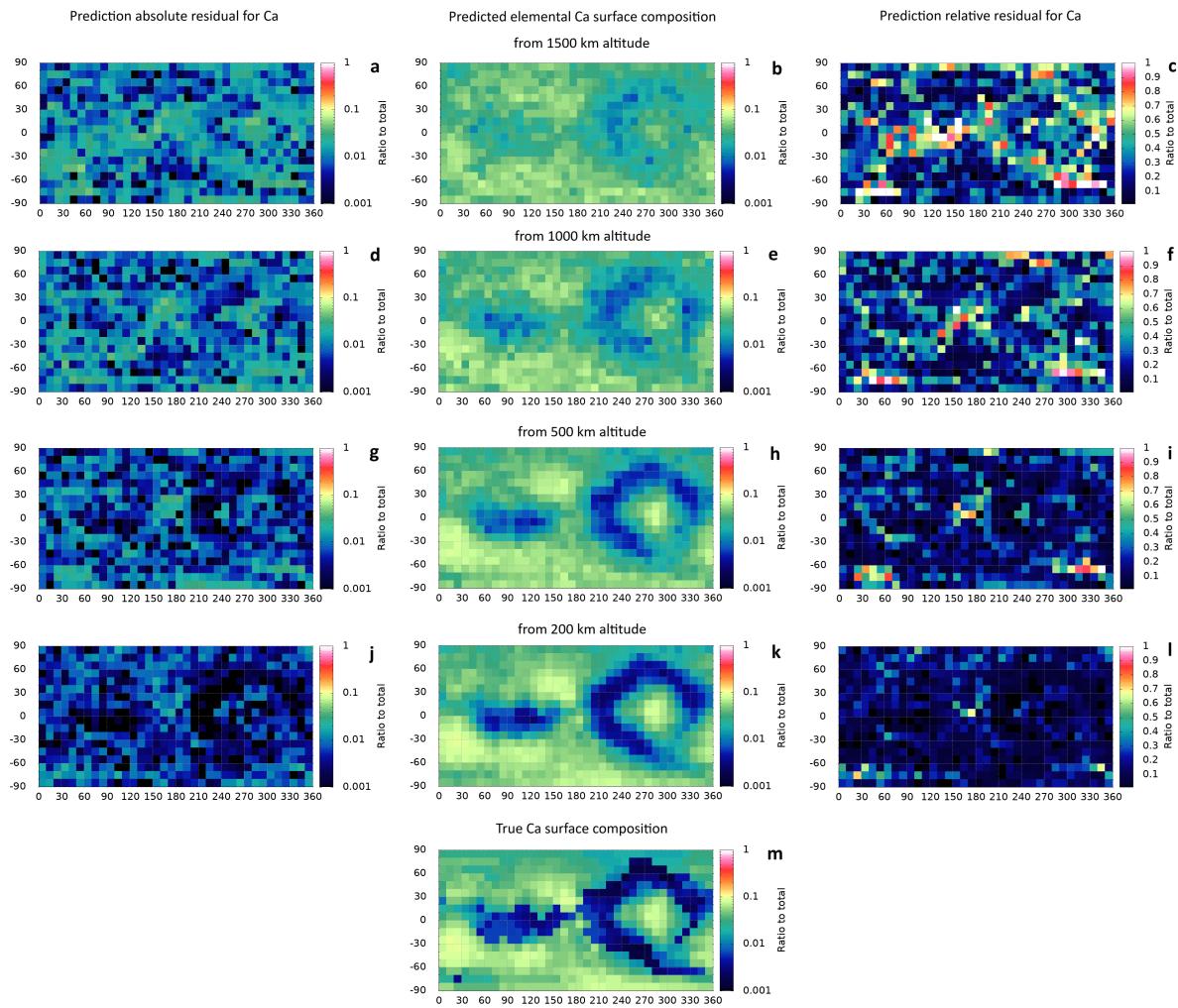


Figure 11. Map representations of the predictions by the ultimately trained MLP DNN of the same sample Calcium surface composition produced from inputs coming from the same simulated exosphere, but at different altitude levels (1500, 1000, 500, and 200 km). The panels in the middle show the reconstructed surface composition. The panels on the left show the absolute residuals to the "ground truth" surface composition (panel m). The panels on the right show the relative residuals per altitude level. Add improvements numbers wrt above panel.

1257 insights gained from the primary campaign to refine our predictions further.

1258 Discuss that the water ice prediction anomaly in the graph is probably due to the low overall
 1259 fraction of this "element" on the surface and/or it's a particularly impactful region in the
 1260 exosphere for the water vapor (?). At the same time, the absolute residuals are fairly close.
 1261 Additionally, the training data points have altitude of 500 and 800 km, but nothing in between,
 1262 so probably this part of the distribution is missed. Discuss that the coupling of Zn and S in our
 1263 model probably helps to predict them better even though their fraction on the surface is also

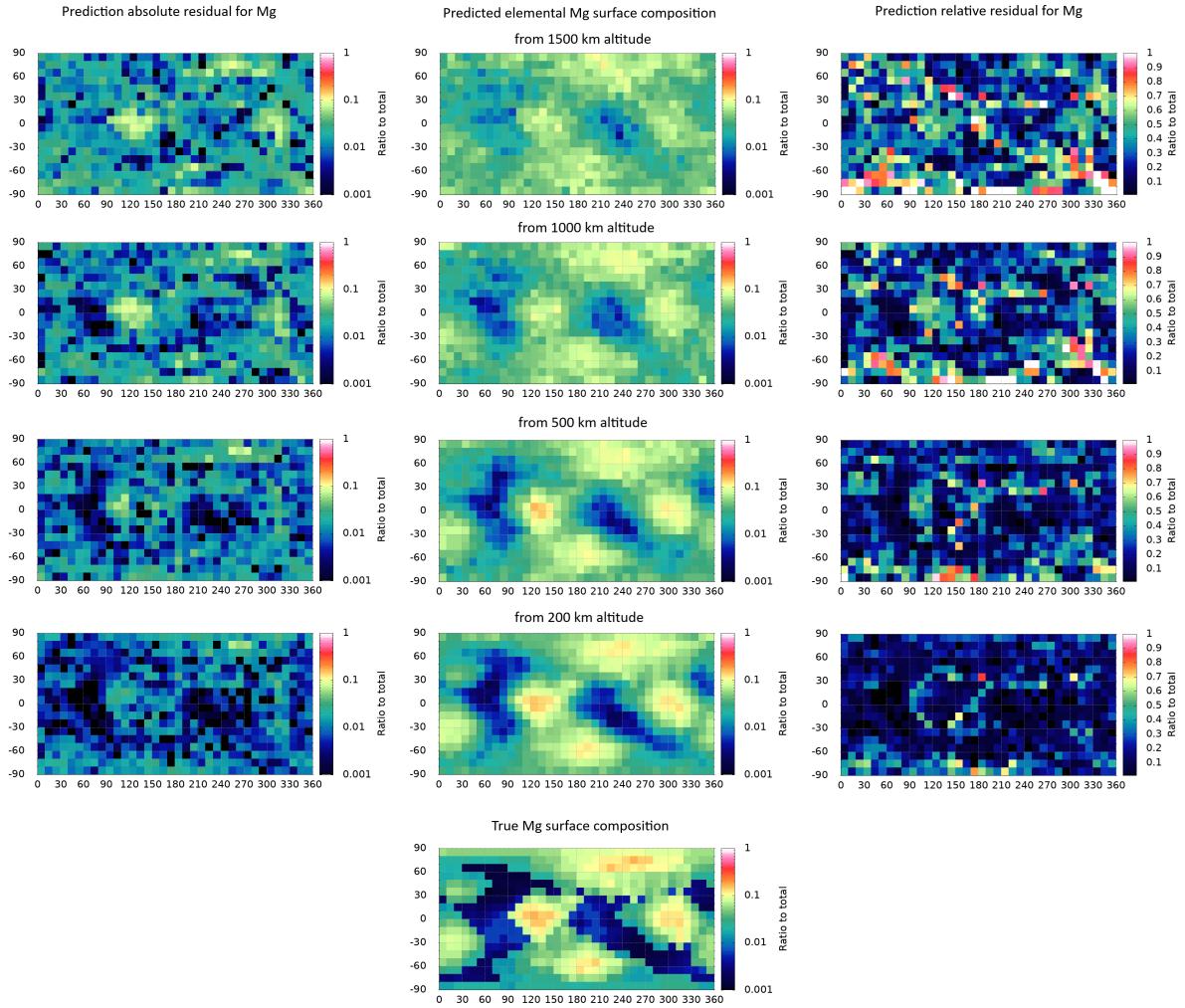


Figure 12. Map representations of the predictions by the ultimately trained MLP DNN of the same sample Magnesium surface composition produced from inputs coming from the same simulated exosphere, but at different altitude levels (1500, 1000, 500, and 200 km). The panels in the middle show the reconstructed surface composition. The panels on the left show the absolute residuals to the "ground truth" surface composition (panel m). The panels on the right show the relative residuals per altitude level. **This figure needs to be polished a little bit - panel letters. Add improvements numbers wrt above panel.**

1264 relatively low.

1265 Explain the map implications: certain large scale patterns are recognized from as far as 1500
 1266 km, even if details are missed by the MLP. The algorithm resolves with a good accuracy even at
 1267 500 km, especially for refractory species (Ca and Mg), but going closer to the planet produces
 1268 the most accurate predictions and reconstructions, even for volatiles (Na). At the same time a
 1269 pattern can be seen in the reconstructed maps and the residual maps, especially for Na, that
 1270 the predictions on the night side are less accurate.

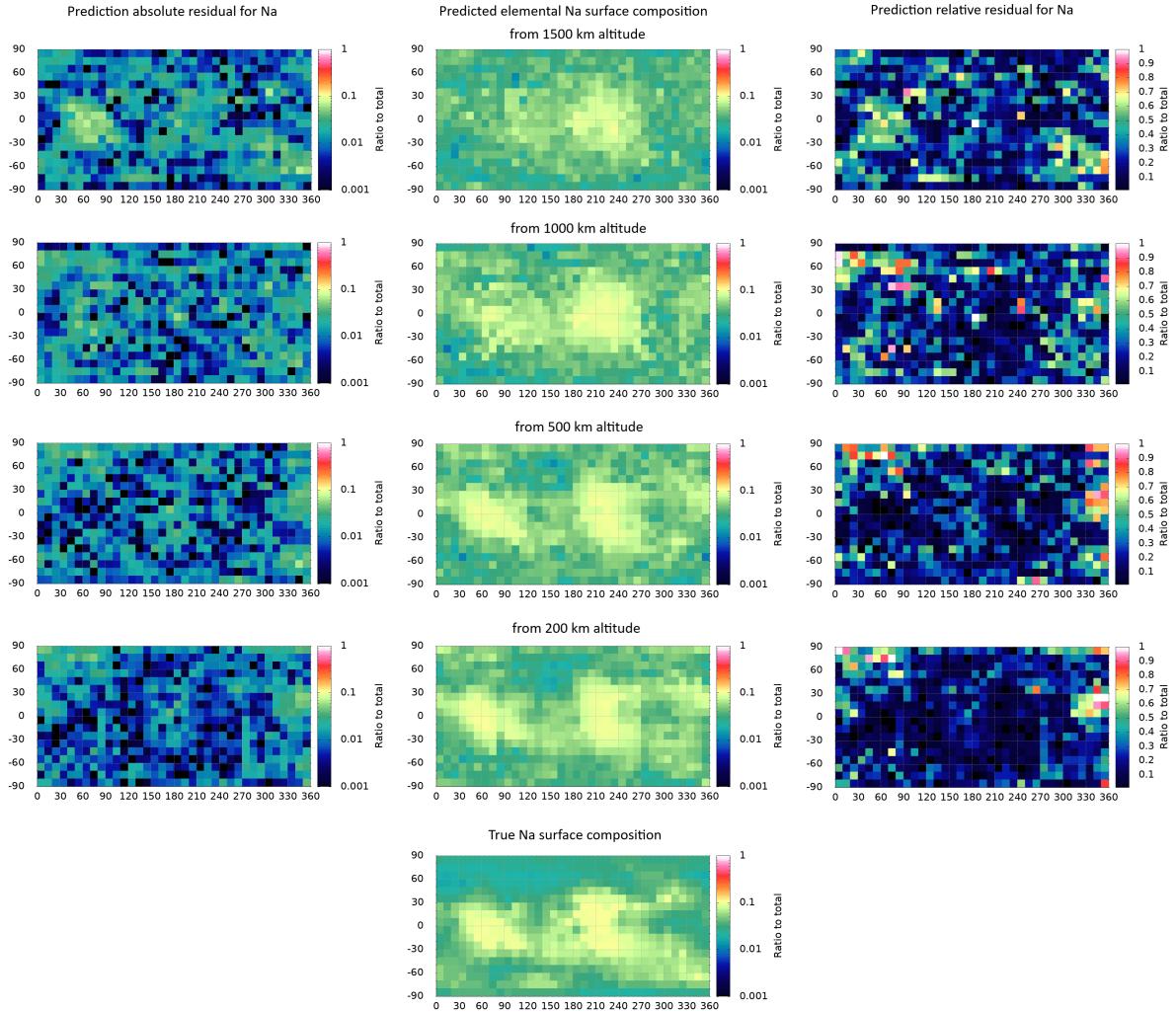


Figure 13. Map representations of the predictions by the ultimately trained MLP DNN of the same sample Sodium surface composition produced from inputs coming from the same simulated exosphere, but at different altitude levels (1500, 1000, 500, and 200 km). The panels in the middle show the reconstructed surface composition. The panels on the left show the absolute residuals to the "ground truth" surface composition (panel m). The panels on the right show the relative residuals per altitude level. **This figure needs to be polished a little bit - panel letters. Add improvements numbers wrt above panel.**

1271 4.3.2. Secondary Test Campaign

1272 The secondary test campaign was strategically designed to leverage the unique rotational
 1273 and orbital dynamics of Mercury, particularly its 3:2 orbit-spin resonance, to observe the
 1274 entire planetary surface under daylight conditions across two successive Mercury years. This
 1275 approach utilized double-simulation compound predictions, focusing on the same 15 surface
 1276 compositions identified in the primary campaign but observed at two consecutive perihelia.

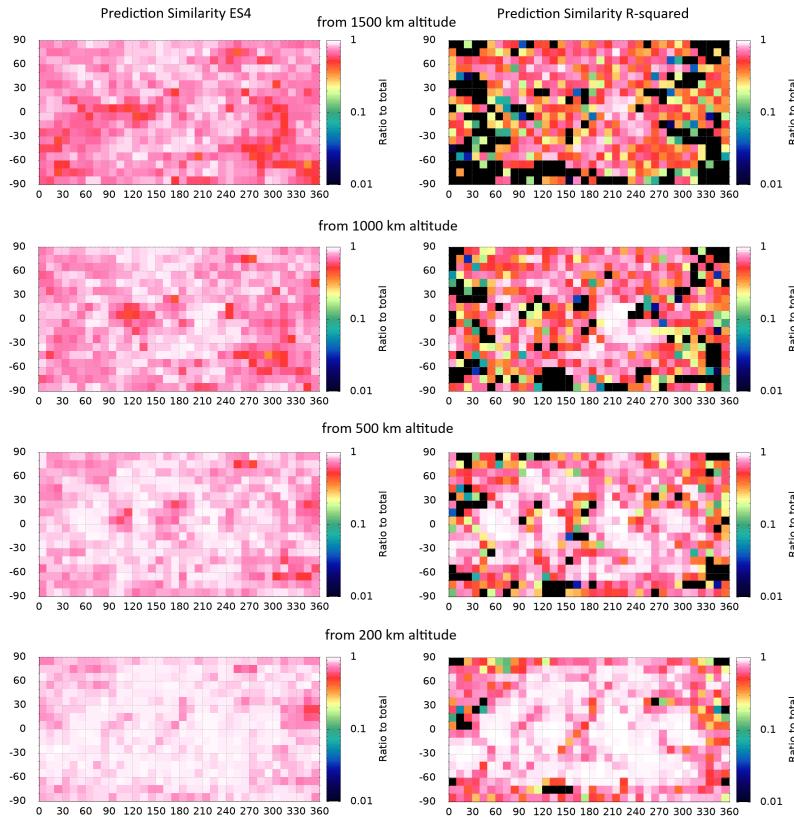


Figure 14. Map representations of the accuracy metrics evaluated for the predictions by the ultimately trained MLP DNN of the same sample total surface composition produced from inputs coming from the same simulated exosphere, but at different altitude levels (1500, 1000, 500, and 200 km). The panels on the left show the ES4 similarity. The panels on the right show the R-squared metrics per altitude level. **This figure needs to be polished a little bit - panel letters. Add improvements numbers from figure above.**

1277 During these periods, different halves of Mercury’s surface were illuminated by the Sun, allowing
 1278 for comprehensive daylight observation of the entire planet over the two simulations. For this
 1279 campaign, predictions specifically targeted sunlit surface tiles, enabling an in-depth analysis of
 1280 surface compositions that were previously obscured by darkness in the initial test phase.

1281 Measurements for this campaign were again taken at a range of altitudes from 200 to
 1282 1500 km, maintaining consistency with the primary campaign’s methodology. This consistent
 1283 approach across varied altitudes allowed for a **nuanced comparison of predictive performance**
 1284 **under different observational conditions.**

1285 A significant outcome of the secondary campaign was the improved accuracy in predicting
 1286 volatile species’ distributions, aligning more closely with the refractory species’ predictions
 1287 observed in the primary campaign. This enhancement in predictive accuracy for volatiles under

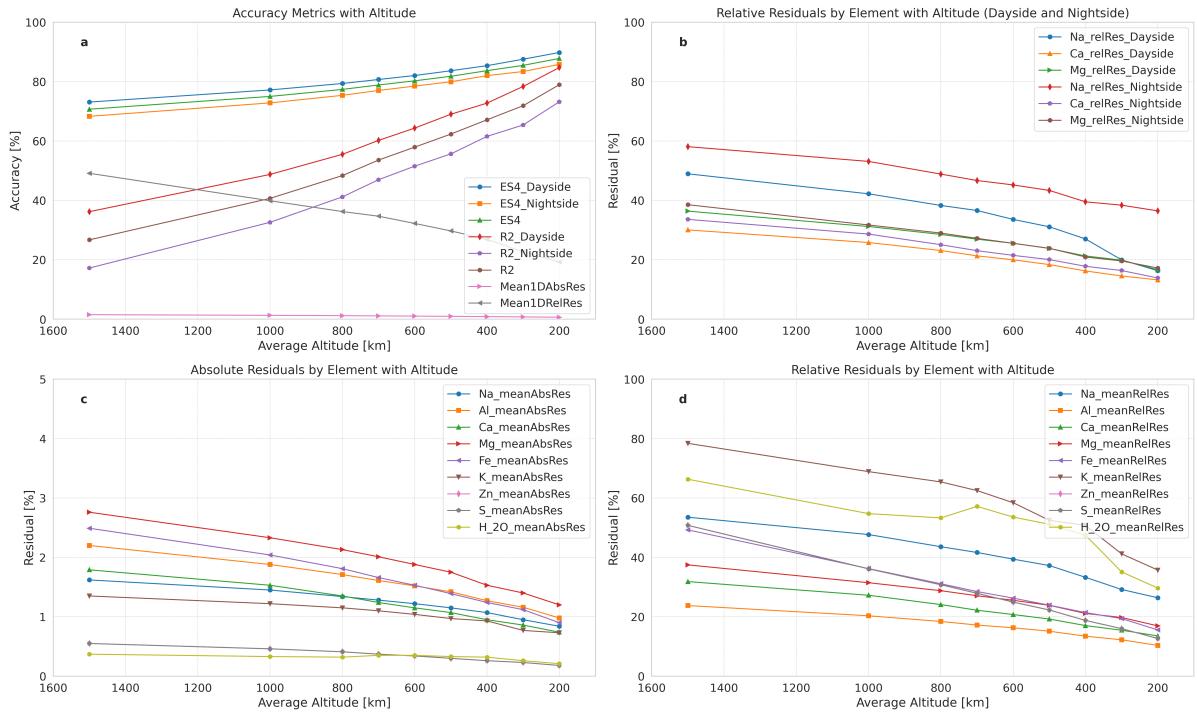


Figure 15. Mean accuracy metrics of the MLP DNN predictions on the 15 test surfaces of the primary test campaign. This figure needs to be described better and possibly compared to figure from Primary campaign.

1288 daylight conditions underscores the importance of solar illumination in accurately assessing
1289 surface compositions.

1290 Add also some figures depicting the improvements in the metrics compared to the primary
1291 campaign - overall and per element.

1292 Moreover, the secondary campaign demonstrated a notable increase in overall prediction
1293 and map reconstruction accuracy, with the average ES4 metric reaching approximately 89.70%
1294 and the R-squared metric reaching 83.41% at the lowest altitude of 200 km (Figure 15).
1295 This marked improvement highlights the efficacy of considering Mercury's solar exposure in
1296 enhancing predictive models' accuracy. By focusing solely on the dayside observations across two
1297 perihelia, the campaign effectively capitalized on the optimal conditions for surface composition
1298 reconstruction.

1299 Add map reconstruction and map reconstruction for Magnesium???

1300 Add map reconstruction and map reconstruction for Calcium???

1301 5. Discussion and Future Work

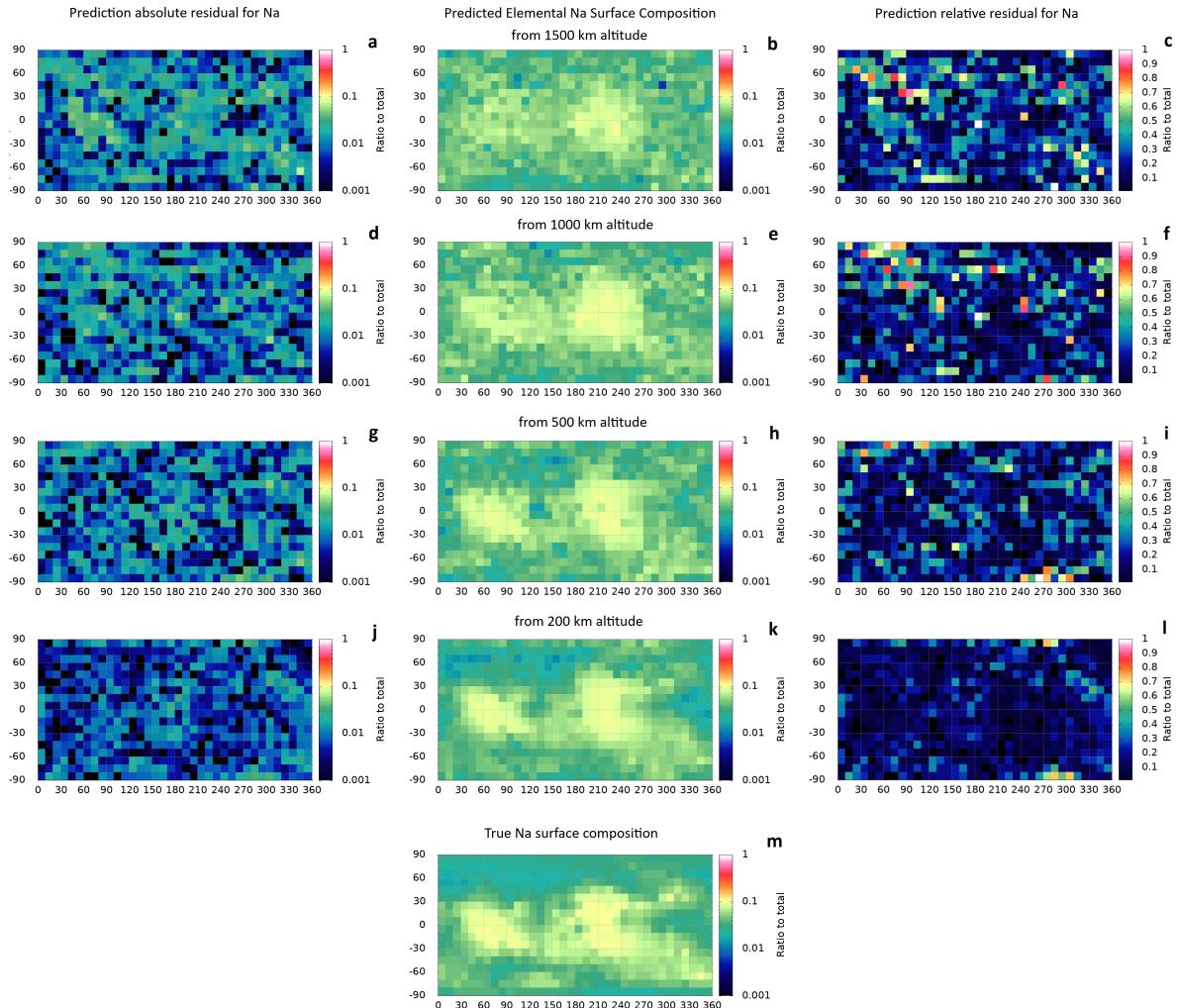


Figure 16. Map representations of the predictions by the ultimately trained MLP DNN of the same sample Sodium surface composition produced from inputs coming from the dayside of two simulated exospheres from two consecutive Mercury perihelia, but at different altitude levels (1500, 1000, 500, and 200 km). The panels in the middle show the reconstructed surface composition. The panels on the left show the absolute residuals to the "ground truth" surface composition (panel m). The panels on the right show the relative residuals per altitude level. Eclipsed areas in the primary campaign (0-90 and 270-360 Lon) are markedly more accurately reconstructed in these dayside only combined maps. Add improvements numbers from figure above and from same altitude level but from only one Mercury perihelion.

1302 The discussion may need to be revised and some repeating statements shortened or removed,
 1303 especially repeating claims on the potential implications to planetary science.

1304 This research presents a novel approach for deducing Mercury's surface composition by
 1305 analyzing exospheric density measurements through advanced deep neural network models. By
 1306 significantly enhancing the methodologies of previous studies, notably the one by Kazakov et al.

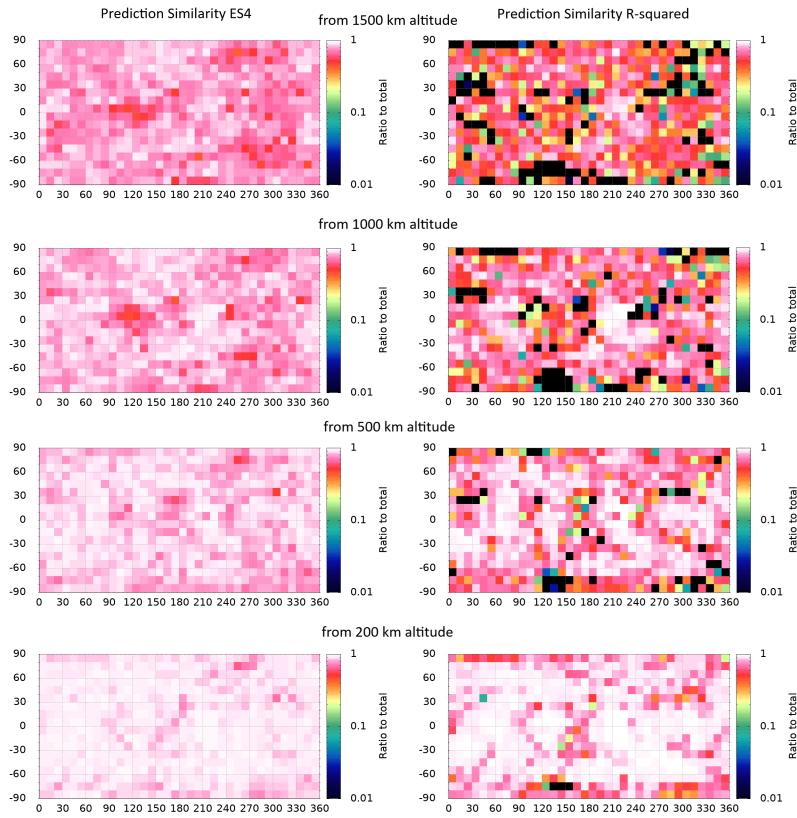


Figure 17. Map representations of the accuracy metrics evaluated for the predictions by the ultimately trained MLP DNN of the same sample total surface composition produced from inputs coming from the dayside of two simulated exospheres from two consecutive Mercury perihelia, but at different altitude levels (1500, 1000, 500, and 200 km). The panels on the left show the ES4 similarity. The panels on the right show the R-squared metrics per altitude level. **This figure needs to be polished a little bit - panel letters. Add improvements numbers from figure above and from same altitude level but from only one Mercury perihelion.**

1307 (2020), this study has pushed the boundaries of surface composition analysis. The improvements
 1308 include a detailed refinement of the simulated model's parameter space, the development of
 1309 a more sophisticated multilayer perceptron DNN architecture, the application of Bayesian
 1310 hyperparameter tuning for optimal network configuration, and the incorporation of domain-
 1311 specific knowledge into the feature selection process.

1312 The core achievement of this study is the formulation of an algorithm that accurately
 1313 estimates surface-exosphere interactions on Mercury. This algorithm is adept at predicting
 1314 elemental surface compositions and reconstructing detailed surface maps beneath exospheric
 1315 density measurements. Such an estimator is pivotal for understanding the mechanisms that
 1316 propel neutral atoms from the planet's regolith into its exosphere.

5 DISCUSSION AND FUTURE WORK

1317 Through rigorous training and testing phases, the study explored the extents of the
1318 hyperparameter space, leading to models that excel in generalizing predictions across previously
1319 unseen datasets. The augmentation of training datasets with additional examples played a
1320 crucial role in enhancing prediction accuracy.

1321 The investigation comprised two main test campaigns. The primary campaign highlighted
1322 the model's superior predictive performance on the planet's dayside, although predictions for
1323 the nightside, particularly for volatile elements like sodium (Na) and potassium (K), were less
1324 precise. This discrepancy led to a secondary campaign, focusing on previously shadowed surface
1325 areas during subsequent Mercury perihelia, when these areas were illuminated. By integrating
1326 simulated daylight observations from two Mercury years, the study significantly improved
1327 predictions for volatile species and achieved a comprehensive reconstruction of Mercury's surface.

1328 The algorithms developed for surface composition reconstruction exhibited high fidelity in
1329 their predictions, with a similarity metric (ES4) reaching 89.70% accuracy, R-squared of 83.41%,
1330 and average residuals of ... %, at an altitude of 200 km. The reconstruction of relative elemental
1331 composition maps, particularly from low-altitude measurements (200km to 500km), underscored
1332 the algorithms' effectiveness.

1333 The extensive training and testing campaigns conducted in this study highlight a vast
1334 potential for refining and enhancing the deep neural network algorithms dedicated to surface
1335 reconstruction tasks. Acknowledging this, we have outlined a future development roadmap to
1336 further advance these algorithms, particularly in preparation for their application to actual
1337 measurements gathered in orbit around Mercury. One of the significant improvements realized
1338 in this study stems from the implementation of an advanced feature set. This success leads us to
1339 explore deeper into feature engineering, applying more domain-specific knowledge to optimize
1340 the input parameters. Such enhancements aim to provide the algorithm with a more informed
1341 starting point for accurately representing the empirical data distribution.

1342 The exploration and elaboration of the hyperparameter space emerge as crucial next steps
1343 in our developmental roadmap. This entails a comprehensive construction of an expanded
1344 hyperparameter space, which will delve into more nuanced aspects previously only briefly
1345 considered. Among these are the connectivity of the network layers, the selection and
1346 optimization of loss functions, the functions used within the hidden and output units, and
1347 other critical parameters that define the algorithm's structure and behavior. Furthermore, the

5 DISCUSSION AND FUTURE WORK

1348 possibility of developing alternative DNN architectures holds promise for pushing the boundaries
1349 of what these models can achieve in terms of application, accuracy and reliability.

1350 Another critical area of future research involves testing the algorithms against data
1351 distributions originating from diverse regions within the physical processes parameter space.
1352 This exploration is vital for assessing the versatility and adaptability of the algorithms, ensuring
1353 they maintain high performance levels even when applied to different models of exospheric
1354 production from the surface. Evaluating the algorithms' efficacy in these new contexts will
1355 provide valuable insights into their potential limitations and highlight opportunities for targeted
1356 improvements.

1357 Expanding and elaborating the physical processes parameter space for our algorithm
1358 represents a crucial frontier for future research. This endeavor is fundamental for aligning
1359 the algorithm more closely with the intricate realities governing the interactions between a
1360 planet's surface and its exosphere. A key aspect of this exploration involves unraveling some of
1361 the compressed dimensions within the current physical parameter space. By doing so, we aim
1362 to refine our model to more accurately mirror the complex dynamics of real-world processes.
1363 Certain dimensions related to surface-exosphere interactions, which were previously simplified or
1364 "flattened" in our model, likely hold other important insights into the mechanisms of planetary
1365 science. The current model's limitation in representing these dimensions means that valuable
1366 information might be overlooked, hampering the estimator's ability to capture the full spectrum
1367 of these interactions.

1368 Moreover, the development of specialized estimators for varying simulated processes presents
1369 an innovative path forward. This strategy entails the creation of multiple, distinct estimators,
1370 each designed to excel within specific segments of the parameter space. Such a tailored approach
1371 allows for adjustments to the inherent variability within the parameter space, promising more
1372 accurate and nuanced predictions.

1373 The contrast between utilizing simulated data and incorporating real observational data into
1374 our algorithm development also warrants attention. Although simulations offer a manageable
1375 setting for examining different scenarios, they fall short of embodying the full complexity and
1376 unpredictability found in actual exospheric data. Consequently, the parameter space defined
1377 by real observations may diverge substantially from that used in our simulations. Despite
1378 the model's capability to closely emulate real processes, the nature of the parameter space

5 DISCUSSION AND FUTURE WORK

1379 could fluctuate over time or involve interdependent variables that are not fully accounted for in
1380 simulations.

1381 Transitioning our focus from simulated to real physical processes represents a bold, yet
1382 potentially transformative, research direction. Developing an estimator that can effectively
1383 process and analyze real-world data from Mercury's exosphere would signify a profound leap
1384 forward in our comprehension of planetary surfaces and their interactions with surrounding
1385 environments. Such advancement would not only enhance our theoretical understanding but
1386 also provide practical insights into the behavior and evolution of planetary exospheres, bridging
1387 the gap between theoretical models and the complex realities of space environments.

1388 In conclusion, the research presented herein has established a robust groundwork for
1389 advancing our understanding of planetary surface-exosphere interactions, particularly around
1390 Mercury. The exploration of novel methodologies for reconstructing surface compositions using
1391 exospheric measurements as inputs to deep neural networks marks a significant step forward.
1392 This work not only enhances the capabilities of estimators for such tasks but also broadens our
1393 comprehension of planetary science as a whole.

1394 The implications of applying these methods to real observational data are profound.
1395 As real data becomes accessible, particularly from missions like BepiColombo with its
1396 suite of instruments (MPO/SERENA, MPO/MIXS, MPO/MGNS, MPO/MERTIS, and
1397 MPO/SIMBIO-SYS), this approach offers a promising avenue for refining models of exosphere
1398 generation. By comparing predicted surface compositions against measurements, we can
1399 constrain models more accurately, thereby advancing our knowledge of planetary processes.

1400 The integration of this method into the data analysis framework of the SERENA ground
1401 system represents an exciting development. With an advanced AI algorithm in place, the
1402 mission's ability to uncover insights into the origins and dynamics of Mercury's exosphere is
1403 significantly enhanced. Furthermore, the continued refinement and sophistication of the deep
1404 neural networks introduced in this study hold the potential to revolutionize our approach to
1405 studying not only Mercury but also other planetary bodies within our Solar System. Through
1406 such advancements, we gain powerful new tools for probing and understanding the myriad
1407 physical processes that govern the environments of celestial objects.

1408 **Acknowledgments**

1409 This work was supported by the Italian Space Agency (ASI) - SERENA contract 2018-8-HH.O

1410 "Scientific participation in the mission BepiColombo SERENA - Phase E1". Is this still the

1411 right contract?

APPENDIX A MEAN MINERAL AND ELEMENTAL FRACTIONS IN THE DATASETS

¹⁴¹² **Appendix A. Mean mineral and elemental fractions in the datasets**

¹⁴¹³ Add table for the mean fractions for 10-300 training datasets, validation dataset, all 15 test

¹⁴¹⁴ datasets.

APPENDIX B TRAINING SUPPLEMENTARY RESULTS

¹⁴¹⁵ **Appendix B. Training Supplementary Results**

¹⁴¹⁶ Add table of trainings?

¹⁴¹⁷ How many total training runs?

¹⁴¹⁸ How much total time and average time?

¹⁴¹⁹ Add some other training graphs and information here.

APPENDIX C TEST CAMPAIGNS SUPPLEMENTARY RESULTS

¹⁴²⁰ **Appendix C. Test Campaigns Supplementary Results**

¹⁴²¹ Add stats graphs for each element?

¹⁴²² Add reconstructed maps for the other 6 elemental species.

1423 **References**

- 1424 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis,
1425 A.and Dean, J., Devin, M., et al. (2015). Tensorflow: Large-scale machine learning on
1426 heterogeneous systems. *Software available from tensorflow.org*.
- 1427 Bengio, Y. (2015). Early inference in energy-based models approximates back-propagation.
1428 Technical Report arXiv:1510.02777, Université de Montréal.
- 1429 Benkhoff, J., van Casteren, J., Hayakawa, H., Fujimoto, M., Laakso, H., Novara, M., Ferri,
1430 P., Middleton, H. R., and Ziethe, R. (2010). Bepicolombo - comprehensive exploration of
1431 mercury: Mission overview and science goals. *Planetary and Space Science*, 58(1):2–20.
- 1432 Berezhnoy, A. (2018). Chemistry of impact events on mercury. *Icarus*, 300:210–222.
- 1433 Bergstra, J., Bardenet, R., Bengio, Y., and Balazs, K. (2011). Algorithms for hyper-parameter
1434 optimization. In *Proceedings of the 24th International Conference on Neural Information
1435 Processing Systems*, pages 2546–2554.
- 1436 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 1437 Cassidy, T. A., Merkel, A. W., Burger, M. H., Sarantos, M., Killen, R. M., McClintock,
1438 W. E., and Vervack, R. J. (2015). Mercury’s seasonal sodium exosphere: Messenger orbital
1439 observations. *Icarus*, 248:547–559.
- 1440 Chollet, F. et al. (2015). Keras. *Github. GitHub repository*.
- 1441 Cintala, M. J. (1992). Impact-induced thermal effects in the lunar and mercurian regoliths. *J.
1442 Geophys. Res.*, 97:947–973.
- 1443 Ciresan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep big simple
1444 neural nets for handwritten digit recognition. *Neural Computation*, 22:1–14.
- 1445 Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-Interscience.
- 1446 Domingue, D. L., Koehn, P. L., Killen, R. M., et al. (2007). Mercury’s atmosphere: A surface-
1447 bounded exosphere. *Space Science Reviews*, 131(1-4):161–186.
- 1448 Gamborino, D., Vorburger, A., and Wurz, P. (2019). Mercury’s subsolar sodium exosphere:
1449 An ab initio calculation to interpret mascs/uvvs observations from messenger. *Annales
1450 Geophysicae*, 37:455–470.
- 1451 Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In

REFERENCES

REFERENCES

- 1452 *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*,
1453 volume 15, pages 315–323.
- 1454 Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge,
1455 MA.
- 1456 Head, T. et al. (2018). scikit-optimize/scikit-optimize: v0.5.2. Zenodo.
- 1457 Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*,
1458 11(10):428–434.
- 1459 James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical
1460 Learning: with Applications in R*. Springer.
- 1461 Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory
1462 and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.
- 1463 Johnson, R., Leblanc, F., Yakshinskiy, B., and Madey, T. (2002). Energy distributions for
1464 desorption of sodium and potassium from ice: the na/k ratio at europa. *Icarus*, 156:136–
1465 142.
- 1466 Kazakov, A. N. et al. (2020). Deep neural networks for analysis of mercury’s planetary exosphere.
1467 *Journal of Physics: Conference Series*, 1548:012–014.
- 1468 Killen, R. (2016). Pathways for energization of ca in mercury’s exosphere. *Icarus*, 268:32–36.
- 1469 Killen, R. M. and Burger, M. H. (2019). Understanding mercury’s exosphere: Models derived
1470 from messenger observations. In *Mercury: The View After MESSENGER*. Cambridge
1471 University Press, London.
- 1472 Killen, R. M., Cremonese, G., Lammer, H., et al. (2007). Processes that promote and deplete
1473 the exosphere of mercury. *Space Science Reviews*, 132(2-4):433–509.
- 1474 Killen, R. M., Potter, A. E., Reiff, P., Sarantos, M., Jackson, B. V., Hick, P., and Giles, B.
1475 (2001). Evidence for space weather at mercury. *J. Geophys. Res.*, 106:20509–20525.
- 1476 Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. In *International
1477 Conference on Learning Representations*.
- 1478 Leblanc, F., Delcourt, D., and Johnson, R. E. (2003). Mercury’s sodium exosphere:
1479 Magnetospheric ion recycling. *Journal of Geophysical Research*, 108:5136.
- 1480 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.

REFERENCES

REFERENCES

- 1481 Mangano, V. et al. (2015). Themis na exosphere observations of mercury and their correlation
1482 with in-situ magnetic field measurements by messenger. *Planetary and Space Science*, 115:102–
1483 109.
- 1484 Merkel, A. W., Vervack, R. J. J., Killen, R. M., Cassidy, T. A., McClintonck, W. E., Nittler,
1485 L. R., and Burger, M. H. (2018). Evidence connecting mercury’s magnesium exosphere to its
1486 magnesium-rich surface terrane. *Geophysical Research Letters*, 45:6709–6797.
- 1487 Milillo, A., Fujimoto, M., Kallio, E., Kameda, S., Leblanc, F., Narita, Y., Cremonese, G., Laakso,
1488 H., Laurenza, M., Massetti, S., McKenna-Lawlor, S., Mura, A., Nakamura, R., Omura, Y.,
1489 Rothery, D., Seki, K., Storini, M., Wurz, P., Baumjohann, W., and Sprague, A. (2010).
1490 The bepicolombo mission: An outstanding tool for investigating the hermean environment.
1491 *Planetary and Space Science*, 58:40–60.
- 1492 Milillo, A., Fujimoto, M., Murakami, G., et al. (2020). Investigating mercury’s environment
1493 with the two-spacecraft bepicolombo mission. *Space Science Reviews*, 216. Article no. 93.
- 1494 Milillo, A. and Wurz, P. (2014). *SERENA Science Performance Report*.
- 1495 Milillo, A., Wurz, P., Orsini, S., Delcourt, D., Kallio, E., Killen, R., Lammer, H., Massetti,
1496 S., Mura, A., Barabash, S., Cremonese, G., Daglis, I., De Angelis, E., Di Lellis, A., Livi, S.,
1497 Mangano, V., and Torkar, K. (2005). Surface-exosphere-magnetosphere system of mercury.
1498 *Space Science Reviews*, 117:397–443.
- 1499 Minsky, M. and Papert, S. A. (2017). *Perceptrons: An Introduction to Computational Geometry*.
1500 MIT Press, Cambridge, MA.
- 1501 Mura, A., Milillo, A., Orsini, S., and Massetti, S. (2007). Numerical and analytical model of
1502 mercury’s exosphere: Dependence on surface and external conditions. *Planetary and Space
1503 Science*, 55:1569–1583.
- 1504 Mura, A., Wurz, P., Lichtenegger, H., Schleicher, H., Lammer, H., Delcourt, D., Milillo, A.,
1505 Orsini, S., Massetti, S., and Khodachenko, M. (2009). The sodium exosphere of mercury:
1506 Comparison between observations during mercury’s transit and model results. *Icarus*, 200:1–
1507 11.
- 1508 Orsini, S., Livi, S., Lichtenegger, H., et al. (2020). Serena: Particle instrument suite for sun–
1509 mercury interaction insights on-board bepicolombo. *Space Science Reviews*. In Press.
- 1510 Orsini, S., Livi, S., Torkar, K., Barabash, S., Milillo, A., Wurz, P., Di Lellis, A. M., and Kallio,

REFERENCES

REFERENCES

- 1511 E. (2010). Serena: A suite of four instruments (elenा, strofio, picam and mipa) on board
1512 bepicolombo-mpo for particle detection in the hermean environment. *Planetary and Space*
1513 *Science*, 58:166–181.
- 1514 Plainaki, C., Mura, A., Milillo, A., Orsini, S., Livi, S., Mangano, V., Massetti, S., Rispoli, R.,
1515 and De Angelis, E. (2017). Investigation of the possible effects of comet encke’s meteoroid
1516 stream on the ca exosphere of mercury. *Journal of Geophysical Research: Planets*, 122:1217–
1517 1226.
- 1518 Pokorný, P., Sarantos, M., and Janches, D. (2017). Reconciling the dawn–dusk asymmetry
1519 in mercury’s exosphere with the micrometeoroid impact directionality. *The Astrophysical*
1520 *Journal Letters*, 842(2):L17.
- 1521 Rothery, D. A., Massironi, M., Alemanno, G., et al. (2020). Rationale for bepicolombo studies
1522 of mercury’s surface and composition. *Space Science Reviews*, 216:66.
- 1523 Rumelhart, D., Hinton, G., and Williams, R. (1986a). Learning representations by back-
1524 propagating errors. *Nature*, 323:533–536.
- 1525 Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). Learning internal representations
1526 by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed*
1527 *Processing*, volume 1, chapter 8, pages 318–362. MIT Press, Cambridge.
- 1528 Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, 3rd edition*.
1529 Prentice Hall Press, USA.
- 1530 Sarantos, M., Slavin, J., Benna, M., Boardsen, S., Killen, R., Schriver, D., and Trávníček, P.
1531 (2009). Sodium-ion pickup observed above the magnetopause during messenger’s first mercury
1532 flyby: Constraints on neutral exospheric models. *Geophysical Research Letters*, 36.
- 1533 Schaible, M. J. et al. (2017). Solar wind sputtering rates of small bodies and ion mass
1534 spectrometry detection of secondary ions. *Journal of Geophysical Research: Planets*, 122:1968–
1535 1983.
- 1536 Wurz, P. and Lammer, H. (2003). Monte-carlo simulation of mercury’s exosphere. *Icarus*,
1537 164:1–13.
- 1538 Wurz, P., Whitby, J. A., Rohner, U., Martin-Fernandez, J. A., Lammer, H., and Kolb, C. (2010).
1539 Self-consistent modelling of mercury’s exosphere by sputtering, micrometeorite impact and
1540 photon-stimulated desorption. *Planetary and Space Science*, 58:1599–1616.