



Unsupervised Learning for Thermophysical Analysis on the Lunar Surface

Ben Moseley^{1,2} , Valentin Bickel^{1,3,4} , Jérôme Burelbach^{1,5} , and Nicole Relatores^{1,6}

¹ NASA Frontier Development Lab, Mountain View, CA, USA; bmoseley@robots.ox.ac.uk

² Department of Computer Science, University of Oxford, Oxford, UK

³ Department of Earth Science, ETH Zurich, Switzerland

⁴ Department Planets and Comets, Max Planck Institute for Solar System Research, 37077 Göttingen, Germany

⁵ Physics and Material Science Research Unit, University of Luxembourg, Luxembourg City, Luxembourg

⁶ Observatories of the Carnegie Institution for Science, Pasadena, CA, USA

Received 2020 May 8; revised 2020 June 5; accepted 2020 June 5; published 2020 July 28

Abstract

We investigate the use of unsupervised machine learning to understand and extract valuable information from thermal measurements of the lunar surface. We train a variational autoencoder (VAE) to reconstruct observed variations in lunar surface temperature from over 9 yr of Diviner Lunar Radiometer Experiment data and in doing so learn a fully data-driven thermophysical model of the lunar surface. The VAE defines a probabilistic latent model that assumes the observed surface temperature variations can be described by a small set of independent latent variables and uses a deep convolutional neural network to infer these latent variables and to reconstruct surface temperature variations from them. We find it is able to disentangle five different thermophysical processes from the data, including (1) the solar thermal onset delay caused by slope aspect, (2) effective albedo, (3) surface thermal conductivity, (4) topography and cumulative illumination, and (5) extreme thermal anomalies. Compared to traditional physics-based modeling and inversion, our method is extremely efficient, requiring orders of magnitude less computational power to invert for underlying model parameters. Furthermore our method is physics-agnostic and could therefore be applied to other space exploration data sets, immediately after the data is collected and without needing to wait for physical models to be developed. We compare our approach to traditional physics-based thermophysical inversion and generate new, VAE-derived global thermal anomaly maps. Our method demonstrates the potential of artificial intelligence-driven techniques to complement existing physical models as well as for accelerating lunar and space exploration in general.

Unified Astronomy Thesaurus concepts: [The Moon \(1692\)](#); [Lunar composition \(948\)](#); [Convolutional neural networks \(1938\)](#)

1. Introduction

Understanding the lunar surface is essential for scientific progress and for the development of the human race. The Moon's surface is the result of a turbulent geologic history and has been shaped by endogenic and exogenic processes over billions of years. It represents a record of the evolution of our Sun and the solar system (Spudis 1996; Fagents et al. 2010; Joy et al. 2016), and the spatial distribution and heterogeneity of surface characteristics, such as composition and relief, can be used to trace back or infer some of the fundamental processes that control the evolution of the Earth–Moon system. Famously, the frequency of occurrence and diameters of impact craters can be used to estimate the age of lunar regions as well as of planetary surfaces all across the solar system (e.g., Shoemaker 1962; Baldwin 1964; Hartmann 1965). Besides informing science, knowledge about the characteristics of the lunar surface is essential for the continued exploration of the Moon. For example, maps of the temperature distribution and variation over the course of a lunar day (e.g., Vasavada et al. 2012; Williams et al. 2017) and maps of rock abundance (Bandfield et al. 2011) are vital for planning future crewed and uncrewed landings and missions. With NASA's goal to establish a permanent and sustainable human presence on the

lunar surface (Neal et al. 2009), lunar surface data could help locate resources for in situ resource utilization such as oxygen; water; rare Earth elements; metals like Fe, Al, Ti, Pt, and 3He; and solar power (ESA 2019).

The creation of scientific products that are required for these types of applications can be challenging, however. Many of these products are derived from raw physical data recorded by scientific instruments on board lunar satellites and require a physical model to interpret the data and to invert for underlying physical parameters. These models are typically either theoretically derived or have been developed during extensive laboratory studies (e.g., Bandfield et al. 2011; Hayne et al. 2017). However, such approaches are often biased toward the assumptions of the physical model, require computationally demanding algorithms, and are driven by time-consuming and tedious human work. With space exploration missions returning increasing amounts of data, new and more efficient means of big data exploration and analysis have to be developed and adopted.

The field of artificial intelligence (AI) has seen explosive growth in the last decade. From being the subject of research, it is now proving invaluable across many real-world problems. It is arguably recent progress in deep learning that has lead to this growth, with breakthroughs in the ability to train deeper neural networks, the availability of bigger data sets, and improvements in computing hardware moving the field forward significantly (Goodfellow et al. 2016). Deep learning algorithms now achieve state-of-the-art results in many tasks, such as computer vision, speech recognition, natural language



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

processing, and robotics (Hinton et al. 2012; Russakovsky et al. 2015; Pierson & Gashler 2017; Vaswani et al. 2017). Modern deep learning techniques are pushing the forefront of research in physics too. For example, they have been successfully used to simulate particle showers in particle colliders (Paganini et al. 2018), optimize the control of quantum dynamical systems (Dalgard et al. 2020), and iteratively reconstruct undistorted images of gravitationally lensed objects (Morningstar et al. 2019).

In lunar and planetary science, however, AI and deep learning have seen limited use to date. For example, Silbert et al. (2019) developed a deep learning-driven tool to detect and measure impact craters on the Moon and Mercury in digital elevation model data. Kodikara & McHenry (2020) showed that machine learning can be used to classify the physical and mineralogical properties of lunar regolith. Bickel et al. (2019, 2020) trained convolutional neural networks (CNNs) to automatically extract the locations and sizes of lunar rockfalls from Lunar Reconnaissance Orbiter's (LRO's) image archive with more than 2 million entries, and Wu et al. (2019) presented a neural network-driven approach to improve the localization accuracy of rovers on planetary surfaces by matching ground-based and space-borne imagery. While these works concentrate on detection and classification, little work has been done on using AI to learn about and to model underlying physical processes on the Moon.

In this work, we investigate the use of unsupervised machine learning to extract physical models from space exploration data. More specifically, we train a variational autoencoder (VAE; Kingma & Welling 2014; Rezende et al. 2014; Doersch 2016) to model over 9 yr worth of surface temperature data from NASA's LRO Diviner Lunar Radiometer Experiment (hereafter Diviner) instrument. The VAE defines a fully data-driven, probabilistic model of the data, and we find it is able to identify underlying, physically interpretable factors of variation in the surface temperature, which strengthens our confidence in existing thermophysical models. Furthermore, it is able to infer these factors of variation orders of magnitude faster than traditional thermophysical inversion. We use the learned representation of the VAE to efficiently generate new, AI-derived global maps, which characterize thermal behavior and highlight thermal anomalies on the lunar surface. Our method is general in that it is physics-agnostic and could therefore be used to extract physical models from other space exploration data sets without the need for an *a priori* physics model. We believe our work demonstrates the potential of big data-driven machine-learning methods for aiding and accelerating lunar science and exploration.

We present our methods in Section 2, results in Section 3, and discussion in Section 4. In more detail, Section 2.1 describes our curation of the Diviner data set, Section 2.2 describes deep learning and VAEs, and Sections 2.3–2.5 describe our machine-learning workflow. A comparison of our approach to an existing thermophysical model is presented in Section 2.6, and we generate global AI-derived thermal anomaly maps in Section 2.7. Sections 3.1–3.3 present the results from all the sections above. Section 4 discusses the interpretation and limitations of our approach and the potential for applying it to other areas of space exploration.

2. Methods

2.1. Diviner Instrument Overview and Data Preprocessing

The Diviner instrument is a passive Vis-infrared (IR) radiometer on board NASA's LRO that has been collecting radiance measurements in nine different wavelength channels (0.3–200 μm) from its near-polar orbit since 2009 (Paige et al. 2010; Williams et al. 2017; Mazarico et al. 2018). Channels 6 through 9, located in the 12.5–200 μm range (middle to far-IR), are particularly sensitive to surface day and night temperatures. Each channel is defined by a linear array of 21 thermopile sensors (i.e., the instrument has a total of 189 sensors) and due to the limited field of view (FOV) of the sensors, Diviner's swath width is very narrow, resulting in quasi-point measurements along LRO's ground track (nadir push broom mapping). Thus, multiple orbits are required to completely cover the entire surface of the Moon and the number and effective FOV of the available measurements governs the spatial and temporal resolution of Diviner's final image or map products. The dominant influence on the size of the effective FOV of each point measurement is the spacecraft altitude, which has varied between \sim 20 and 200 km over Diviner's operational period; at 30 km altitude, it is expected to be an ellipse approximately 160 \times 120 m in size with its major axis oriented in the in-track direction, while at 200 km altitude, these dimensions are expected to double or triple (Sefton-Nash et al. 2017). Two other effects on the FOV include in-track smearing due to LRO's motion and footprint elongation caused by sensor memory (Paige et al. 2010; Williams et al. 2016). The in-track smearing can result in slight inaccuracies in the location of maximum probability of the source of back-scattered photons, but its significance weakens with increasing spacecraft altitude.

There are multiple Diviner data sets at different levels of processing publicly available from NASA's Planetary Data System. For this study, we collect the Diviner Level 1 Reduced Data Record (RDR) data (Paige et al. 2011), which consists of tables of calibrated radiance point measurements and their derived surface temperature values, as well as their associated ephemeris and geometry information, including the local lunar time at which each point measurement was recorded. We choose this data set because it is minimally processed from the raw Diviner telemetry data and contains the point measurements before any gridding is applied.

We download data from Diviner's entire operational period 2010 January–2019 March, which contains data recorded during LRO's nominal science mission phase as well as its three extended science mission phases. We use all available data for a number of reasons. First, because we are using a data-driven approach, we want to absorb as much data as possible and do not wish to bias our analysis to a narrow observational window. Second, as noted above, including more orbits increases the spatial-temporal resolution of the data. An important consideration is that LRO changed from a near-circular to an elliptical orbit in 2011 December, with its periapsis oriented over the lunar south pole. Thus, the effective FOV of point measurements is expected to vary with latitude after this date (Sefton-Nash et al. 2017).

A number of preprocessing steps are carried out before inserting this data into our unsupervised learning workflow. First, for each RDR data table, we only keep measurements that have instrument activity flags equal to 110 (indicating "on

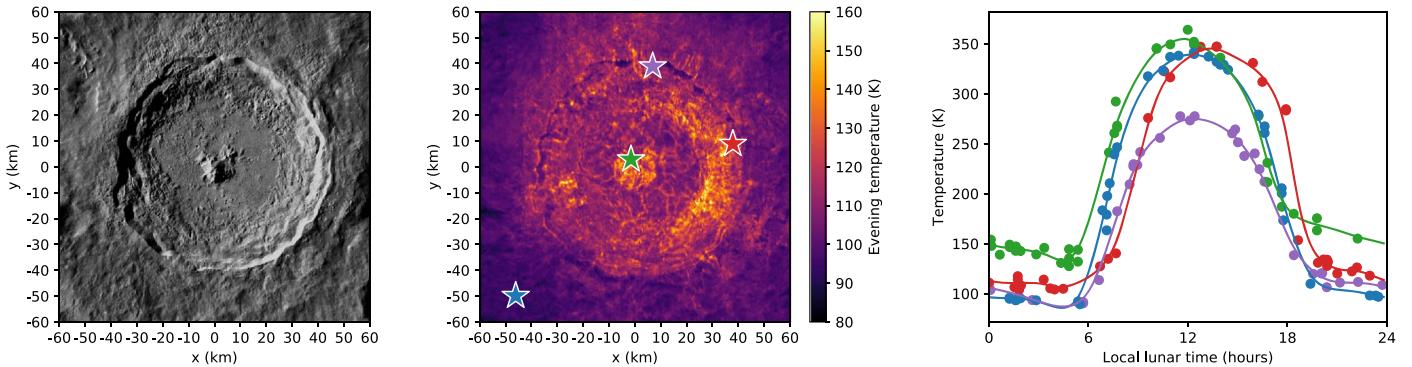


Figure 1. Left: optical image of Tycho (LROC Wide Angle Camera (WAC) Global Morphologic base map; Speyerer et al. 2011; 348.7° lon./−43.3° lat.). Middle: evening temperature map extracted from the preprocessed Diviner data set, plotting the average temperature over 12 am–4 am in local lunar time. Right: four example temperature profiles extracted over Tycho. Points show temperature measurements, and solid lines show the GP fits of the profiles used for interpolation. Color-coded stars in the middle plot indicate the locations of the profiles.

“moon” orientation, standard nadir observation, and a “nominal” instrument mode), calibration flags equal to 0 (indicating an in-bounds interpolated measurement), geometry flags equal to 12 (indicating tracking data was used to generate geometry information), miscellaneous flags equal to 0 (indicating no miscellaneous observations), and emission angles less than 10° (the angle between the vector from the surface FOV center to Diviner and the “normal” vector to the Moon’s surface). Second, for simplicity, in this study we only use data from channel 7, which records radiance in the 25–50 μm band, is the most sensitive to temperature values in the range 69–178 K, and has a high single-to-noise ratio over the majority of surface temperatures (Vasavada et al. 2012; Williams et al. 2017; Feng et al. 2020). Finally, we sort the data into $0.5^\circ \times 0.5^\circ$ latitude and longitude bins, which allows temperature values at each point in space to be easily extracted in our subsequent workflows. This preprocessing step takes considerable computational resources; 400 CPU cores are used to process over 425 billion point measurements contained in 35 TB of uncompressed Diviner RDR data. Example preprocessed data are shown in Figure 1.

2.2. Deep Learning, CNNs, and VAEs

Deep learning is a subfield of machine learning, which itself a subfield of AI, and offers a class of machine-learning algorithms that use neural networks with many layers to learn complex, nonlinear relationships in data. A quintessential example is the deep feed-forward network (Goodfellow et al. 2016). Given a set of training samples $\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ from a function $y = f^*(x)$: $X \rightarrow Y$, where X and Y are vector spaces, the deep feed-forward network defines a mapping $y = f(x; \theta)$ and attempts to learn the values of the parameters θ that best approximate f^* . The mapping is typically composed of multiple functions, or layers, of the form $f(x) = f^{(d)} \circ \dots \circ f^{(3)} \circ f^{(2)} \circ f^{(1)}(x)$, where the output dimensionality (or “width”) of each function and the number of functions (or “depth”) can vary. One standard form is the fully connected network, where each layer function has the form $f^{(i)}(x; W^{(i)}, b^{(i)}) = g(W^{(i)\top} x + b^{(i)})$, where the weight matrix $W^{(i)}$ and bias vector $b^{(i)}$ contain learnable parameters and g is an activation function, often chosen to be the rectified linear unit, given by $g(x) = \max\{0, x\}$. Rather remarkably, the universal approximation theorem (Hornik et al. 1989) states that a fully connected feed-forward network with 2 or more layers can approximate any continuous function for inputs within a specific

range. It is this flexibility that potentially allows deep neural networks to learn across many different problems.

CNNs are a specific type of deep neural network that have become hugely popular in recent years. In a CNN, each component in a layer’s output vector (or “neuron”) shares the same weight values as the other output components. Furthermore, each neuron has a limited FOV of the layer’s input vector. Mathematically, the layer function is defined as $f^{(i)}(x; w^{(i)}, b^{(i)}) = g(w^{(i)} \star x + b^{(i)})$, where \star denotes cross-correlation and $w^{(i)}$ is a vector of a fixed size, or filter, which does not depend on the size of the layer’s input, and $b^{(i)}$ is a scalar. CNNs are nearly always extended to allow multiple filters (or “channels”) per layer and 2D and 3D cross-correlation over its inputs. An important feature of CNNs is that they are equivariant to translations of their inputs, which is extremely valuable when learning about data that contains spatial correlations, such as images. CNNs have been successfully applied in computer vision, where they achieve state-of-the-art performance (Russakovsky et al. 2015), as well as many other areas.

VAEs use deep neural networks to carry out unsupervised learning (Kingma & Welling 2014; Rezende et al. 2014; Doersch 2016). Rather than estimating a function, they use deep neural networks to carry out approximate inference in a latent probabilistic model. The goal is to (implicitly) learn a distribution $p(x)$, $x \in X$, where X is a vector space, which is as similar as possible to an unknown target distribution $p^*(x)$, given a set of samples from this distribution $\mathcal{D} = \{x_1, \dots, x_n\}$. VAEs make the assumption that x depends on some set of underlying latent variables $z \in Z$, which are distributed according to some known distribution $p(z)$, and use deep neural networks to model the likelihood $p(x|z)$ and an approximation of the posterior distribution $p(z|x)$. The latent representation is learned by training the VAE to reconstruct the training data. One typically chooses the dimensionality of the latent representation to be much lower than the dimensionality of X , forcing the VAE to learn a salient representation of the input data. Furthermore, in a standard VAE, these latent variables are assumed to be independent and normally distributed, which has the effect of keeping input samples that are similar in X close together in the latent space Z and is very useful when one wants to learn “meaningful” representations that vary smoothly over the input space. In computer science, VAEs have successfully been applied to disentangle hair styles, poses, and facial expressions from images of people’s faces

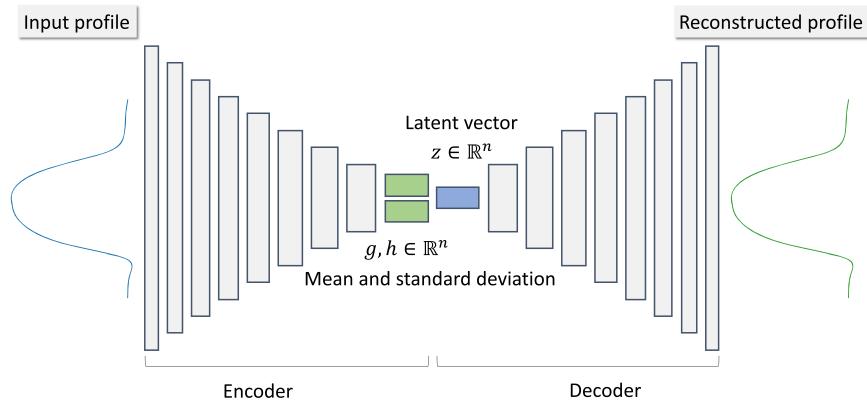


Figure 2. VAE workflow. We train a VAE to compress the observed surface temperature variations over the lunar day into a sparse set of latent values. The input to the VAE is the 1D profile of the surface temperature variation over the lunar day extracted at a single point on the lunar surface. A convolutional encoder computes the mean and standard deviation of the VAE’s approximate latent posterior distribution, given these observations. During training a sample from this posterior distribution is feed into a convolutional decoder, which outputs a reconstruction of the input profile.

(Larsen et al. 2015; Higgins et al. 2017). By using deep neural networks to infer $p(x|z)$ and approximate $p(z|x)$, VAEs are able to learn highly nonlinear mappings between the input and latent space. A detailed mathematical description of VAEs is included in Appendix A.

2.3. Overview of VAE Workflow

Our basic idea is to use a VAE to learn the underlying factors of variation in the Diviner surface temperature measurements. The input to the VAE is the observed surface temperature variations over the lunar day at each point on the surface, and we learn a latent representation that is able to reconstruct this input. In doing so, we learn a model of the thermophysical data, without the use of an a priori physics model. We choose a VAE framework (as specified in Equation (A3)–(A6)) because it attempts to learn a model that (1) accurately reconstructs the input profiles, (2) disentangles independent factors of variation in the surface temperature, and (3) has a latent representation, which varies smoothly over the input space. We stress that these are very general assertions and therefore this method could be applied to other similar space science and exploration tasks.

A major benefit of this approach is that it is fully data driven. In general, methods that rely on physical models to analyze physical data may be unreliable if one is not confident on the accuracy of the model (e.g., Hayne et al. 2017). Furthermore, we learn salient features directly from the data; other traditional approaches that uses handcrafted features (such as averaged or evening temperature maps; e.g., Bandfield et al. 2011) to interpret physical data risk discarding important information. We interpret the VAE by comparing its latent variables to parameters estimated using inversion with an existing thermophysical model. Once trained, the VAE is run over the entire lunar surface to map its latent variables and these maps are used to search for thermal anomalies. We name the VAE “Lunar ORbiter-Derived VAE,” or LordVAEder.

Our VAE workflow is shown in more detail in Figure 2. The input to the VAE is the 1D profile of the surface temperature variation over the lunar day extracted at a single point on the lunar surface. The VAE has multiple “encoder” convolutional layers that reduce this profile into its latent representation and multiple “decoder” convolutional layers that reconstruct it, and

it is trained using example profiles from many different locations on the lunar surface.

2.4. Training Data Generation

To train the VAE, millions of example 1D profiles of the surface temperature variation over the lunar day are extracted from 48 areas of interest (AOI). We curate the AOIs so that they encapsulate a wide range of surface temperature variation. The AOIs include impact craters of different ages, large pyroclastic deposits, and swirls (magnetic anomalies), with areas ranging from 1 to 10,000 km². They are listed in Table B1 and described in more detail in Appendix B. The locations of all the AOIs are shown in Figure B1.

We extract the 1D temperature profiles by locally projecting and binning the observed Diviner surface temperature measurements at each AOI onto a 200 × 200 m grid and sorting the points in each bin by the local lunar time at which they were recorded, obtaining a profile at each bin location. An orthographic projection centered over each AOI with a radius of $R_{\text{moon}} = 1737.4 \times 10^3$ m is used to define each grid. To ensure sufficient sampling of each temperature profile, we reject profiles whose maximum spacing in local lunar time between points is less than 4 hr. We also reject some profiles from background areas to ensure they do not dominate the training set. This workflow results in 1,985,693 training profiles extracted over all AOIs and examples of extracted profiles are shown in Figure 1.

Before inputting the profiles into our VAE workflow, we interpolate their temperature measurements onto a regular grid by using Gaussian process (GP) regression (Rasmussen & Williams 2005). A GP with a Matern-1.5 kernel, a maximum length scale of 6 hr, and an assumption of 10 K standard deviation of uncertainty in each temperature measurement is fit to each profile. We sample every 0.2 hr, resulting in a 1D input profile with 120 samples. Example GP interpolations are also shown for the profiles in Figure 1.

2.5. VAE Architecture and Training

A convolutional encoder-decoder design is used for our VAE architecture, shown in Figure 2, which consists of nine convolutional layers in the encoder and eight convolutional layers in the decoder. The encoder takes an input profile of dimensionality 120 × 1 and slowly reduces it to two output

vectors each of dimensionality $1 \times n$, which represent the mean, g , and standard deviation, h , of the approximate latent posterior distribution in Equation (A5). The number of latent variables, n , is a hyperparameter of the VAE model we test. The decoder takes a sample from the latent posterior distribution and slowly expands it to output a reconstruction of the input profile.

We use 32 hidden channels in the encoder and decoder and batch normalization (Ioffe & Szegedy 2015) after all hidden layers to help training converge. For $n = 4$, the total number of trainable parameters in the model is 28,713. A complete specification of the network is given in Table B2 and the network architecture is described in more detail in Appendix B. We design the network architecture by hand to maximize its reconstruction performance.

During training, the mean and standard deviation vectors are used to generate a single sample (or vector of latent values) from the approximate posterior distribution, which is fed into the decoder so that the expectation term in the loss function (Equation (A6)) can be estimated. During inference, only the mean vector is passed to the decoder and used to define the latent values. We use a β -value of 0.2 and normalize the training profiles to have a collective mean of 0 and a standard deviation of 1 before inputting them to the VAE. The network is trained using the Adam stochastic gradient descent optimization algorithm (Kingma & Ba 2014), using an exponentially decaying learning rate starting at 1×10^{-3} and a mini-batch size of 200 profiles, reserving 20% of the example data set for cross-validation. We use the PyTorch framework (Paszke et al. 2018) to define and train the VAE.

2.6. Physics Modeling and Interpretation

A downside of our approach is that the VAE model lacks a clear physical interpretation. To aid its interpretation, we compare the VAE's latent representation to a physics-based inversion. We modify the thermophysical model proposed by Hayne et al. (2017), which consists of numerically solving the heat equation over time t and depth z in a 1D solid medium:

$$\rho(z, t) C_p(z, t) \frac{\partial T(z, t)}{\partial t} = \frac{\partial}{\partial z} \left(K(z, t) \frac{\partial T(z, t)}{\partial z} \right), \quad (1)$$

where $\rho(z, t)$ is the mass density, $C_p(z, t)$ is the specific heat capacity, $T(z, t)$ is the temperature, and $K(z, t)$ is the thermal conductivity. The boundary conditions and functions $\rho(z, t)$, $C_p(z, t)$, and $K(z, t)$ must be known in order to solve this differential equation for the temperature profile. We make a few important alterations to the model used by Hayne et al. (2017). First, we assume that the density ρ is constant and equal to the surface density ρ_s given by Hayne et al. (2017). Second, we use the phonon conductivity K_c as a free fitting parameter, rescaled with respect to the surface conductivity K_s given by Hayne et al. (2017). For this purpose, we introduce the thermal conductivity factor as K_c/K_s . Third, we assume that the solar incidence angle approximately coincides with the hour angle and absorb the effect of the latitude and slope on solar incidence in an effective Bond albedo, which we use as another free fitting parameter. Hence, our model treats the lunar regolith as a solid medium with a constant density but with a varying effective albedo and thermal conductivity factor.

We run inversion using this model by tuning the thermal conductivity factor, albedo, and onset time of solar radiation to best match each input interpolated profile. Bayesian optimization with an expected improvement acquisition function and a L2 loss function is used as the inversion algorithm (Frazier 2018). We then compare the estimated parameters from the inversion to the latent variables generated by the VAE.

2.7. Generation of Global Maps

As a final step, we run the trained VAE with $n = 4$ over the entire lunar surface and generate maps of its latent variables to search for thermal anomalies. This allows us to compute global statistics of the learned latent variables and interpret these new features alongside existing physical maps.

The first step in our mapping workflow is to run VAE inference and to obtain latent values over the Moon's entire sphere. To do this, we tessellate the sphere into many latitude-longitude boxes of approximately equal areas. For each box, we bin the observed Diviner surface temperature measurements onto a 200×200 m grid, using an orthographic projection centered at the center of each box, obtaining a profile at each bin location. Each profile is interpolated using GP interpolation, rejecting profiles whose maximum spacing in local lunar time between points is less than 4 hr. The VAE is run on each interpolated profile, obtaining a set of latent values at each bin location, and the central coordinate of each bin is unprojected to obtain a set of latent value latitude-longitude points for each box.

The second step is to project and interpolate the latent value latitude-longitude points onto a global map. We use the standard plate-carrée map projection (equirectangular) and linearly interpolate the points onto a 100×100 m grid. Similar to the first step, we do this by tessellating the sphere into many latitude-longitude boxes of approximately equal area. For each box, we project the latent value latitude-longitude points onto the global map and use linear interpolation with Delaunay triangulation to interpolate the points onto the global grid. Interpolated points are set to null values where their Delaunay triangle has a maximum distance between its vertices on the surface of the Moon's sphere greater than 2000 m to ensure that the map does not include over-interpolated points.

It is important to note that geometrical error in our workflow is introduced in the first step when binning temperature measurements onto the 200×200 m grid, because the bins in the orthographic projection are not exactly equally sized when unprojected onto the surface of the sphere. However, we choose small enough box sizes when tessellating the sphere so that this effect is negligible; 379,448 boxes of approximately 10×10 km shape on the Moon's surface are used for the first step and 10,322 boxes of approximately 61×61 km shape are used for the second step. For both steps, we ensure the borders of the boxes are sufficiently overlapping during the processing of each box to ensure that no edge effects are present. Part of the tessellation used for the second step is plotted in Figure C1. Both steps require considerable computational resource, but tessellation allows the data to be processed in parallel using 100 CPU cores. Our final output products are four latent maps and one quality control map that shows the L1 reconstruction loss of the VAE. The maps are cropped between -70° and 70° latitude to avoid excessive deformation and areas of high VAE

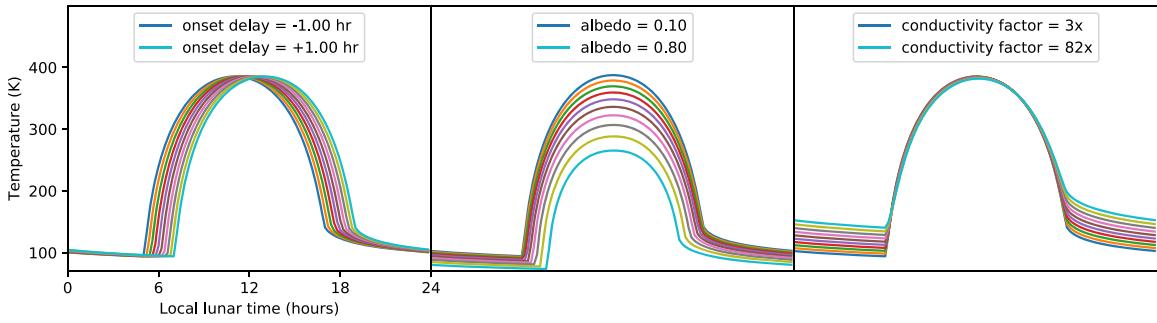


Figure 3. 1D numerical simulations of the lunar surface temperature against local lunar time. Our baseline model uses thermal properties estimated for the lunar regolith. The three plots show the sensitivity of the baseline model when varying the onset delay, albedo, and thermal conductivity factor independently.

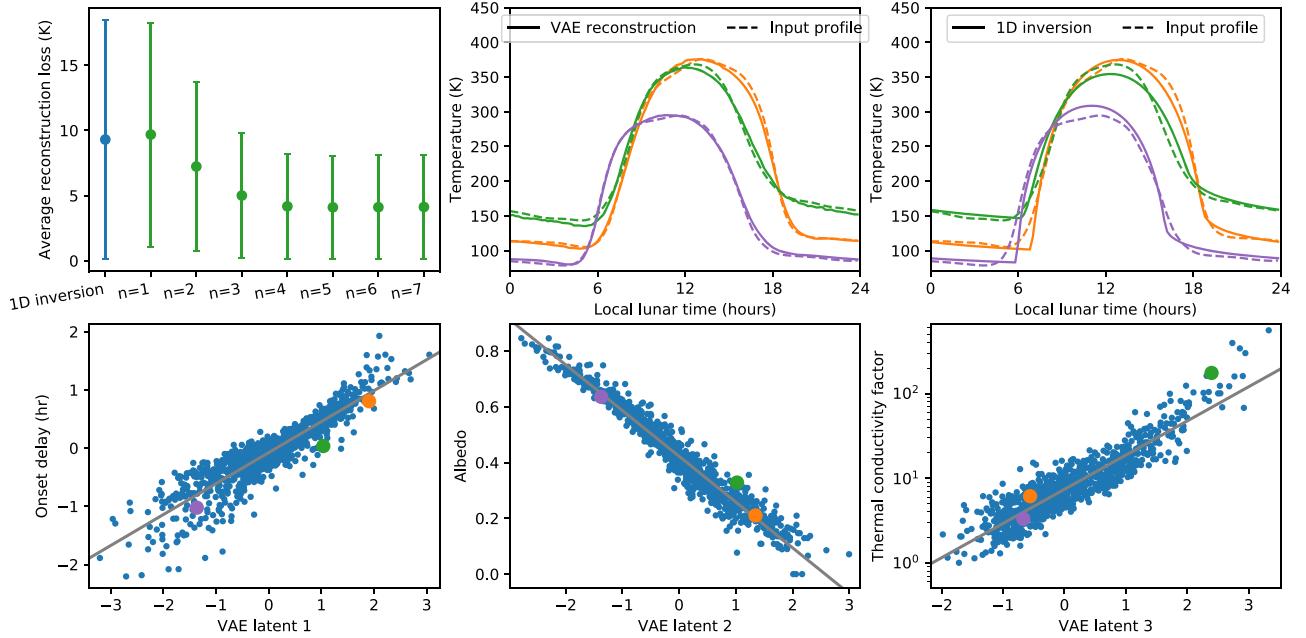


Figure 4. Top left: average L1 reconstruction loss of VAEs with varying latent vector lengths n and the 1D physics inversion, over 1000 profiles from the validation set. Error bars show ± 1 standard deviation. Top middle and right: example reconstructions for the VAE with $n = 4$ and the 1D physics inversion, for three selected profiles in the validation set. Bottom: plots showing the correlation between the first three VAE latent values when $n = 4$ and the estimated parameters from physics inversion, over 1000 profiles from the validation set. Thermal conductivity factor is relative to the regolith. The larger colored-coded points correspond to the profiles shown in the top middle and top right plots.

reconstruction loss observed at the poles (see Appendix C and Figure C2 for more details).

3. Results

3.1. Data Preprocessing

In total 425,745,195,120 Diviner point measurements are extracted, of which 326,190,541,311 (77%) remain after the first preprocessing filtering step, and 36,483,372,924 remain after filtering channel 7, requiring 2.9 TB of storage. We plot the number of points after preprocessing in each latitude-longitude bin in Figure D1. There is a higher point density closer to the poles, which is expected because LRO's orbit means the Diviner instrument samples the poles more frequently than the equator.

Example data after preprocessing over the Tycho crater are plotted in Figure 1. We observe a wide variation in temperature around this crater. Outside of the crater, the surface temperature increases to around 350 K at midday and decreases to 100 K at night. On the eastern slope of the crater the profiles are shifted in local time by approximately 2 lunar hours, which is

consistent with the slope's shadow delaying the incidence time of the Sun's radiation. On the northern slope the peak temperature is lower at 270 K, which is consistent with the slope receiving less thermal radiation per unit area from the Sun's elevation at Tycho's latitude. In the center of the crater we observe a profile which has a higher evening temperature around 150 K. Figure 3 (right) plots our physics model simulation when varying thermal conductivity factor; one explanation is that the material at Tycho's center has a higher thermal conductivity than its surroundings, but it should be noted that its central peak has complex topography and surface roughness that are likely to influence the measured temperature profiles.

3.2. VAE Results and Physical Interpretation

While training the VAE, the training and validation reconstruction loss values converge to similar values, suggesting that the VAE generalizes well and does not over-fit to the training data set. Training takes approximately 1 hr using one Nvidia Tesla T4 GPU. Example reconstructed profiles from the

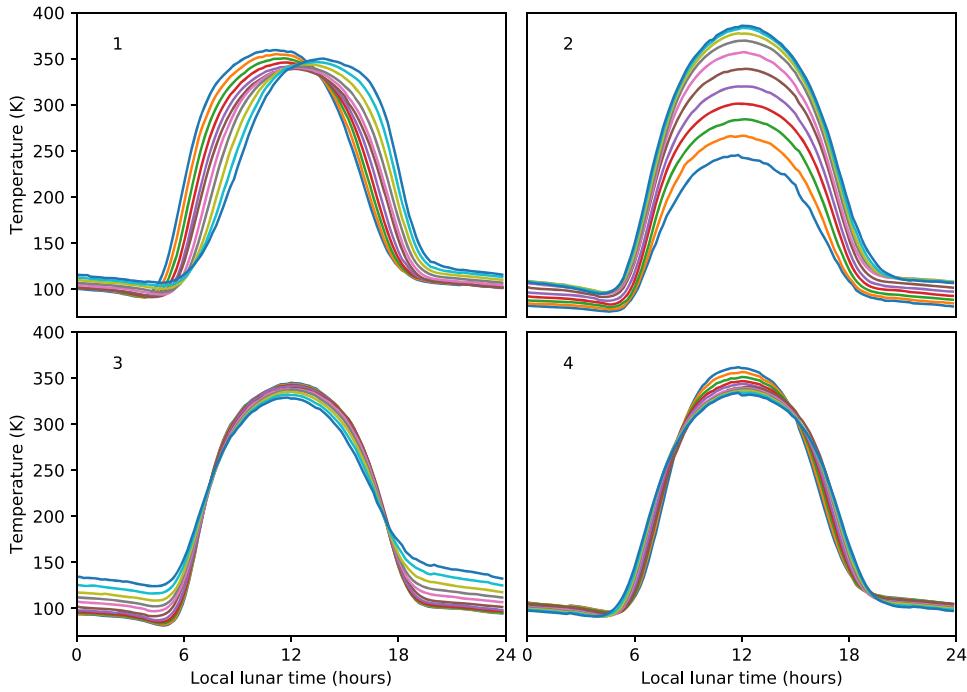


Figure 5. Reconstructed profiles generated from the VAE with $n = 4$ when sampling each latent variable independently across its range and fixing the other latent variables to zero.

validation data set after training are plotted in Figure 4 (top middle panel). We find that the VAE is able to accurately reconstruct the profiles from the validation data set. Figure 4 (top left panel) shows the average reconstruction loss over the validation data set after training the VAE when varying the number of latent variables. The reconstruction loss asymptotes at $n = 4$, suggesting that only four latent variables are needed to accurately represent the input profiles.

Figure 5 shows example profiles generated by the VAE when varying each latent variable independently and fixing the others to zero for $n = 4$. We observe that latent variable 1 responds to the onset of the profile, latent 2 responds to the peak temperature, latent 3 responds to the evening temperature “sidelobe,” and latent 4 responds to the width of the temperature profile. To aid our interpretation of the VAE further, we compare its latent variables with estimated parameters from the physics inversion. The VAE and physics inversion are run over the validation data set and the latent values are plotted against the estimated physics parameters in Figure 4 (bottom). We find that latent variables 1, 2, and 3 have a strong linear correlation with the onset of solar radiation, effective albedo, and the logarithm of the thermal conductivity factor, respectively. This correlation can also be seen in Figure 3; scanning the physics simulation across the physical parameters results in similar profile variations to those seen when scanning the latent variables in Figure 5.

A straight line fit is shown for each correlation plot in Figure 4 (bottom panel). As thermal inertia is proportional to the square root of thermal conductivity, we expect that applying the transform $\hat{I} = e^{mc_3/2}$ to the latent 3 values will obtain a quantity that is linearly correlated with the thermal inertia at constant density and temperature, under the assumptions of our physics model. Here m is the gradient of the latent 3 and log thermal conductivity factor correlation and is estimated to be 0.93.

Maps of the latent variables and the VAE L1 reconstruction loss over Tycho are plotted in Figure 6. The VAE takes approximately 0.001 s to compute latent values for each profile when running on a single CPU, while each physics inversion takes of the order of 60 s to run using the same CPU. This significant efficiency improvement (factor of 60,000 or 4 orders of magnitude) allows the VAE to be run over large areas.

3.3. Global Maps

The global maps produced by our mapping workflow are shown in Figures 7 and 8. We find that our observations of the latent variables are consistent on a global scale; the latent 1 map consistently shows higher values on the eastern sides of craters and lower values on their western sides, indicating that latent 1 responds to the onset delay of solar radiation as a function of slope (east–west) aspect. The latent 2 map shows higher values at the equator and lower values closer to the poles, suggesting that it represents a combination of an effective incidence angle and surface albedo. The latent 3 map mostly shows a background value with some craters and recent impact basins being either anomalously bright (e.g., the Tycho crater) or showing a brightness level slightly above the background (e.g., Mare Orientale) and could plausibly indicate changes in thermal conductivity. The latent 4 map is generally darker within smaller and narrower craters, particularly toward the poles. Besides their dominant trends, both the latent 1 and 2 maps show minor, local variations in the equatorial regions (for example, around the main maria (centered at 315° lon./ 10° lat. or 45° lon./ 0° lat.)) that could represent the influence of terrain roughness on latent 1 and varying albedo on latent 2. We define an anomalously high value of the latent 3 variable to be one that is greater than two standard deviations from its global mean and find that approximately 4% of the surface area of the Moon (within -70° to 70° latitude) contains these values.

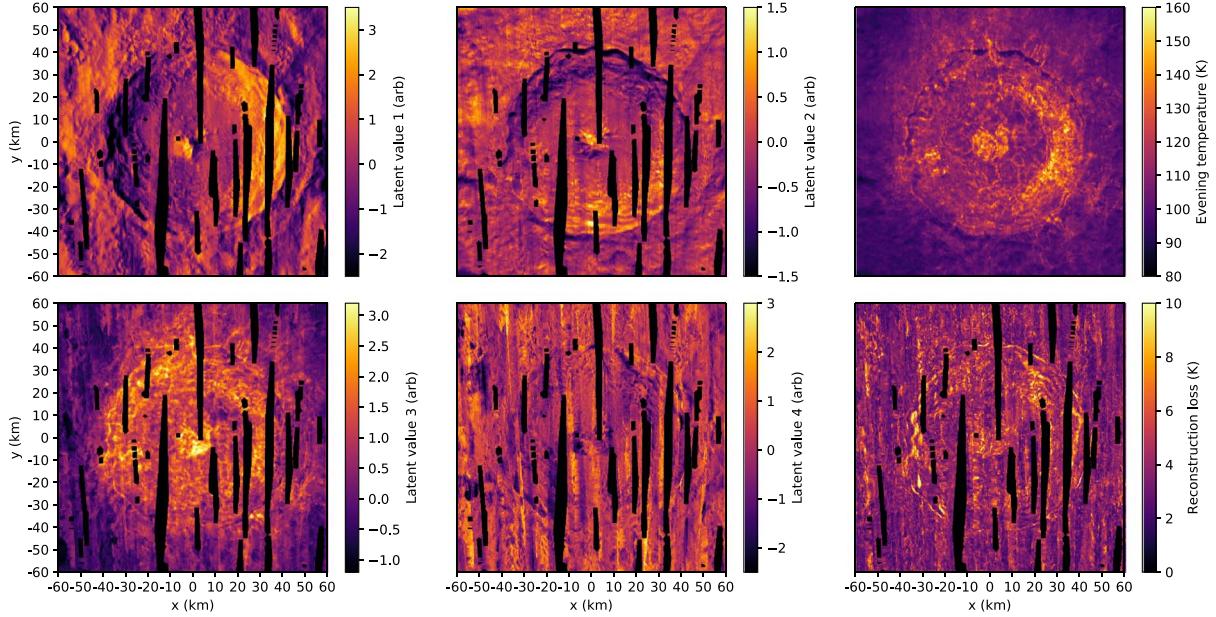


Figure 6. Maps of the VAE latent variables over Tycho, using the VAE with $n = 4$. The left two columns show maps of the four VAE latent variables, without any transforms applied. The top right panel is a repeat of the Diviner evening temperature map shown in Figure 1. The bottom right panel shows a map of the L1 reconstruction loss of the VAE.

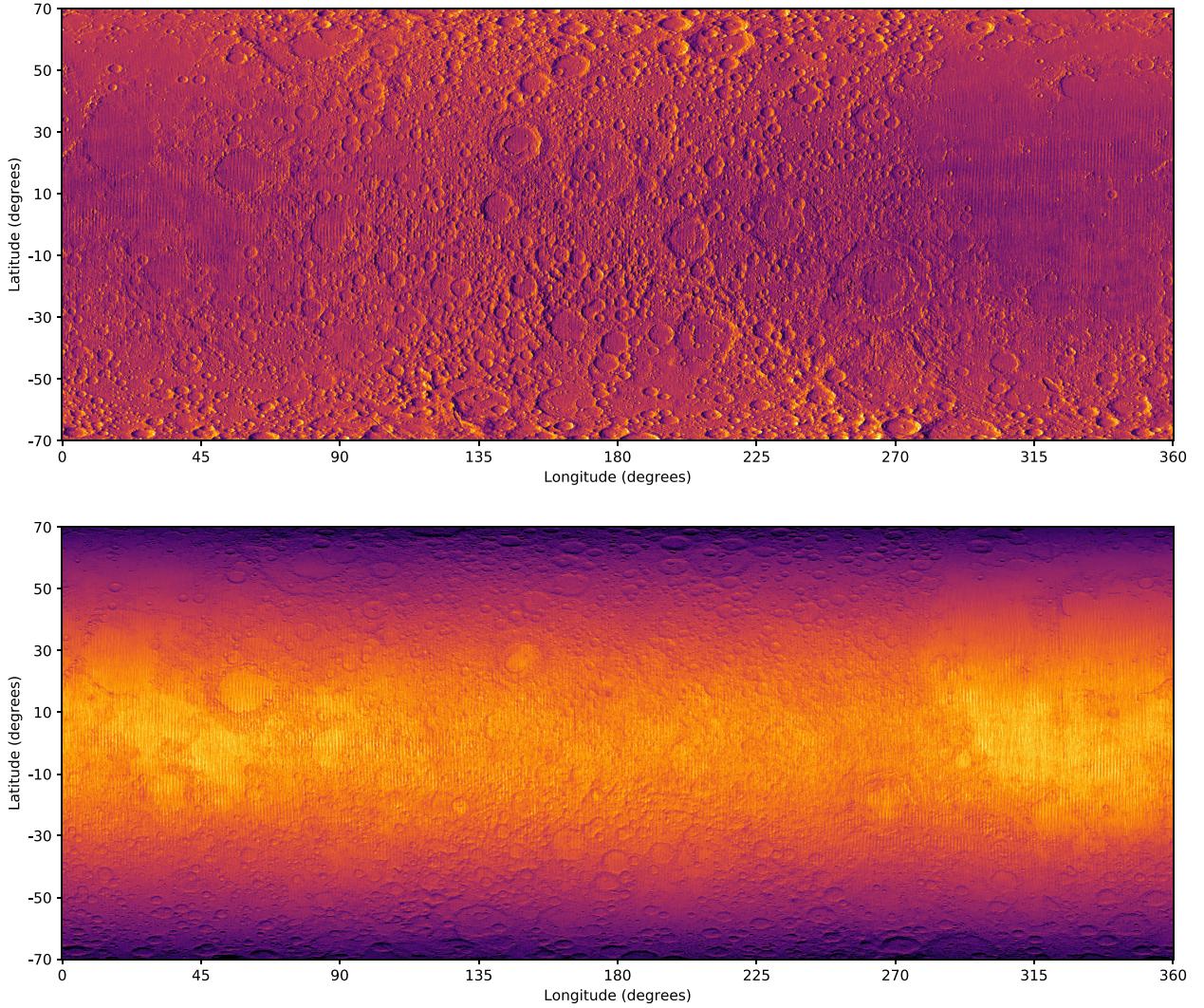


Figure 7. Global maps of the VAE latent variables 1 and 2. The top map shows latent variable 1 and the bottom shows latent variable 2.

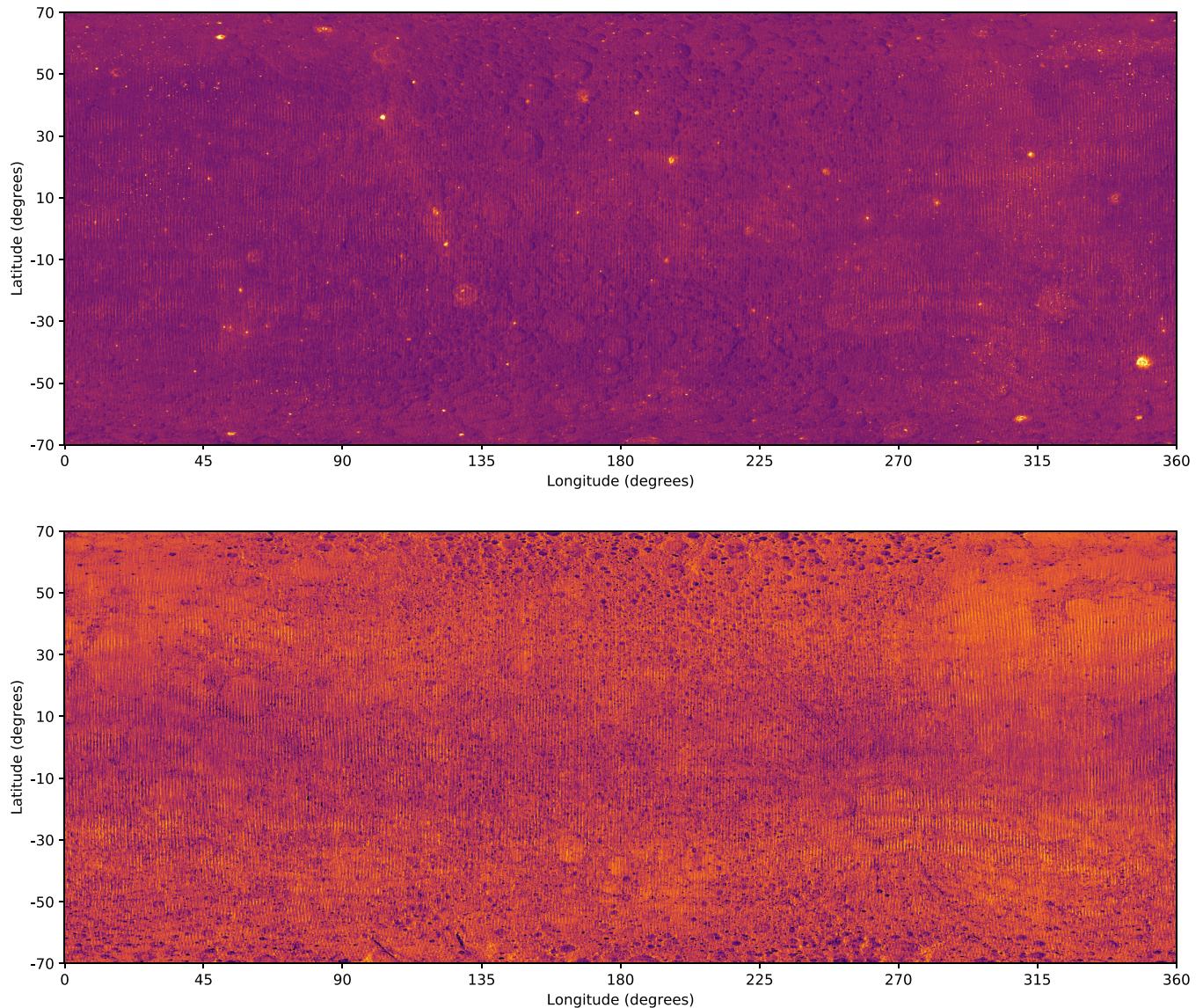


Figure 8. Global maps of the VAE latent variables 3 and 4. The top map shows latent variable 3 with the thermal inertia transform ($\hat{I} = e^{0.93z_3/2}$) applied and the bottom shows latent variable 4.

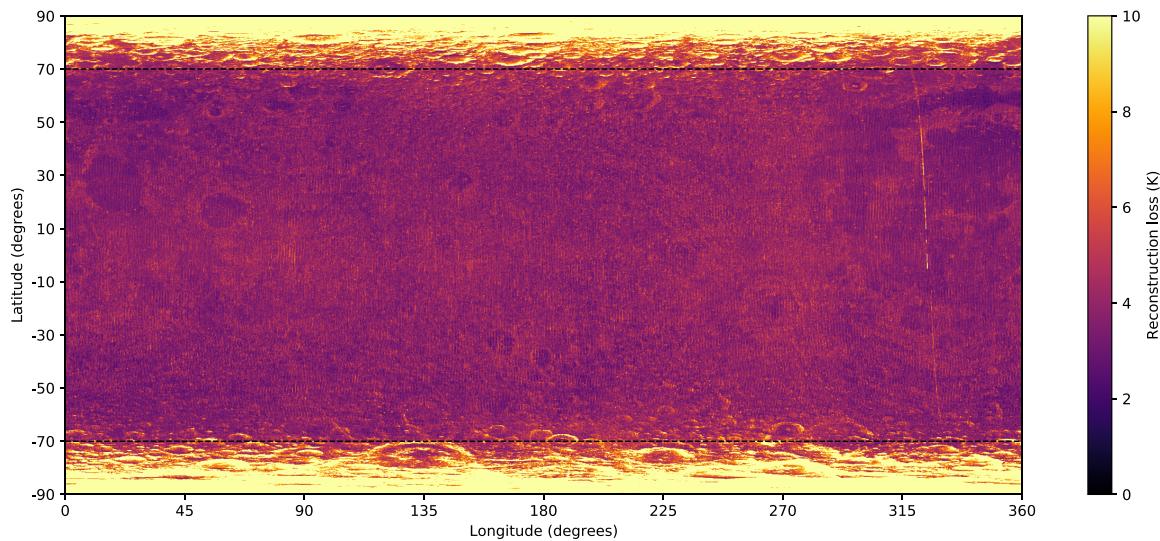


Figure 9. Global map of the VAE L1 reconstruction loss. Dashed lines show the -70° to 70° latitude cutoff used when generating the latent maps shown in Figures 7 and 8.

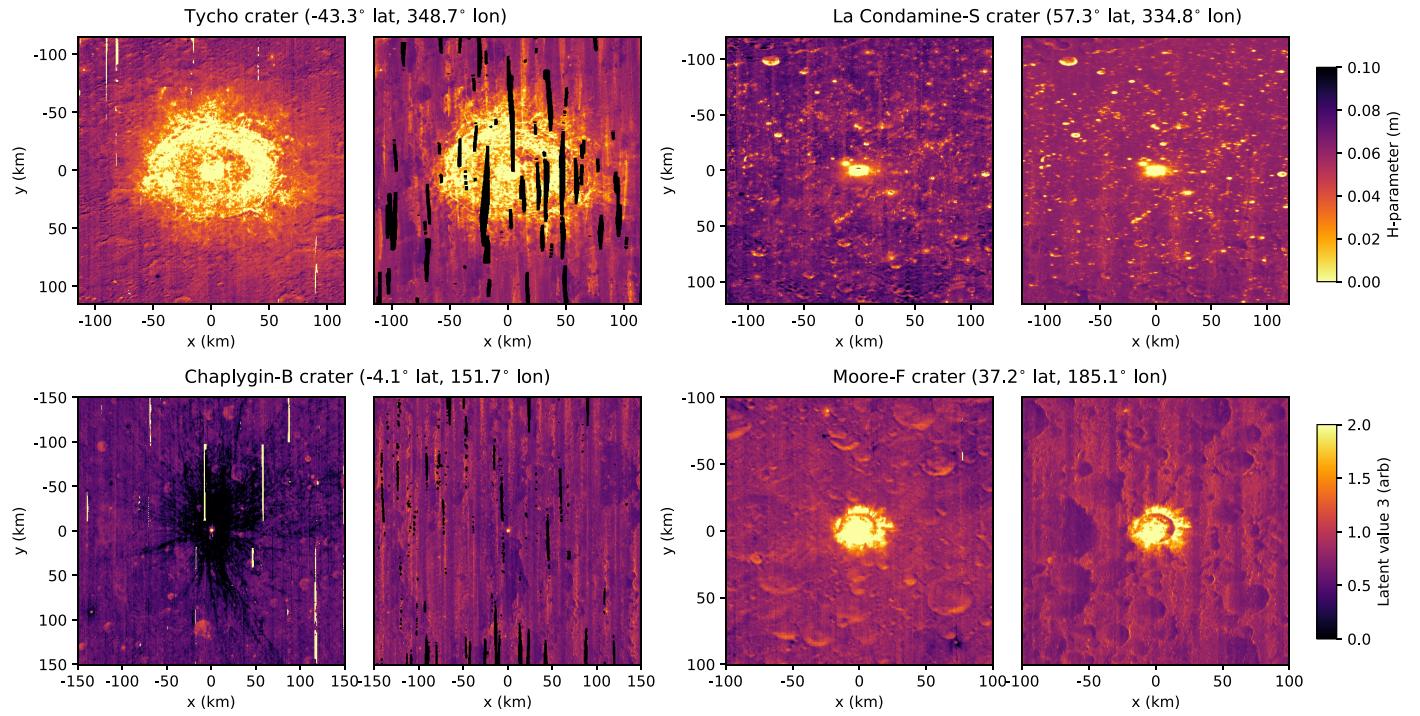


Figure 10. Comparison of the H-parameter map generated by Hayne et al. (2017) to our latent 3 variable map at four example crater locations. For each crater, the left shows the H-parameter map and the right shows the latent 3 variable map. The latent 3 variable map has the thermal inertia transform applied. Top left: high thermal inertia at the Tycho crater. Bottom left: the cold spot around Chaplygin-B crater unrecognized by latent 3. Top right: latent 3 slightly less affected by noise. Bottom right: H-parameter and latent 3 exhibit north/south and east/west coherent noise, respectively.

The global map of the VAE L1 reconstruction loss is shown in Figure 9. The loss is low (typically less than 10 K) in the range of -70° to 70° latitude, while it increases dramatically closer to the poles. In general, the loss slightly increases along topographic gradients such as the rims of craters and in the more heavily cratered highland regions. Locally, the loss increases at east–west-facing slopes of some impact features such as the Chaucer or Rydberg craters. A regional exemption is the north-western and central portion of Mare Smythii (85° lon./ 0° lat.), which show higher than average loss values. Some observations in the loss map are discussed in more detail in Appendix C.

Approximately 80% of the surface area of the Moon is covered by our extraction workflow. The missing data are where the extracted profiles do not meet our minimum temporal sampling criteria and are mostly located around the equator where the density of point measurements is lowest due to LRO’s near-polar orbit.

4. Discussion

We have shown that the latest unsupervised machine-learning techniques are able to extract physically interpretable models of thermophysical processes on the Moon. Our VAE is able to disentangle the effects of solar thermal onset delay caused by slope aspect, effective albedo, and surface thermal conductivity from the raw Diviner surface temperature measurements. This fully data-driven model strengthens our confidence in existing thermophysical models and furthermore it is four orders of magnitude faster than traditional inversion, allowing us to efficiently generate global maps characterizing these thermal processes.

To analyze the latent representation of the VAE further, we compare its latent 3 values to the H-parameter values estimated

by Hayne et al. (2017) at four different crater locations in Figure 10. The H-parameter has a one-to-one mapping with thermal inertia at a constant temperature and is estimated using a thermophysical model where it governs the assumed density variation with depth. There are multiple similarities and differences between the latent 3 map and the H-parameter map at each crater. At Tycho, the maps are largely similar within the crater, suggesting a strong thermal inertia anomaly. At Chaplygin-B, there is a “cold spot” in the H-parameter map and no apparent anomaly in the latent 3 map. Bandfield et al. (2014) define cold spots as extensive regions of low thermal inertia and highly insulating surface material that usually surrounds small, recent impact events. This appears to be a limitation of the VAE, which struggles to differentiate between the subtler colder temperatures that produce these types of anomalies. At both Tycho and Chaplygin-B, there is an acquisition footprint in the latent 3 map (vertically striped lines), which is likely a result of the differences in the point density of each LRO orbit affecting the GP interpolation and VAE reconstruction of each temperature profile. At La Condamine-S the latent 3 map is in strong agreement with the H-parameter map and arguably has a lower noise content; this may be because the higher density of point measurements at higher latitudes makes the VAE inference more stable. At Moore-F, we notice that both the H-parameter map and latent 3 map show coherent noise outside of the crater. While the H-parameter map appears to erroneously correlate to the north–south slope, the latent 3 map correlates to the east–west slope. Looking carefully at the sensitivity of our latent variables in Figure 5 and comparing them to the physical model sensitivity in Figure 3, we notice that the VAE latent representation does not manage to fully disentangle the effect of solar onset with thermal conductivity factor and albedo, which could explain

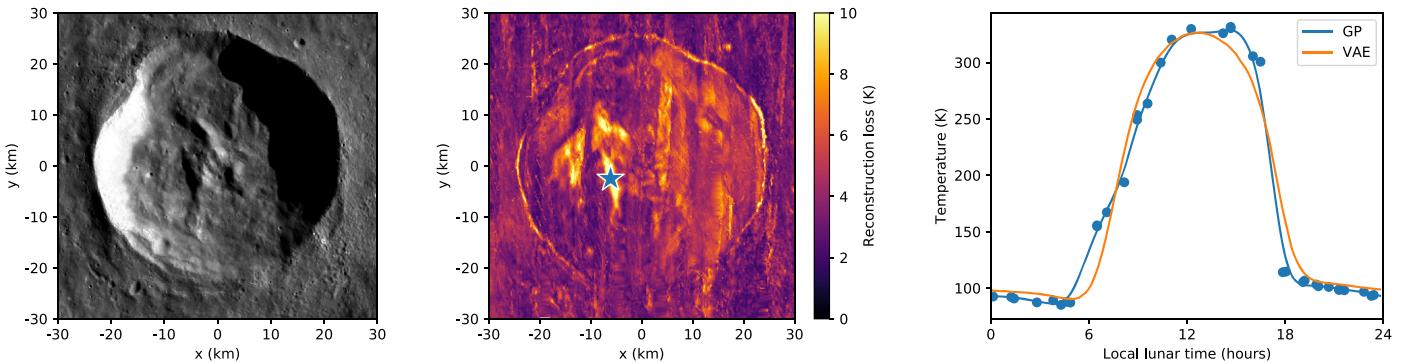


Figure 11. Example anomaly in our VAE reconstruction loss map, located within the Rydberg crater (263.6° lon./ -46.5° lat.). The left panel shows an optical image of the crater (LROC WAC Global Morphologic base map; Speyerer et al. 2011), the middle shows a map of the L1 reconstruction loss of the VAE, and the right shows an example interpolated profile and its resulting VAE reconstruction extracted at the star location in the loss map. The loss anomaly here is most likely caused by geometric errors, predominantly from the east–west-facing slope of Rydberg’s central peak.

this observation. This may be a result of the fundamental assumptions of the VAE model (for example, the Gaussian constraints on the posterior and prior latent distributions) or hint at some other physical process(es) affecting the latent representation, and further research is required to understand this effect.

As shown in Figure 7, the latent 1 variable is sensitive to the onset of solar radiation and thus depends on the aspect of the terrain, where a west-facing slope would heat up earlier in the day, while an east-facing slope would heat up later but also cool down later (resulting in a shifted thermal profile). The thermal behavior of north and south-facing slopes is represented by the latent 2 variable, which is sensitive to the peak temperature or effective albedo of the lunar surface. Closer to the equator, its behavior is governed by the albedo of the surface, whereas closer to the poles, it is increasingly controlled by topography as the angle between the incoming solar illumination and the surface increases. Interestingly, both the latent 1 and 2 maps feature local variations close to the equator, which indicates different physical properties of the surface in these regions (such as surface roughness) and is particularly evident when comparing equatorial mare and highland regions. Such local variations are less abundant in the latent 3 map, highlighting how it is largely decoupled from geometric artifacts caused by illumination and topography.

The latent 4 map shown in Figure 8 is generally darkest inside small and narrow craters and brighter for flatter regions. One possible explanation is that it is sensitive to cumulative illumination, i.e., to the relief or a large-baseline surface roughness, as a narrow crater would be illuminated for a shorter period of time than a flat mare region. Another explanation is that it is sensitive to a dependence of solar incidence angle on albedo, as modeled by Vasavada et al. (2012). As observed in the latent 1 and 2 maps, the latent 4 map also exhibits minor local variations; in particular, it has generally lower values close to the equator. This appears counterintuitive, as the equatorial region would receive more direct sunlight on average; this variable seems to be sensitive to and is not fully decoupled from the effective albedo to some extent.

The loss map generated by the VAE shown in Figure 9 indicates the input profiles that the VAE has struggled to reconstruct. These profiles are likely to be outside of the VAE’s training distribution and the representative set of temperature variations observed on the lunar surface, and therefore this map is a useful resource because it can be used to search for extreme

thermal anomalies. We show an example of a high loss anomaly within the Rydberg crater in Figure 11, along with a temperature profile extracted from within this anomaly. We find that the profile has very gradual heating in the morning that the VAE is unable to model. There are multiple explanations for this profile; it could likely be a result of complex topology/geometry being integrated over Diviner’s FOV and our 200×200 m bin size or from extreme underlying thermal parameters. We observe another regional loss anomaly in Mare Smythii, which is on the lunar nearside. The reason for the slightly above-average loss values in this region is unknown but could represent anomalous thermophysical behavior. Many more analyses could be done with this loss map.

There are many possible ways to extend this work. First, it could be powerful to combine traditional physical models and the VAE together to combine their relative strengths. For example, one could manipulate the latent representation of an input profile before reconstructing it to remove unwanted factors of variation, such as effective albedo, and then use an existing thermophysical model to invert for underlying thermal parameters. This may enable a simpler physical model to be used and combine the strengths of both approaches. Traditional thermophysical models (such as Bandfield et al. 2011 and Hayne et al. 2017) either require terrain modeling to remove the effect of slopes and isolate thermophysical effects or restrict themselves to low slope angles (Feng et al. 2020). Our model is able to remove the effect of slopes very efficiently without the need for terrain modeling. Going the other way and using existing derived Diviner products, such as soil temperature, instead of the raw temperature measurements could help the VAE disentangle different thermophysical effects. Physical constraints could also be encoded directly into the VAE: for example, by changing the assumed prior distributions of the latent variables in Equation (A4) to those expected of thermophysical parameters on the lunar surface or by adding physics constraints directly into the VAE’s loss function.

A downside of the standard VAE used in this work is that the encoding functions $g(x)$, $h(x)$, and decoding function $f(z)$ (as defined in Equations (A3)–(A6)) are in general not identifiable—that is, many different latent representations could exist that produce global optima when training. Applying recent work that investigates this may improve the disentangling ability of our VAE (Khemakhem et al. 2020).

Another direction would be to incorporate the rest of the wavelength channels from the Diviner instrument into the VAE

workflow. The VAE framework is readily extendable in this regard and it may be able to disentangle cross-channel effects such as anisothermality to estimate rock abundance (channels 6–8; Bandfield et al. 2011; Williams et al. 2016, 2017) or to recognize the Christiansen feature (channels 3–5) that is indicative of bulk silicate (Greenhagen et al. 2010). We discarded the standard deviation vector of the approximate latent posterior distribution produced by the VAE when carrying out inference, but this could also be used to obtain uncertainty estimates of the underlying thermal parameters. The ability of the VAE to carry out fast inference may mean it is useful for real-time inference (for example, in detecting transient temperature anomalies from fresh impacts).

Finally, as our method is physics-agnostic, it could be applied to other similar planetary science and exploration data sets. For example, other useful applications may be to use the VAE to model differences in optical illumination of the lunar surface over the lunar day or to localize regions with overnight temperature flattening that might indicate multilayer subsurfaces, such as cryptomare.

5. Conclusion

We have shown that AI can benefit space science and exploration by offering novel ways to extract physical models from big data sets in an extremely accelerated timeframe. We found that a VAE was able to learn about multiple different thermophysical processes on the lunar surface from over 9 yr of Diviner surface temperature data by reconstructing the observed variation in surface temperature over the lunar day. Our AI-derived model complemented and increased our confidence in existing thermophysical models of the lunar surface. Furthermore, the VAE was four orders of magnitude faster at inferring underlying model parameters than traditional inversion, which allowed us to efficiently create global thermal anomaly maps. Our method is fully data driven and could therefore be used to fast-track other similar space science and exploration workflows, without the need for an a priori physics model to be developed. Further work is needed to understand how best to combine traditional physical modeling with the VAE and how to fully disentangle physical processes using the VAE. We believe our work demonstrates the potential of big data-driven AI for aiding and accelerating planetary science and exploration.

The initial results of this study were produced during NASA’s 2019 Frontier Development Lab (FDL). We would like to thank the FDL and its partners (Luxembourg Space Agency, Google Cloud, Intel AI, HP Enterprise, Element AI, NVIDIA, and USGS); the SETI Institute for hosting the FDL; Jean-Pierre Williams, Paul Hayne, and the Diviner team for their help working with the Diviner data as well as for providing the H-parameter map; as well as our FDL 2019 lead mentors Dennis Wingo and Allison Zuniga and mentors Daniel Angerhausen, Frank Soboczenski, Abigail Calzada Diaz, and Ignacio Lopez-Francos.

Appendix A VAE Definition

The goal of a VAE is to (implicitly) learn a distribution, $p(x)$, $x \in X$, where X is a vector space, which is as similar as possible to an unknown target distribution $p^*(x)$, given a set of

samples from this distribution $\mathcal{D} = \{x_1, \dots, x_n\}$. VAEs make the assumption that x depends on some set of underlying latent variables $z \in Z$, which are distributed according to some known distribution $p(z)$, and model the likelihood $p(x|z)$ and an approximation of the posterior distribution $p(z|x)$.

To help them learn, VAEs use variational Bayes inference (Gelman et al. 2013). In this approach, instead of computing the posterior $p(z|x)$, which is often intractable, one searches for a more tractable distribution $q(z|x)$ that approximates the posterior. This is most commonly found by minimizing the Kullback–Leibler (KL) divergence between $q(z|x)$ and $p(z|x)$, which can be written as

$$\begin{aligned} D_{\text{KL}}(q(z|x) \| p(z|x)) &= \mathbb{E}_{q(z|x)} \left[\log \frac{q(z|x)}{p(z|x)} \right] \\ &= -\mathbb{E}_{q(z|x)} [\log p(x|z)] + D_{\text{KL}}(q(z|x) \| p(z)) + \log p(x). \end{aligned} \quad (\text{A1})$$

As $p(x)$ is fixed with respect to $q(z|x)$, maximizing the quantity $\mathbb{E}_{q(z|x)} [\log p(x|z)] - D_{\text{KL}}(q(z|x) \| p(z))$ is equivalent to minimizing the KL divergence between $q(z|x)$ and $p(z|x)$. Furthermore, as the KL divergence is always positive this quantity is a lower bound on $\log p(x)$ and is therefore known as the variational or evidence lower bound (ELBO).

VAEs assume parameterized models of the likelihood and the approximate posterior, given by $p(x|z, \theta)$ and $q(z|x, \theta)$, where θ represent model parameters. By rewriting Equation (A1), one can write an expression bounding the log-likelihood of the training data,

$$\begin{aligned} \log p(\mathcal{D}|\theta) &\geq \sum_{i=1}^n \mathbb{E}_{q(z_i|x_i, \theta)} [\log p(x_i|z_i, \theta)] \\ &\quad - D_{\text{KL}}(q(z_i|x_i, \theta) \| p(z_i)), \end{aligned} \quad (\text{A2})$$

assuming the training samples are independently and identically distributed. Therefore we can use the maximum likelihood method to simultaneously learn θ and ensure $q(z|x, \theta)$ approximates $p(z|x)$, given the training data. To be able to compute the ELBO, VAEs typically further assume that

$$p(x|z, \theta) = \mathcal{N}(x; f(z; \theta_f), \sigma^2 I) \quad (\text{A3})$$

$$p(z) = \mathcal{N}(z; 0, I) \quad (\text{A4})$$

$$q(z|x, \theta) = \mathcal{N}(z; g(x; \theta_g), \text{diag}(h(x; \theta_h)^2)), \quad (\text{A5})$$

where f , g , and h are all deep neural networks parameterized by $\theta = \{\theta_f, \theta_g, \theta_h\}$, σ is a hyperparameter, and \mathcal{N} denotes a normal distribution. From a deep learning standpoint, the networks $g(x; \theta_g)$ and $h(x; \theta_h)$ can be seen as a nonlinear encoder that maps an observation x to possible values of z while $f(z; \theta_f)$ can be seen as a nonlinear decoder that maps a latent vector z back to possible values of x . We note that in this definition (Equation (A5)), the latent variables z are assumed to be independent from one another; in general it is also possible to parameterize $p(z)$ along with $p(x|z, \theta)$ and $q(z|x, \theta)$. Under the definitions above, the task of maximum likelihood

estimation becomes

$$\hat{\theta}_f, \hat{\theta}_g, \hat{\theta}_h = \arg \min_{\theta_f, \theta_g, \theta_h} \left\{ \sum_{i=1}^n \mathbb{E}_{q(z_i|x_i, \theta_g, \theta_h)} \left[\frac{\|x_i - f(z_i; \theta_f)\|^2}{2\sigma^2} \right] + D_{\text{KL}}(q(z_i|x_i, \theta_g, \theta_h) \| p(z_i)) \right\}. \quad (\text{A6})$$

The KL term in Equation (A6) can easily be computed analytically as it involves two normal distributions. When calculating the expectation over $q(z|x, \theta)$, which is usually carried out via sampling, a “reparameterization trick” is used. One notes that sampling $z \sim q(z|x, \theta)$ is equivalent to sampling $\epsilon \sim \mathcal{N}(\epsilon; 0, I)$ and computing $z = h(x; \theta_h)\epsilon + g(x; \theta_g)$, which allows gradient descent methods to be used to find $\hat{\theta}$.

The expectation term in Equation (A6) can be interpreted as a reconstruction loss between the VAE and the training data, while the KL term can be interpreted as a regularizer that tries to ensure that the predicted latent variables z are normally distributed. This has the effect of keeping input samples that

are similar in X close together in the latent space Z . This regularization is very useful when one wants to learn “meaningful” representations that vary smoothly over the input space. Typically, one chooses the dimensionality of Z to be much lower than the dimensionality of X , forcing the VAE to learn a salient representation of the input data. While we present a “vanilla” VAE here, many variations have been proposed. One variation is the β -VAE (Higgins et al. 2017). In this approach, the KL term in Equation (A6) is weighted by a hyperparameter $\beta \in \mathbb{R}^+$, which allows more control on the balance between the regularization and reconstruction terms during training.

Appendix B VAE Training

Table B1 lists the AOIs used for training the VAE. The anomalies are classified as (1) thermal anomalies, (2) microwave anomalies, (3) magnetic anomalies, or (4)

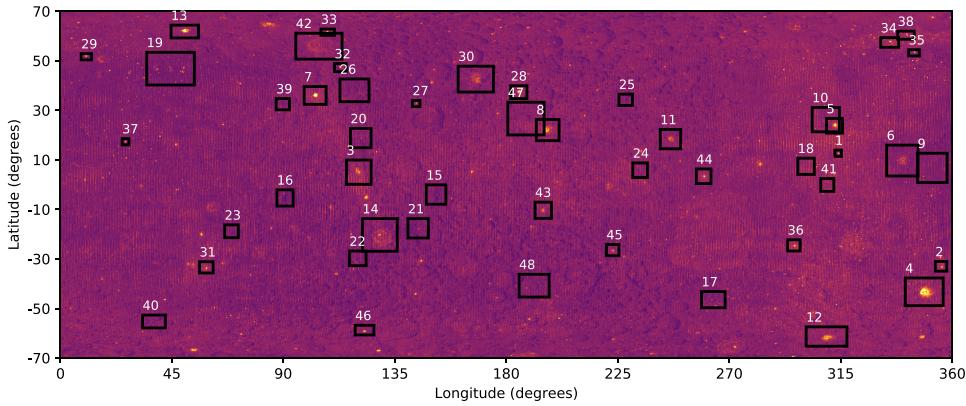


Figure B1. Locations of the AOIs used for training the VAE, overlaid on the latent variable 3 map repeated from Figure 8.

Table B1
AOIs Used for Training the VAE

Index	Name	Lat	Lon	Class					
1	Marius-A crater	12.58	313.96	1	25	Unnamed crater	34.21	228.28	1
2	Hell-Q crater	-33.00	355.53	1	26	Cantor crater	37.98	118.70	2
3	King crater	4.91	120.47	1	27	Unnamed crater	32.66	143.68	1, 2
4	Tycho crater	-43.29	348.67	1, 2	28	Moore-F crater	37.23	185.05	1
5	Aristarchus crater	23.70	312.52	1, 2	29	Egede-A crater	51.55	10.50	1
6	Copernicus crater	9.60	339.93	1, 2	30	Chandler crater	42.46	167.70	1
7	Giordano Bruno crater	35.95	102.88	1, 2	31	Furnerius-A crater	-33.51	58.92	1
8	Jackson crater	21.98	196.69	1	32	Rayet-Y crater	47.19	113.05	1
9	Unnamed region	6.72	351.97	1	33	Dugan-J crater	61.45	107.88	1
10	Aristarchus pyroclastic deposit	26.14	308.99	1	34	La Condamine-S crater	57.35	334.78	1
11	Ohm crater	18.30	246.29	1	35	Plato-M crater	53.08	344.54	1
12	Zucchius crater	-61.39	309.32	1	36	Byrgius-A crater	-24.57	296.05	1
13	Thales crater	61.70	50.28	1	37	Dawes crater	17.21	26.36	1
14	Tsiolkovsky crater	-20.38	129.03	1	38	Unnamed crater	60.31	341.33	1
15	Chaplygin-B crater	-4.08	151.67	1	39	Unnamed crater	32.40	89.84	1
16	Unnamed crater	-5.40	90.76	1	40	Hommel-B crater	-55.34	37.79	1
17	Rydberg basin	-46.45	263.59	3	41	Unnamed crater	-0.25	309.53	1
18	Reiner Gamma swirl	7.39	300.97	3	42	Compton crater	55.82	104.44	1
19	Atlas crater	46.78	44.50	1	43	Crookes crater	-10.38	194.98	1
20	Unnamed crater	18.68	121.31	1	44	Pierazzo crater	3.34	259.69	1
21	Unnamed crater	-17.69	144.40	1	45	Das crater	-26.50	222.94	1
22	Unnamed crater	-29.74	120.11	1	46	Fechner-T crater	-58.74	122.80	1
23	Unnamed crater	-18.93	69.14	1	47	Fitzgerald crater	26.62	187.96	4
24	Unnamed crater	5.82	233.99	1	48	Maksutov crater	-40.87	191.31	4

Note. See the appendix text for a description of the assigned classes for each AOI.

Table B2
VAE Network Parameters

Layer	Operation	In Channels	Out Channels	Filter Size	Stride	Padding							
1e	Wrap1D	1	1	1	1d	ConvT1D	<i>n</i>	32	2	2	0
2e	Pad1D	1	1	3	2d	ConvT1D	32	32	2	2	0
3e	Conv1D	1	32	3	1	1	3d	ConvT1D	32	32	2	2	0
4e	Conv1D	32	32	2	2	0	4d	ConvT1D	32	32	2	2	0
5e	Conv1D	32	32	2	2	0	5d	ConvT1D	32	32	2	2	0
6e	Conv1D	32	32	2	2	0	6d	ConvT1D	32	32	2	2	0
7e	Conv1D	32	32	2	2	0	7d	ConvT1D	32	32	2	2	0
8e	Conv1D	32	32	2	2	0	8d	Conv1D	32	1	1	1	0
9e	Conv1D	32	32	2	2	0	9d	Crop1D	1	1	-4
10e	Conv1D	32	32	2	2	0
11e	Conv1D	32	<i>n</i>	1	1	0
11e	Conv1D	32	<i>n</i>	1	1	0

Note. Each entry shows the parameterization of each layer in the network. The padding column shows the padding or cropping on each side of the first dimension of each layer’s input tensor. The profile input to the network has dimension 120×1 and is padded by 1 at its start and end by wrapping its end values to ensure no edge effects are present at the first and last samples in the first convolutional layer. It is also zero padded by 3 on each side to expand its dimension to 128×1 before being reduced by the encoder. The last two layers of the encoder are two separate convolutional filters that output the mean and standard deviation vectors of the VAE’s approximate latent posterior distribution each with dimension $1 \times n$, where n is the number of latent variables. The output of the decoder is a tensor of shape 128×1 that is cropped to 120×1 to match the input profile dimensions.

background. The selection of these training sites is based on surface temperature measurements and their variations in the thermal range (Hayne et al. 2017) and in the microwave range (Zheng et al. 2014), surface roughness and rock abundance (Bandfield et al. 2011), as well as on magnetic measurements (Tsunakawa et al. 2010). Thermal and microwave anomalies can either be “hot” or “cold” relative to their surroundings and general environment. Magnetic anomalies show an increased total force in comparison to their immediate surroundings and the general surface of the Moon. The background class contains locations that do not feature any thermal, microwave or magnetic anomalies. Figure B1 shows the locations of the AOIs overlain on the latent variable 3 map generated using the VAE.

Table B2 details the architecture of the VAE. The first convolutional layer of the encoder expands the number of channels (the second dimension) of the input profile from 1 to 32. The seven subsequent convolutional layers gradually reduce the number of samples (the first dimension) of the

input profile to one, keeping the number of channels constant. The final layer of the encoder consists of two separate convolutional filters that output the mean and standard deviation vectors. The decoder takes a vector of latent values of dimensionality $1 \times n$ and outputs a reconstruction of the input profile. It consists of seven transposed convolutional layers with 32 channels that gradually increase the number of samples in the latent vector. A final convolutional layer is used to reduce the number of channels from 32 to 1, outputting the reconstructed profile.

Appendix C Global Maps

Figure C1 shows the tessellation of the Moon’s sphere used for generating global maps. A tessellation is used to both extract 1D surface temperature profiles over the sphere and to interpolate the generated latent variables onto the 2D map projection. Figure C2 shows VAE reconstructions of profiles

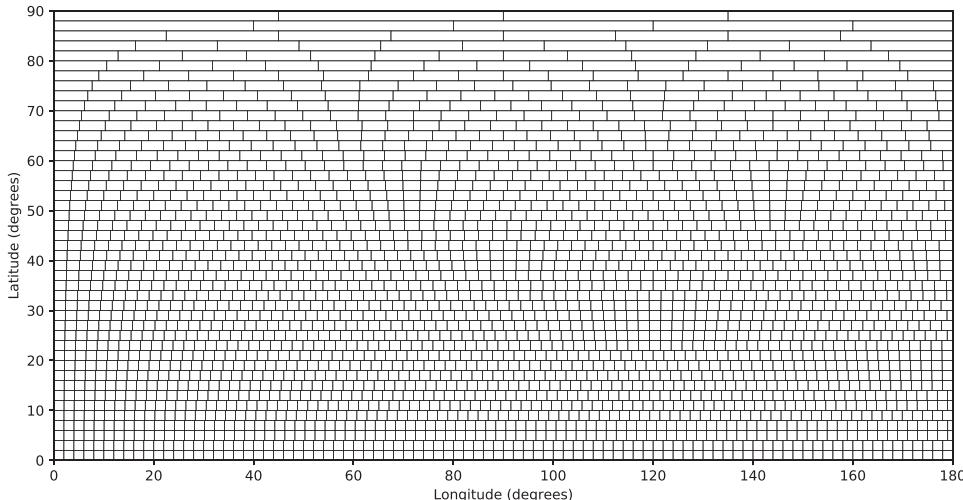


Figure C1. Tessellation of the Moon’s sphere used for generating global maps. We tessellate the sphere into approximately equal area latitude–longitude boxes, keeping the latitude length constant and rounding the longitudinal length in each latitude strip so that an integer number of boxes fit into each strip. Of the boxes of approximately 10×10 km shape, 379,448 boxes of approximately 10×10 km shape on the Moon’s surface are used for the first VAE inference step and 10,322 boxes of approximately 61×61 km shape are used for the second interpolation step. The first quadrant of boxes for the second step is plotted.

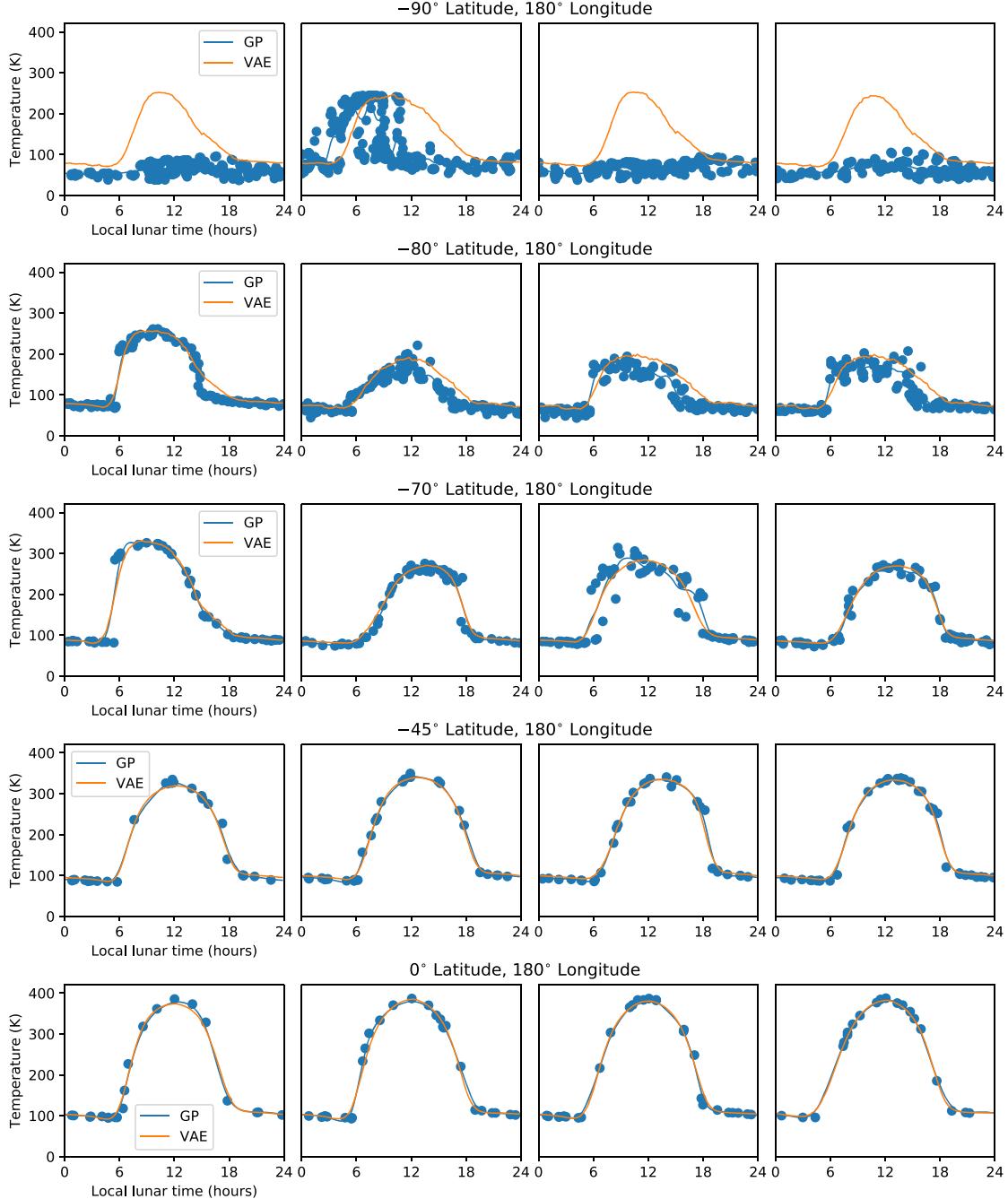


Figure C2. Example input surface temperature profiles (blue points), GP interpolations (blue lines), and VAE reconstructions (orange lines) with varying latitudes. Each row shows four randomly selected example profiles close to the coordinates shown in the row's title. As shown in Figure 9, we find that the VAE reconstruction loss is higher closer to the poles. The poles are also more densely sampled than the equator, as can be seen in Figure D1.

with varying latitudes. We find that the reconstruction quality reduces close to the poles. This is expected as the temperature variation at the poles is more complex as a result of the high latitude and the topography present (Williams et al. 2019).

Appendix D Diviner Data Preprocessing

Figure D1 shows the population of the 0.5×0.5 degree latitude and longitude bins after preprocessing. We observe a

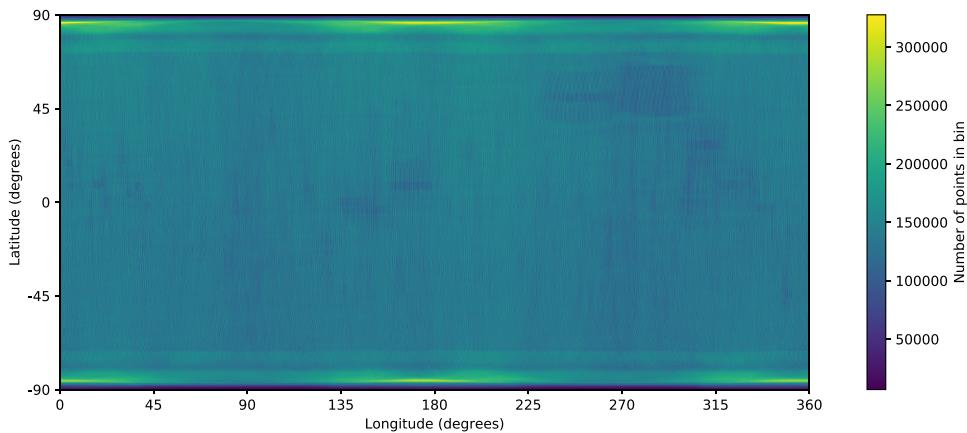


Figure D1. Population of the 0.5×0.5 degree latitude and longitude bins after preprocessing the Diviner data. Over 36 billion point measurements of the lunar surface are extracted from the Diviner instrument spanning 9 yr of acquisition. The band of lower coverage at 80N and 80S is caused by systematic Diviner calibration operations.

higher point density closer to the poles, which is expected because LRO's orbit means the Diviner instrument samples the poles more frequently than the equator.

ORCID iDs

- Ben Moseley <https://orcid.org/0000-0003-2238-1783>
 Valentin Bickel <https://orcid.org/0000-0002-7914-2516>
 Jérôme Burelbach <https://orcid.org/0000-0002-7509-9637>
 Nicole Relatores <https://orcid.org/0000-0002-6041-6388>

References

- Baldwin, R. B. 1964, *AJ*, **69**, 377
 Bandfield, J. L., Ghent, R. R., Vasavada, A. R., et al. 2011, *JGRE*, **116**, E00H02
 Bandfield, J. L., Song, E., Hayne, P. O., et al. 2014, *Icar*, **231**, 221
 Bickel, V. T., Aaron, J., Manconi, A., Loew, S., & Mall, U. 2020, *NatCo*, **11**, 2862
 Bickel, V. T., Lanaras, C., Manconi, A., Loew, S., & Mall, U. 2019, *ITGRS*, **57**, 3501
 Dalgaard, M., Motzoi, F., Sørensen, J. J., & Sherson, J. 2020, *npjQI*, **6**, 1
 Doersch, C. 2016, arXiv:[1606.05908](https://arxiv.org/abs/1606.05908)
 ESA 2019, ESA Space Resources Strategy (Paris: ESA), https://sci.esa.int/documents/34161/35992/1567260390250-ESA_Space_Resources_Strategy.pdf
 Fagents, S. A., Rumpf, M. E., Crawford, I. A., & Joy, K. H. 2010, *Icar*, **207**, 595
 Feng, J., Siegler, M. A., & Hayne, P. O. 2020, *JGRE*, **125**, e06130
 Frazier, P. I. 2018, arXiv:[1807.02811](https://arxiv.org/abs/1807.02811)
 Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (Boca Raton, FL: CRC Press), doi:[10.1201/b16018](https://doi.org/10.1201/b16018)
 Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge, MA: MIT Press)
 Greenhagen, B. T., Lucey, P. G., Wyatt, M. B., et al. 2010, *Sci*, **329**, 1507
 Hartmann, W. K. 1965, *Icar*, **4**, 157
 Hayne, P. O., Bandfield, J. L., Siegler, M. A., et al. 2017, *JGRE*, **122**, 2371
 Higgins, I., Matthey, L., Pal, A., et al. 2017, in 5th Int. Conf. on Learning Representations, ICLR 2017, ed. S. Mohamed et al. (La Jolla, CA: ICRL)
 Hinton, G., Deng, L., Yu, D., et al. 2012, *ISPM*, **29**, 82
 Hornik, K., Stinchcombe, M., & White, H. 1989, *NN*, **2**, 359
 Ioffe, S., & Szegedy, C. 2015, in Proc. Int. Conf. Machine Learning, ICML 15, ed. F. Bach & D. Blei (New York: ACM), 448
 Joy, K. H., Crawford, I. A., Curran, N. M., et al. 2016, *EM&P*, **118**, 133
 Khemakhem, I., Kingma, D. P., Monti, R. P., & Hyvärinen, A. 2020, in Proc. 23rd Int. Conf. on Artificial Intelligence and Statistics, ed. S. Chiappa & R. Calandra (New York: PMLR), 2207, <http://proceedings.mlr.press/v108/khemakhem20a>
 Kingma, D. P., & Ba, J. 2014, arXiv:[1412.6980](https://arxiv.org/abs/1412.6980)
 Kingma, D. P., & Welling, M. 2014, in Proc. 2nd Int. Conf. on Learning Representations (Banff: ICLR) (arXiv:[1312.6114](https://arxiv.org/abs/1312.6114))
 Kodikara, G. R., & McHenry, L. J. 2020, *Icar*, **345**, 113719
 Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. 2015, in Proc. 33rd Conf. Machine Learning, ed. M. F. Balcan & K. Q. Weinberger (New York: PMLR), 1158, <http://proceedings.mlr.press/v48/larsen16.html>
 Mazarico, E., Neumann, G. A., Barker, M. K., et al. 2018, *AdSpR*, **62**, 3214
 Morningstar, W. R., Levasseur, L. P., Hezaveh, Y. D., et al. 2019, *ApJ*, **883**, 14
 Neal, C. R., Bradley, B., Beaty, D., et al. 2009, The Lunar Exploration Roadmap. Exploring the Moon in the 21st Century: Themes, Goals, Objectives, Investigations, and Priorities (Washington, DC: NASA), <https://solarsystem.nasa.gov/studies/182/the-lunar-exploration-roadmap-exploring-the-moon-in-the-21st-century-themes-goals-objectives-investig/>
 Paganini, M., De Oliveira, L., & Nachman, B. 2018, *PhRvL*, **120**, 042003
 Paige, D., Schofield, J., Calcutt, S., et al. 2011, NASA Planetary Data System, LRO-L-DLRE-4-RDR-V1.0 (St. Louis, MS: Washington Univ. St. Louis), <https://pds-geosciences.wustl.edu/lro/lro-l-dlre-4-rdr-v1/>
 Paige, D. A., Foote, M. C., Greenhagen, B. T., et al. 2010, *SSRV*, **150**, 125
 Paszke, A., Gross, S., Massa, F., et al. 2018, in Adv. Neural Info. Process. Systems, Vol. 32, ed. H. Wallach et al. (Red Hook, NY: Curran Associates), 8024
 Pierson, H. A., & Gashler, M. S. 2017, *Adv. Robotics*, **31**, 821
 Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (Cambridge, MA: The MIT Press)
 Rezende, D. J., Mohamed, S., & Wierstra, D. 2014, in Proc. 31st Int. Conf. Machine Learning, ICML 14, ed. E. P. Xing & T. Jebara (New York: ACM), 1278
 Russakovsky, O., Deng, J., Su, H., et al. 2015, *Int. J. Comput. Vision*, **115**, 211
 Sefton-Nash, E., Williams, J. P., Greenhagen, B. T., Aye, K. M., & Paige, D. A. 2017, *Icar*, **298**, 98
 Shoemaker, E. M. 1962, in Physics and Astronomy of the Moon, ed. Z. Kopal (New York: Elsevier), 283
 Silburt, A., Ali-Dib, M., Zhu, C., et al. 2019, *Icar*, **317**, 27
 Speyerer, E. J., Robinson, M. S., Denevi, B. W. & LROC Science Team 2011, *LPI*, **42**, 2387
 Spudis, P. D. 1996, The Once and Future Moon (Washington, DC: Smithsonian Inst. Press)
 Tsunakawa, H., Shibuya, H., Takahashi, F. & others 2010, *SSRV*, **154**, 219
 Vasavada, A. R., Bandfield, J. L., Greenhagen, B. T., et al. 2012, *JGRE*, **117**, OH18V
 Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Proc. 31st Int. Conf. Neural Info. Process. Systems, 41, ed. U. von Luxburg & I. Guyon (Red Hook: Curran Associates), 6000
 Williams, J. P., Greenhagen, B. T., Paige, D. A., et al. 2019, *JGRE*, **124**, 2505
 Williams, J. P., Paige, D. A., Greenhagen, B. T., & Sefton-Nash, E. 2017, *Icar*, **283**, 300
 Williams, J. P., Sefton-Nash, E., & Paige, D. A. 2016, *Icar*, **273**, 205
 Wu, B., Ross, W. K., Ludvig, P., Chung, P., & Seabrook, A. S. 2019, in Proc. 2019 IEEE/RSJ Int. Conf. Intell. Robots and Systems, IROS (Piscataway, NJ: IEEE), 3262, (doi:[10.1109/IROS40897.2019.8968124](https://doi.org/10.1109/IROS40897.2019.8968124))
 Zheng, Y.-C., Chan, K., Zhu, Y., et al. 2014, *LPI*, **45**, 2208