

Enhancing 3D planetary atmosphere simulations with a surrogate radiative transfer model

Tara P. A. Tahseen¹, João M. Mendonça^{2,3,4}, Kai Hou Yip¹ and Ingo P. Waldmann¹

¹Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

²Department of Space Research and Space Technology, Technical University of Denmark, Elektrovej 328, DK-2800 Kgs. Lyngby, Denmark

³Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

⁴School of Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, UK

Accepted 2024 October 27. Received 2024 October 18; in original form 2024 July 3

ABSTRACT

This work introduces an approach to enhancing the computational efficiency of 3D atmospheric simulations by integrating a machine-learned surrogate model into the OASIS global circulation model (GCM). Traditional GCMs, which are based on repeatedly numerically integrating physical equations governing atmospheric processes across a series of time-steps, are time-intensive, leading to compromises in spatial and temporal resolution of simulations. This research improves upon this limitation, enabling higher resolution simulations within practical time frames. Speeding up 3D simulations holds significant implications in multiple domains. First, it facilitates the integration of 3D models into exoplanet inference pipelines, allowing for robust characterization of exoplanets from a previously unseen wealth of data anticipated from *JWST* and post-*JWST* instruments. Secondly, acceleration of 3D models will enable higher resolution atmospheric simulations of Earth and Solar system planets, enabling more detailed insights into their atmospheric physics and chemistry. Our method replaces the radiative transfer module in OASIS with a recurrent neural network-based model trained on simulation inputs and outputs. Radiative transfer is typically one of the slowest components of a GCM, thus providing the largest scope for overall model speed-up. The surrogate model was trained and tested on the specific test case of the Venusian atmosphere, to benchmark the utility of this approach in the case of non-terrestrial atmospheres. This approach yields promising results, with the surrogate-integrated GCM demonstrating above 99.0 per cent accuracy and factor of 147 speed-up of the entire simulation executed on one graphics processing unit (GPU) compared to using the matched original GCM under Venus-like conditions.

Key words: radiative transfer – planets and satellites: atmospheres.

1 INTRODUCTION

3D atmospheric models, commonly known as *global circulation models* (GCMs), are key tools for studying the climates of Solar system planets including the Earth, and are becoming increasingly important in the characterization of exoplanet atmospheres. GCMs consist of multiple components, which individually model different atmospheric processes: each component numerically solves equations governing an atmospheric process across elements of a grid and across many time-steps until simulation convergence criteria are reached. Due to the large number of numerical integrations involved in GCM simulations, producing a simulated atmospheric state for a given input set of planetary and stellar parameters is incredibly computationally expensive and time-consuming.

GCMs are applied in exoplanet science in the forward modelling component of Bayesian atmospheric retrieval pipelines. To retrieve posterior distributions on planetary and stellar parameters from a single transit spectrum, the Bayesian retrieval framework requires of the order of tens of thousands of forward simulations corresponding

to different samples of input parameter space. With current state-of-the-art 3D modelling techniques, the compute resources and time needed to produce the required number of 3D simulations prohibit statistically rigorous inference of data from the *JWST* (Gardner et al. 2006) and further next-generation observational instruments yet to come, namely the *Ariel Space Telescope* (Tinetti et al. 2022). Acceleration without compromising the accuracy of GCMs is thus one clear method of facilitating inference using *JWST* and post-*JWST* data in exoplanet science.

Reducing GCM simulation time would also incur benefits in climate science of Earth and other Solar system planets. Increased speed of computation would enable simulations to be run at greater resolution, and/or with more physics included, thus enabling more realistic climate simulations to be achieved.

The past few years have yielded work in both exoplanet science and Earth climate science to accelerate 3D models. Much of this work in exoplanet science involves extending 1D models with extra parameters to reflect certain 3D atmospheric variations deemed necessary to account for 3D effects in phase-curve data (Changeat & Al-Refaie 2020; Feng, Line & Fortney 2020; Irwin et al. 2020; Chubb & Min 2022; Nixon & Madhusudhan 2022; Himes, Harrington & Baydin 2023). This approach is prone to introducing biases to the simulated

* E-mail: tara.tahseen.22@ucl.ac.uk

atmospheric states produced, and a fully 3D model parametrization is essential to ensuring that simulations are robust against both known and unknown biases resulting from model oversimplifications (Pluriel et al. 2020). Earth climate science, benefiting from a wealth of high-resolution observational measurements, has had a different set of methods employed, namely machine-learned surrogate models trained on such observations (Ukkonen 2022; Yao et al. 2023). Surrogate models have not been broadly explored outside of Earth climate science, but demonstrate the potential of speed-up without requiring an oversimplified model parametrization (Ukkonen 2022; Yao et al. 2023).

Of the components within GCMs, radiative transfer (RT) is often the slowest or least resolved process. In the OASIS GCM (Mendonça et al. 2014), the RT component contributes to 50–99 per cent of the total simulation runtime (depending on the temporal resolution of the RT, see Section 4.3 for more details) for massive and complex atmospheres such as Venus (Mendonça & Buchhave 2020). There is thus key scope to substantially improve GCM computational efficiency by targeting the RT component specifically.

To assess the effectiveness of our new approach, we are testing it on Venus’s atmospheric conditions. Simulating the Venus atmosphere in 3D is very computationally intensive due to the need for a complex RT model to accurately represent the energy balance in the atmosphere (e.g. Eymet et al. 2009; Lee & Richardson 2012; Mendonça et al. 2015). Venus has a substantial amount of CO₂, which generates a strong greenhouse effect (Sagan 1962) and is covered by highly reflective sulphuric acid clouds, obscuring the planet’s surface. Only about 2.5 per cent of the incoming solar radiation reaches the surface (e.g. Tomasko et al. 1980; Mendonça et al. 2015). The radiation model also requires fine spectral resolution to capture the spectral windows impacting energy exchange between the deep atmosphere and the upper layers above the clouds. Additionally, due to the high thermal inertia of the massive CO₂ atmosphere, the models need to be integrated over a long period to reach a statistically steady state (Mendonça & Read 2016). Using a surrogate model to represent the RT is key to enhancing the performance of 3D simulations while enabling a more realistic depiction of radiative processes at minimal cost. Therefore, Venus presents a significant challenge and serves as a benchmark for the complexities our new modelling approach may encounter in future applications.

2 DATA AND CODES

2.1 OASIS

OASIS is a planetary climate model composed of different coupled modules representing physical and chemical processes within planetary atmospheres (Mendonça & Buchhave 2020). The mathematics and assumptions of the RT component of OASIS are detailed in Mendonça et al. (2015).

The equations in OASIS are discretized over concentric icosahedral spatial grids (Mendonça et al. 2016; Deitrick et al. 2020). For the 3D simulations in this study, we set the grid to approximately 2 degrees in the horizontal and 49 vertical layers. This grid configuration results in a total of 10 242 columns covering the entire model domain. Our simulations started from the converged state obtained in Mendonça & Buchhave (2020) and were integrated for 5 Venus solar days (approximately 117 Earth days) using a time-step of 15 s. The model configuration is similar to the simulations in Mendonça & Buchhave (2020), and in the following sections, we describe the main physical modules relevant to this work.

2.1.1 OASIS-RT

The RT module of OASIS (henceforth referred to as OASIS-RT) models the interaction of radiation with gas and cloud species in the atmosphere in a two-stream manner. Radiation from two different sources are modelled separately: these are *solar* radiation (0.1–5.5 µm) and *thermal* radiation (1.7–260 µm). The solar radiation code utilizes the δ -Eddington approximation (Joseph, Wiscombe & Weinman 1976) in combination with an adding-layer method (Liu & Weng 2006; Mendonça et al. 2015), while the thermal code considers absorptivity and emissivity (Mendonça et al. 2015). The radiation scheme uses the k -distribution method to represent the gas absorption cross-sections, which are integrated over 353 spectral bands and 20 Gaussian points (Mendonça & Buchhave 2020).

In the OASIS code used for this project, the spatial distribution and radiative properties of the clouds, which are composed of sulphuric acid and water, were taken from Crisp (1986) and Mendonça et al. (2015). The main cloud deck is located between roughly 45 and 65 km altitude with layers of sub-micron particles below and above (Knollenberg & Hunten 1980). More details on the cloud properties can be found in Mendonça et al. (2015).

The RT code involves two steps of computation for each time-step; these are

- (i) **Computing the gas optics:** computing the optical properties of layer boundaries from the thermodynamic variables.
- (ii) **Computing the flow of radiation:** computing the upward- and downward-welling fluxes at each layer boundary, from the incident flux at the top of the column in combination with the optical properties at each layer boundary.

The model uses a k -distribution table (Lacis & Oinas 1991) to calculate the optical properties of each layer across the wavelength bands using pre-computed wavelength-dependent absorption coefficients for CO₂, SO₂, and H₂O. These coefficients are combined with a continuum absorption, mostly from CO₂–CO₂ collisions, and Rayleigh scattering from CO₂ and N₂ (Mendonça et al. 2015).

The RT model takes inputs of the density ρ , pressure p , temperature T , and chemical composition per grid element as outputted from the dynamical core THOR¹ (Mendonça et al. 2016). Heating rates are then calculated per grid element from fluxes outputted from the RT model, and these heating rates update the temperature profile of the atmosphere, which serves as input back into the dynamical core. Flux is used to compute the heating rate per layer according to equation (1).

$$\frac{dT}{dt} = \frac{1}{\rho c_p} \frac{dF^{\text{net}}}{dz}, \quad (1)$$

where dF^{net} is the spectral-integrated net radiative flux (W m^{−2}), ρ is the atmospheric density (kg m^{−3}), and c_p is the specific heat capacity at constant pressure (900 J kg^{−1} K^{−1}).

In order to improve the computational efficiency of 3D simulations with RT, Venus GCMs traditionally do not update the radiative fluxes from the solar and thermal schemes at every time-step (Lebonnois, Sugimoto & Gilli 2016; Mendonça & Read 2016; Mendonça & Buchhave 2020). In the 3D simulations with explicit RT used in this study, the fluxes calculated from the solar radiation scheme were updated every 2880 steps, and the thermal radiation fluxes were updated every 320 steps. These values are adjusted by the model user and are specific to Venus’s simulations. Larger values could

¹The model component that governs the resolved 3D fluid flow evolution.

potentially cause model instabilities. Although this approach introduces some inaccuracies in the heating/cooling rates, the robustness of the simulation is not compromised because the composition of the atmosphere and clouds remains constant over time. With the efficiency of the new surrogate model described in the next section, we are now able to update radiative fluxes at every time-step. To enhance the stability of the model with the new surrogate model, we also apply the three-step Adams–Bashforth method to the heating rate calculated from the surrogate radiative fluxes.

2.2 Data

The data used for this project are 3D data simulated by OASIS corresponding to input parameters of Venus (Mendonça & Buchhave 2020). The data used in this work are from a simulation of Venus using OASIS-RT. The data consist of 1000 time-steps of simulation, with the time interval between consecutive time-steps within the recorded data set, Δt , set to 12 h. These data correspond to time-steps of a full simulation for which the RT update has been executed. The data cover the entire icosahedral grid with dimensions $10\,242$ columns \times 49 layers.

At the sample level (per atmospheric column), the data comprise the following quantities: pressure p , temperature T , and gas density ρ , all defined per layer; upwelling short-wave flux $F^{\text{SW},\uparrow}$, downwelling short-wave flux $F^{\text{SW},\downarrow}$, upwelling long-wave flux $F^{\text{LW},\uparrow}$, and downwelling long-wave flux $F^{\text{LW},\downarrow}$, all defined per layer boundary; and cosine of the solar zenith angle μ , short-wave surface albedo α_{SW} , long-wave surface albedo α_{LW} , and surface temperature T_0 , all defined per column.

Fig. E1 and Fig. E2 illustrate the distribution of samples within the data set as a function of $\cos \mu$ and surface temperature.

3 METHODS

3.1 Surrogate modelling

Surrogate models, or *emulators*, are approximate mathematical models that model outcomes of interest, whereby the emulator mechanism does not necessarily reflect the physical mechanism that produces the outcomes. Surrogate models are useful in cases where the physical mechanism producing such outcomes is not well understood, or in cases where modelling the physical mechanism is excessively computationally demanding; in the case of the latter, the aim is to produce a surrogate model that is computationally efficient compared to the physical model while maintaining accuracy. Machine learning provides a framework by which surrogate models can be produced: deep neural networks adhere to the Universal Approximation Theorem, and can (in theory) model any arbitrarily complex non-linear relationship, thus providing a function space whereby a suitable surrogate function is almost certain to exist (Goodfellow, Bengio & Courville 2016).

Research into surrogate modelling of exoplanetary atmospheres is relatively nascent (Himes et al. 2022, 2023; Unlu et al. 2023). The use of surrogate models within Earth climate science, though still new, is much more established and well explored (Ukkonen 2022; Mukkavilli et al. 2023; Yao et al. 2023). The development of surrogate models for atmospheric models of exoplanets can thus be informed by the development of such surrogate models for the parameter space of Earth.

An example where surrogate modelling has already proved valuable within 3D modelling of exoplanetary atmospheres is work by Schneider et al. (2024), who utilized DeepSets to achieve fast and

accurate mixing of correlated- k opacities. Their work exemplifies how machine learning can be leveraged to successfully speed up a single process of a GCM while retaining accuracy, thus establishing a basis for further exploration into the integration of surrogates within GCM frameworks.

In this work, two surrogate models were produced for 3D modelling of Venus using OASIS: one surrogate model to emulate the short-wave RT schema, and one to emulate the long-wave RT schema (see Section 2.1.1 for details on the computations for the two radiation schemas).

3.2 Data pre-processing

Data were pre-processed separately for the long-wave and short-wave regimes. Each model took input of two types of data: variables defined at the grid-element level, referred to as *vector* variables of dimension $(n_{\text{columns}},) = (49,)$, and variables defined at the column level, referred to as *scalar* variables.

The short-wave surrogate model took scalar inputs of surface temperature, T_0 ; gas density of the lowest altitude layer, ρ_S ; pressure of the lowest altitude layer, p_S ; cosine of the solar zenith angle μ ; and short-wave surface albedo α_{SW} . The long-wave surrogate model took scalar inputs of surface temperature, T_0 ; gas density of the lowest altitude layer, ρ_S ; pressure of the lowest altitude layer, p_S ; and long-wave surface albedo α_{LW} .

Both models took input of the same vector variables, which were as follows: temperature T ; pressure p ; and gas density ρ . Vector variables were scaled as

$$x_{i,j} = \frac{\log_e(x_{i,j})}{\log_e(x_{i,0})} \quad (2)$$

for $x \in \{T, p\}$, and

$$x_{i,j} = \left(\frac{x_{i,j}}{x_{i,0}} \right)^{0.25} \quad (3)$$

for $x = \rho$, for the i th column and j th atmospheric level.

Scalar variables were then re-scaled as

$$x_i^{\text{scaled}} = \frac{x_i - \min_{i \in S_{\text{train}}} x_i}{\max_{i \in S_{\text{train}}} x_i - \min_{i \in S_{\text{train}}} x_i} \quad (4)$$

for $x \in \{T_0, p_0, \rho_0\}$, where S_{train} is the training set.

Targets were scaled as follows:

$$u_j^{\text{LW}} = \frac{y_j^{\text{LW}}}{A^{\text{LW}}} \quad (5)$$

for $y_j \in \{F_j^{\text{LW},\uparrow}, F_j^{\text{LW},\downarrow}\}$ and

$$u_j^{\text{SW}} = \frac{y_j^{\text{SW}}}{B^{\text{SW}}} \quad (6)$$

for $y_j \in \{F_j^{\text{SW},\uparrow}, F_j^{\text{SW},\downarrow}\}$, for the j th altitude level ($j \in [0, 49]$ where 0 indexes the ground level and 49 indexes the top level of the atmospheric column), where A^{LW} and B^{SW} are scaling factors linear in T_0 and μ , respectively, and fitted from the data:

$$A^{\text{LW}}(T_0) = a_1 T_0 + a_2, \quad (7)$$

$$B^{\text{SW}}(\mu) = b_1 \mu + b_2, \quad (8)$$

where (a_1, a_2) are constants fitted using $y_0^{\text{LW},\uparrow}$ and (b_1, b_2) are fitted using $y_{49}^{\text{SW},\downarrow}$ across all columns of the training set S_{train} . Values of (a_1, a_2, b_1, b_2) are retained for model prediction post-processing. Residuals between the targets $y_0^{\text{LW},\uparrow}$, $y_{49}^{\text{SW},\downarrow}$, and the

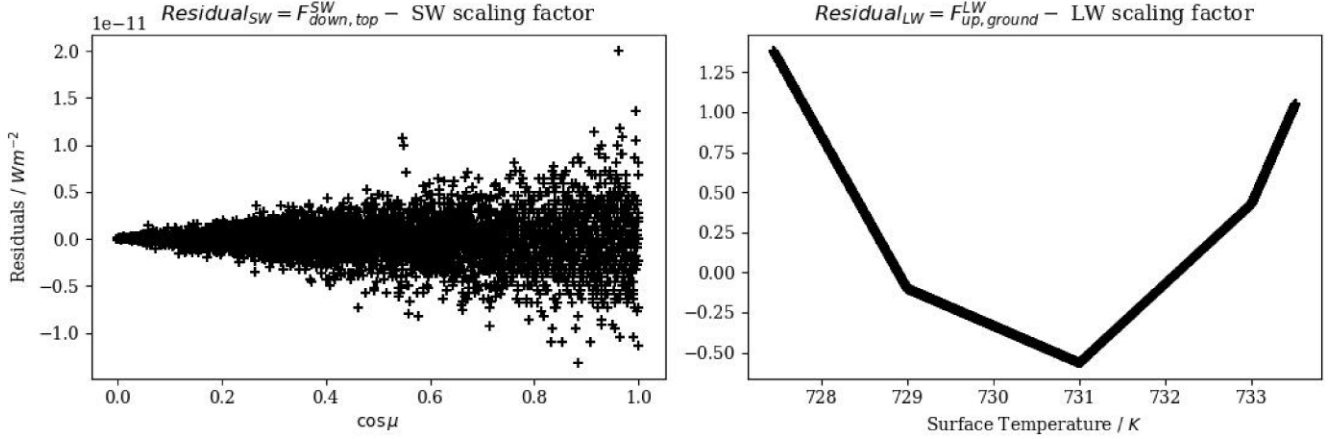


Figure 1. The figures above display scatter plots illustrating the residuals between scaling factors $A^{LW}(T_0)$ and $B^{SW}(\mu)$ and their respective targets $y_0^{LW,\uparrow}$ and $y_{49}^{SW,\downarrow}$, plotted over 1 epoch of 10242 test columns covering the entire icosahedral grid. Left: This panel shows the residuals between the scaling factors $B^{SW}(\mu_i)$ used for the short-wave targets of a given column i and the downward-welling short-wave flux at the top of the column $F_{i,49}^{SW,\downarrow}$. $B^{SW}(\mu_i)$, a linear function of the cosine of the solar zenith angle μ_i , approximates $F_{i,49}^{SW,\downarrow}$. Right: This panel displays the residuals between the scaling factors $A^{LW}(T_{i,0})$ used for the long-wave targets of a given column i and the upward-welling long-wave flux at the ground level $F_{i,0}^{LW,\uparrow}$. Here, $A^{LW}(T_{i,0})$ approximates $F_{i,0}^{LW,\uparrow}$ as a linear function of surface temperature $T_{i,0}$. Both: In both the long-wave and short-wave cases, the preferred quantities $F_{i,0}^{LW,\uparrow}$ and $F_{i,49}^{SW,\downarrow}$ to use for scaling flux profiles across atmospheric levels involve complex calculations. The figures demonstrate that simple linear functions $A^{LW}(T_0)$ and $B^{SW}(\mu)$ yield close approximations with low residuals, making them suitable scaling factors instead.

respective scaling factors approximating the value of these targets, are displayed in Fig. 1. These simple linear scaling methods were chosen for data pre-processing in this work instead of using exact computations of $y_0^{LW,\uparrow}$ and $y_{49}^{SW,\downarrow}$ as the latter computations are much more involved, and the more simple computations produce results with an acceptably small marginal difference in the values of the fluxes.

Columns across all epochs were shuffled and split into train, test, and validation data sets, in the ratio 70:15:15.

3.3 Model architecture

Model architecture was chosen to be based on recurrent neural networks (RNNs),² as RNNs structurally incorporate the spatial dependence of the training data, which fits naturally in this scenario. RNN layers were implemented in the form of gated recurrent units (GRUs; for further details on GRUs, see Cho et al. 2014). Simple RNNs are susceptible to short-term memory problems, whereby information propagated forwards diminishes quickly. A common implementation of RNNs is the GRU that utilizes a more sophisticated mechanism for propagating information ‘memory’ forwards through a sequence: this is the base of the model architecture we employ in this work. The specific architecture of the benchmark was chosen to be that used by Ukkonen (2022) (illustrated in Fig. 2), which utilized a bidirectional RNN-based architecture to create a two-stream RT emulator for Earth, trained using observational data. Ukkonen’s model performed with ≤ 0.5 per cent mean absolute error (MAE) for the upwelling and downwelling fluxes on the test set (Ukkonen 2022), suggesting its potential efficacy for developing surrogates trained on analogous simulated data. The models used in this work were constructed and trained using TENSORFLOW version 2.12.0 (Abadi et al. 2016), and converted into Open Neural Network Exchange (ONNX) format for integration within OASIS. Input data

to the surrogate model are detailed in Table 1, and surrogate model parameters are detailed in Table C1 and Table C2.

3.4 Model training

Both models were trained in a supervised, end-to-end fashion. An Adam (Kingma & Ba 2017) optimizer was used in combination with a cyclical learning rate. Models were trained using TENSORFLOW on an NVIDIA A100 GPU.

3.4.1 Loss function

The loss function was constructed as a combination of the mean percentage error (MPE) between the predictions and targets, per output. The mean error (ME) for the i th test column and k th target variable is defined as follows:

$$ME_{i,k} = \sum_{j=0}^{n_{\text{levels}}-1} \frac{|\hat{y}_{i,j,k} - \tilde{y}_{i,j,k}|}{n_{\text{levels}}}, \quad (9)$$

where $\hat{y}_{i,j,k}$ is the target for the i th test column, j th atmospheric level, and k th target variable, where $k = 0$ corresponds to downwelling flux and $k = 1$ corresponds to upwelling flux, and n_{levels} is the number of atmospheric levels ($n_{\text{levels}} = 50$ in this work). Normalization factors were defined as

$$\text{norm}_{i,k} = \frac{\sum_{j=0}^{n_{\text{levels}}-1} |\hat{y}_{i,j,k}|}{n_{\text{levels}}}, \quad (10)$$

such that MPE of the i th test column and k th target variable can be expressed as

$$\text{MPE}_{i,k} = \frac{ME_{i,k}}{\text{norm}_{i,k}}. \quad (11)$$

The loss function per sample was then defined as

$$\text{loss}_i = \frac{1}{2} \sum_k \text{MPE}_{i,k}, \quad (12)$$

²See Appendix A for more details on neural networks and RNNs.

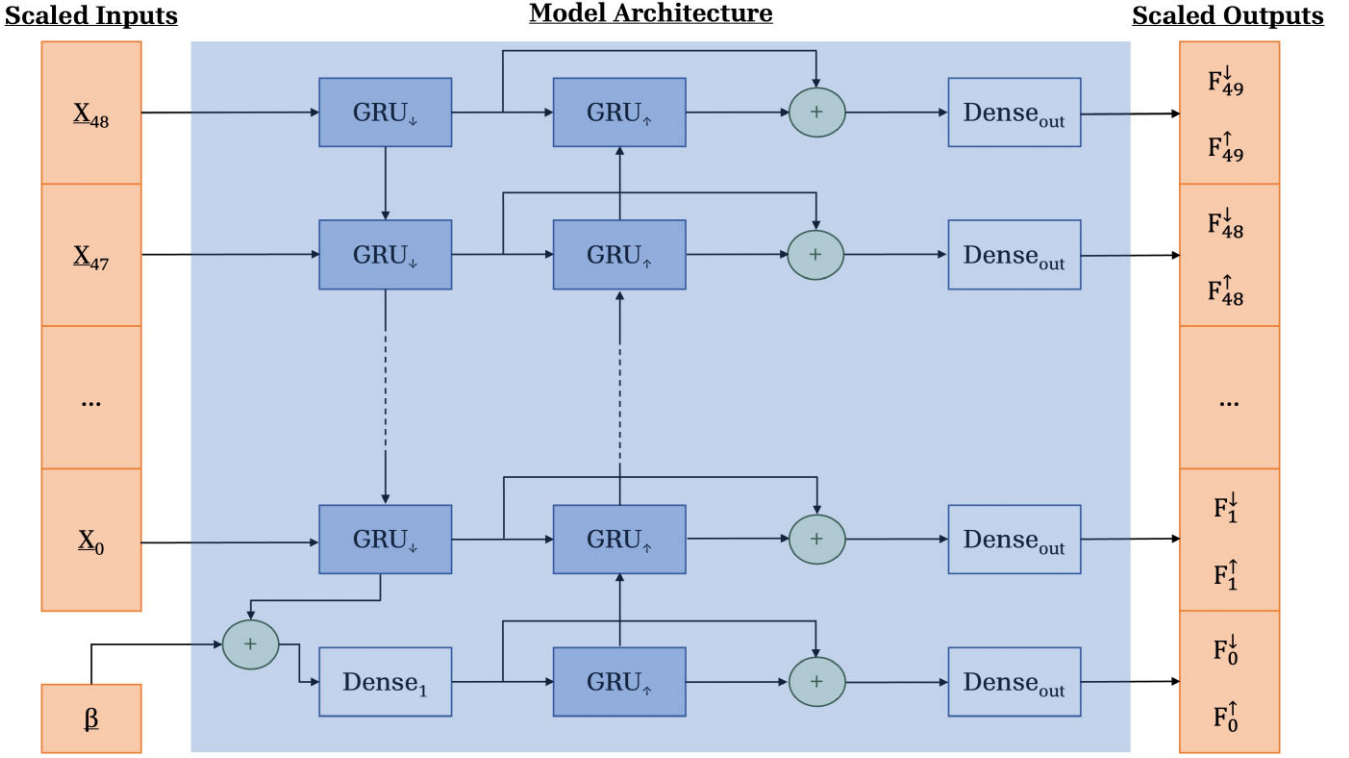


Figure 2. Figure illustrating the architecture of the learned surrogate models used in this work. Surrogate models for both the long-wave and short-wave schemas took scaled vector inputs $\mathbf{X}_j = (\tilde{p}_j, \tilde{T}_j, \tilde{\rho}_j)$ for the j th layer where \tilde{p}_j , \tilde{T}_j , and $\tilde{\rho}_j$ are scaled pressure, scaled temperature, and scaled density of the j th layer of the input atmospheric column, respectively (see Section 3.2 for details on scaling). The surrogate models took input of scalar inputs β , where $\beta_{\text{SW}} = (\mu, p_0, T_0, \rho_0, \alpha_{\text{SW}})$ for the short-wave surrogate model, and $\beta_{\text{LW}} = (p_0, T_0, \rho_0, \alpha_{\text{LW}})$ for the long-wave surrogate model, where μ and α denote the cosine of the solar zenith angle and surface albedo, respectively. Vector inputs are passed through a GRU layer starting from the top atmospheric layer (layer 48) and moving downwards to the ground layer (layer 0). Scalar inputs are concatenated with the output of the downward GRU, and inputted into a dense layer, which then serves as the first input to an upward-moving GRU layer. The outputs of the two GRU layers are then concatenated and passed through a dense layer, which produces the model outputs of shape $(n_{\text{levels}}, n_{\text{outputs}}) = (50, 2)$, corresponding to scaled upwards and downwards flux for each atmospheric level. Rectified linear unit (ReLU, Agarap 2019) activation functions were used for both GRU layers and Dense_1 , while a sigmoid activation function was used for the $\text{Dense}_{\text{out}}$ layer.

Table 1. *Surrogate model inputs:* Surrogate models for both the long-wave and short-wave schemas took scaled vector inputs $\mathbf{X}_i = (\tilde{p}_i, \tilde{T}_i, \tilde{\rho}_i)$ for the i th layer where \tilde{p}_i , \tilde{T}_i , and $\tilde{\rho}_i$ are scaled pressure, scaled temperature, and scaled density of the i th layer of the input atmospheric column, respectively (see Section 3.2 for details on scaling). The surrogate models took input of scalar inputs β , where $\beta_{\text{SW}} = (\mu, p_0, T_0, \rho_0, \alpha_{\text{SW}})$ for the short-wave surrogate model, and $\beta_{\text{LW}} = (p_0, T_0, \rho_0, \alpha_{\text{LW}})$ for the long-wave surrogate model, where μ and α represent the cosine of the solar zenith angle and surface albedo, respectively.

Surrogate schema	Short-wave	Long-wave
Vector inputs	$\mathbf{X}_i = (\tilde{p}_i, \tilde{T}_i, \tilde{\rho}_i)$	$\mathbf{X}_i = (\tilde{p}_i, \tilde{T}_i, \tilde{\rho}_i)$
Scalar inputs	$\beta_{\text{SW}} = (\mu, p_0, T_0, \rho_0, \alpha_{\text{SW}})$	$\beta_{\text{LW}} = (p_0, T_0, \rho_0, \alpha_{\text{LW}})$

with the total loss defined as the sum over all test samples:

$$\text{loss} = \frac{1}{2} \sum_i \sum_k \text{MPE}_{i,k}. \quad (13)$$

3.4.2 Hyperparameter tuning

Multiple models were trained corresponding to different hyperparameter values. Number of neurons of all RNN layers was varied across the range of values [16, 32, 64, 128] for both surrogate models. The best candidate models were chosen as having 128

neurons per RNN layer for the short-wave surrogate model, and 32 neurons per RNN layer for the long-wave surrogate model.

3.5 Performance analysis

Below, we detail the metrics used to analyse the performance of the surrogate models on the test set. In the results (Section 4), different aggregations of absolute error (equation 15 for raw model outputs and equation 14 for post-processed model outputs) are used to investigate the performance of both the long-wave and short-wave surrogate models.

$$\text{Absolute Error} \equiv \text{AE}_{i,j,k} = |\hat{y}_{i,j,k} - \tilde{y}_{i,j,k}|, \quad (14)$$

where $\hat{y}_{i,j,k}$ are the post-processed model predictions, and $\tilde{y}_{i,j,k}$ are the unscaled target variables.

$$\widehat{\text{AE}}_{i,j,k} = |\hat{u}_{i,j,k} - \tilde{u}_{i,j,k}|, \quad (15)$$

where $\hat{u}_{i,j,k}$ are the raw model predictions, and $\tilde{u}_{i,j,k}$ are target variables, which have been scaled to lie in the interval [0, 1] using the scaling methods detailed in Section 3.2.

Column-aggregated error quantities are defined as follows:

$$\text{CAE}_{i,k} = \sum_{j=0}^{n_{\text{levels}}-1} |\hat{y}_{i,j,k} - \tilde{y}_{i,j,k}|, \quad (16)$$

Table 2. This table summarizes the MAE (MAE_k , defined in equation 21) across the four target variables, and relative to the mean values \bar{F}_k of these four variables (as defined in equation 20), across the test set.

Regime	Stream	Mean flux \bar{F}_k (W m^{-2})	MAE_k (W m^{-2})	MAE_k/\bar{F}_k (per cent)
Long-wave	Upwelling	4211.0	18.8	0.45
	Downwelling	4139.3	16.9	0.41
Short-wave	Upwelling	577.9	6.4	1.11
	Downwelling	707.8	7.7	1.09

$$\widehat{\text{CAE}}_{i,k} = \sum_{j=0}^{n_{\text{levels}}-1} |\hat{u}_{i,j,k} - \tilde{u}_{i,j,k}|, \quad (17)$$

where n_{levels} is the number of atmospheric levels.

Error quantities averaged across samples per altitude level are defined as follows:

$$\text{MAE}_{j,k} = \frac{\sum_{i=0}^{N-1} |\hat{y}_{i,j,k} - \tilde{y}_{i,j,k}|}{N}, \quad (18)$$

$$\widehat{\text{MAE}}_{j,k} = \frac{\sum_{i=0}^{N-1} |\hat{u}_{i,j,k} - \tilde{u}_{i,j,k}|}{N}, \quad (19)$$

where N is the number of test samples.

Mean flux for the k th target variable is defined as

$$\text{Mean Flux}_k \equiv \bar{F}_k = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{n_{\text{levels}}-1} \tilde{y}_{i,j,k}}{N}, \quad (20)$$

and the MAE for the k th target variable aggregated across all altitude levels is calculated as

$$\text{MAE}_k = \sum_{j=0}^{n_{\text{levels}}-1} \text{MAE}_{j,k}. \quad (21)$$

4 RESULTS AND DISCUSSION

4.1 Model performance on test set

Table 2 summarizes the MAE (equation 21) across the four target variables, and relative to the mean values of these four variables, across the test set. These MAE values are in line with those achieved using similar surrogate modelling methods for RT within the Earth's atmosphere, with Ukkonen (2022) quoting MAE for short-wave fluxes of around 1 per cent or less. For both the long-wave and short-wave regimes, the MAE is higher for the upwelling fluxes as compared to downwelling fluxes: this is expected as physical computation of the upwelling flux depends on the computation of the downwelling flux, thus meaning that this is a more complicated mapping to emulate. The percentage errors for the short-wave targets are a factor of 2–3 greater than those for the long-wave targets: this is to be expected as the magnitude of the short-wave targets is smaller, and the supervised learning task set for the short-wave model in this work is a more complex mapping from inputs to outputs as compared to that for the long-wave model.

In the following sections, we visualize and interrogate the variation of model prediction errors with altitude and with scalar variables: $\cos \mu$ for both long-wave and short-wave model predictions, and surface temperature for long-wave model predictions only. For completeness, further plots of error as a function of the remaining input scalar variables (surface temperature, surface pressure, and surface gas density) are contained in Appendix F.

4.1.1 Short-wave surrogate model

Variation of error versus altitude: Fig. 3 displays the average absolute error of predictions at different altitude levels across test samples. Below around 65 km, the average error increases roughly in proportion to the average target fluxes. Above this altitude, the average error decreases for both target variables, even though the average values of these variables continue to rise with altitude. This change in the trend of the average error occurs around the top of the cloud deck.

Potential explanations as to why there is lower error in predicting targets above 65 km are as follows: above the top of the cloud deck, there is less complexity in mapping from input variables to output fluxes, and so this can be naively assumed to be an easier task to learn; there is also tighter variance in the target variables within the test set above this altitude threshold.

Figs 3(A) and (B) display MEs of less than 3 per cent across all altitude levels, averaged across test samples in the [5, 95] percentile interval of column-aggregated errors. This accuracy falls within an acceptably small margin of error.

Variation of error versus $\cos \mu$: Fig. 5 displays the distribution in test errors as a function of $\cos \mu$ of the test column. Data in these plots have been binned to more simply display the spread of errors for a given $\cos \mu$ interval. For values of $\cos \mu$ close to 1, there is a narrower distribution of error, with test samples being predicted reliably more accurately as compared to test samples corresponding to lower values of $\cos \mu$.

This variation in test error distribution as a function of $\cos \mu$ may be attributable to the approximations used during target pre-processing when training the model (see Section 3.2); if indeed this is the case, then there is scope to mitigate against this by refining data pre-processing methods when producing future iterations of this surrogate model. This error variation with $\cos \mu$ may otherwise be due to the model not exactly capturing the complexities in how $\cos \mu$ is used in mapping inputs to outputs, which may be an acceptable and necessary trade-off in using a surrogate model for the purposes of model speed-up.

4.1.2 Long-wave surrogate model

Variation of error versus altitude: Fig. 4 displays the absolute error of predictions averaged across altitude levels across test samples. A similar trend can be seen in these plots as compared to the trends described in Section 4.1.1: average magnitude of error roughly follows the same trend as average magnitude of the target variables, except for between 40 and 70 km altitude (roughly in the interval of the main cloud deck) whereby the error rises significantly at the top of the cloud deck and decreases going deeper into the cloud deck, for both target variables.

Below the cloud deck, it can be seen from Figs 4(A) and (B) that the ME in target predictions is less than 5 per cent of mean target magnitude, for the given percentile interval of test samples. Above

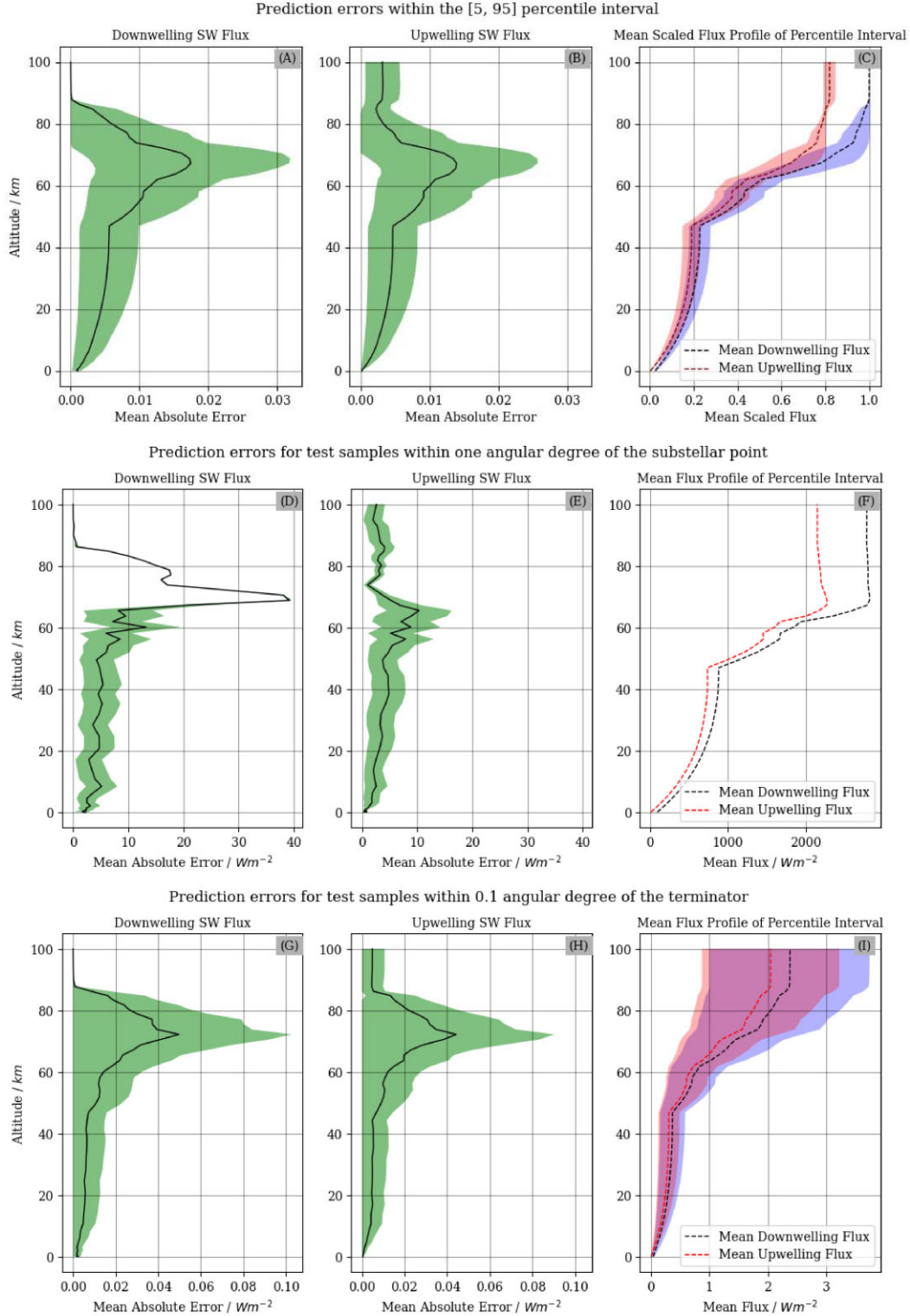


Figure 3. Plots A and B display the $\widehat{MAE}_{j,k}$ (defined in equation 19) for test samples that have column-aggregated absolute error ($\widehat{CAE}_{i,k}$, defined in equation 17) in the [5, 95] percentile interval (that is, test samples with errors falling in the smallest 5 per cent and largest 5 per cent of the test set have been discarded from this aggregate statistic, in order to better illustrate typical model performance). Plots D and E display the $MAE_{j,k}$ (defined in equation 18) of surrogate model predictions across the two short-wave target variables, for test samples that lie within 1 angular degree of the substellar point. Plots G and H display the $MAE_{j,k}$ of surrogate model predictions across the two short-wave target variables, for test samples that lie within 0.1 angular degree of the day-night terminator. Plot C displays the scaled target short-wave flux profiles averaged across the test samples with $\widehat{CAE}_{i,k}$ within the chosen percentile interval. Plot F displays the target short-wave flux profiles averaged across samples that lie within 1 angular degree of the substellar point. Plot I displays the target short-wave flux profiles averaged across samples that lie within 1.1 angular degree of the terminator. The shaded regions in all plots represent the interval $[\max(0, \theta_j - \sigma_j), \theta_j + \sigma_j]$, where θ_j and σ_j denote the mean and standard deviation of the plotted quantities for the j th altitude level.

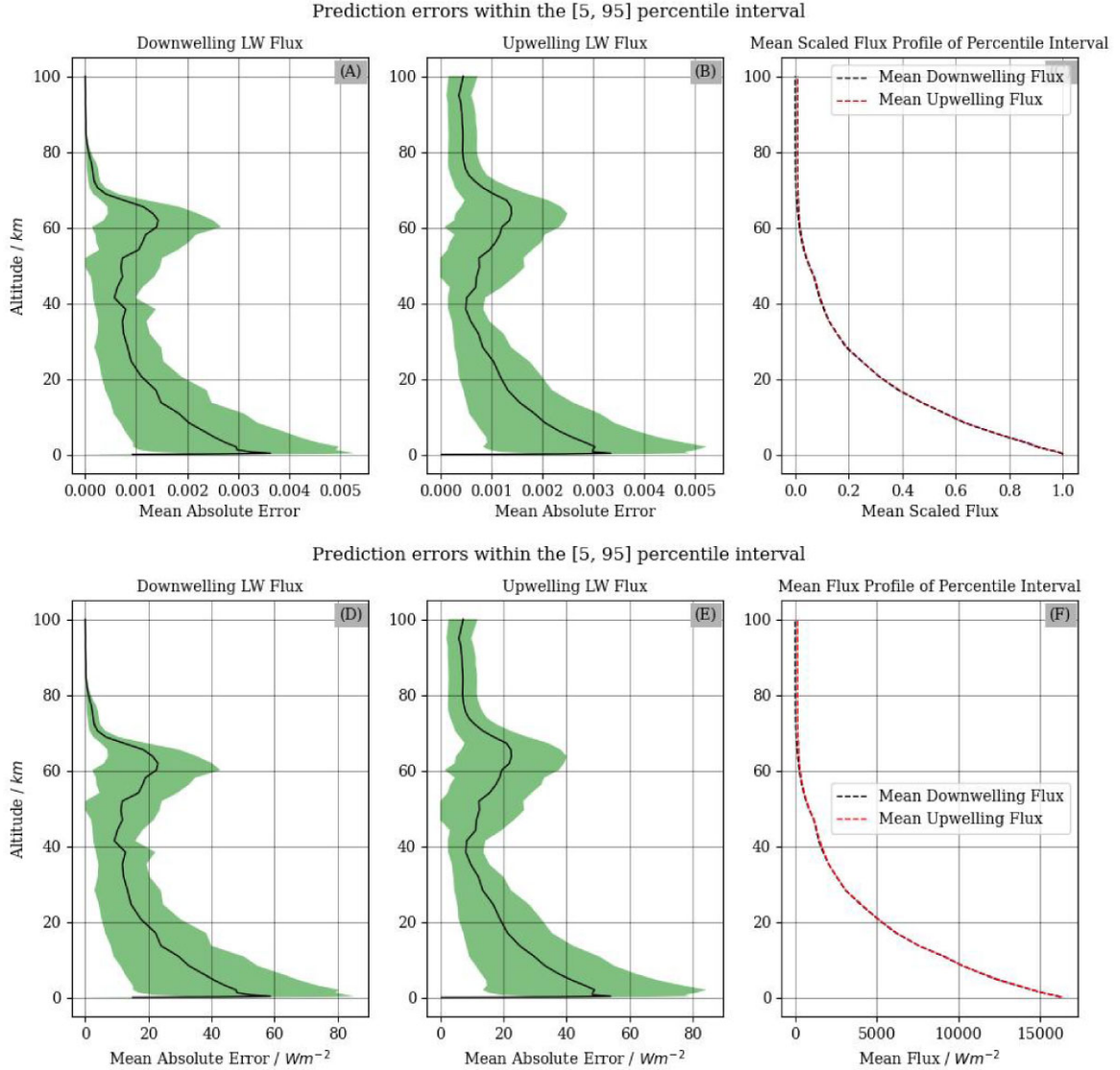


Figure 4. Plots A and B display the $\widehat{MAE}_{j,k}$ (defined in equation 19) and plots D and E display the $MAE_{j,k}$ (defined in equation 18) of surrogate model predictions across the two long-wave target variables, for test samples that have column-aggregated absolute error ($\widehat{CAE}_{i,k}$, $CAE_{i,k}$) in the [5, 95] percentile interval (that is, test samples with errors falling in the smallest 5 per cent and largest 5 per cent of the test set have been discarded from this aggregate statistic, in order to better illustrate typical model performance). Plot C displays the scaled target long-wave flux profiles averaged across the test samples with $\widehat{CAE}_{i,k}$ within the chosen percentile interval. Plot F displays the target flux long-wave profiles averaged across the test samples with $CAE_{i,k}$ within the chosen percentile interval (not necessarily the same subset displayed in plot C). The shaded regions in all plots represent the interval $[\max(0, \theta_j - \sigma_j), \theta_j + \sigma_j]$ where θ_j and σ_j denote the mean and standard deviation of the plotted quantities for the j th altitude level.

the cloud deck, the long-wave fluxes tend to zero; so, though the percentage errors increase for increasing altitude, these accuracies still fall within the reasonable margin of error for the model.

Variation of error versus $\cos \mu$ /surface temperature: Variations in test error versus surface temperature (displayed in Fig. 5) appear to follow the same trend as compared to residuals in the scaling factor used in pre-processing (displayed in Fig. 1). This is promising as it may indicate that scaling residuals are the limiting factor in model accuracy, and these scaling residuals were initially deemed as acceptably small for the purpose of this work. Considering variation in test error versus cosine of the solar zenith angle ($\cos \mu$) (also displayed in Fig. 5), there is not a discernible trend in the error variation across $\cos \mu$, though it appears that the error distribution is most narrow for $\cos \mu$ approaching 1. This may reflect the variability in target flux profiles in both the train and test set for different $\cos \mu$

bins, as can be seen plotted in Appendix G, whereby targets for higher values of $\cos \mu$ (i.e. approaching the substellar point) are more constrained, and thus are intuitively easier to predict.

4.2 Model performance in simulation

To evaluate the performance of our new surrogate model on 3D simulations of Venus's atmosphere, we have run two OASIS simulations: one using OASIS-RT for the RT scheme and the other using the new surrogate model presented in this work. In order to efficiently run the 3D simulation with OASIS-RT, as detailed in Section 2.1.1, we updated the radiative fluxes for solar radiation every 2880 steps, and the thermal radiation fluxes were updated every 320 steps. Both simulations run for 5 Venus solar days,

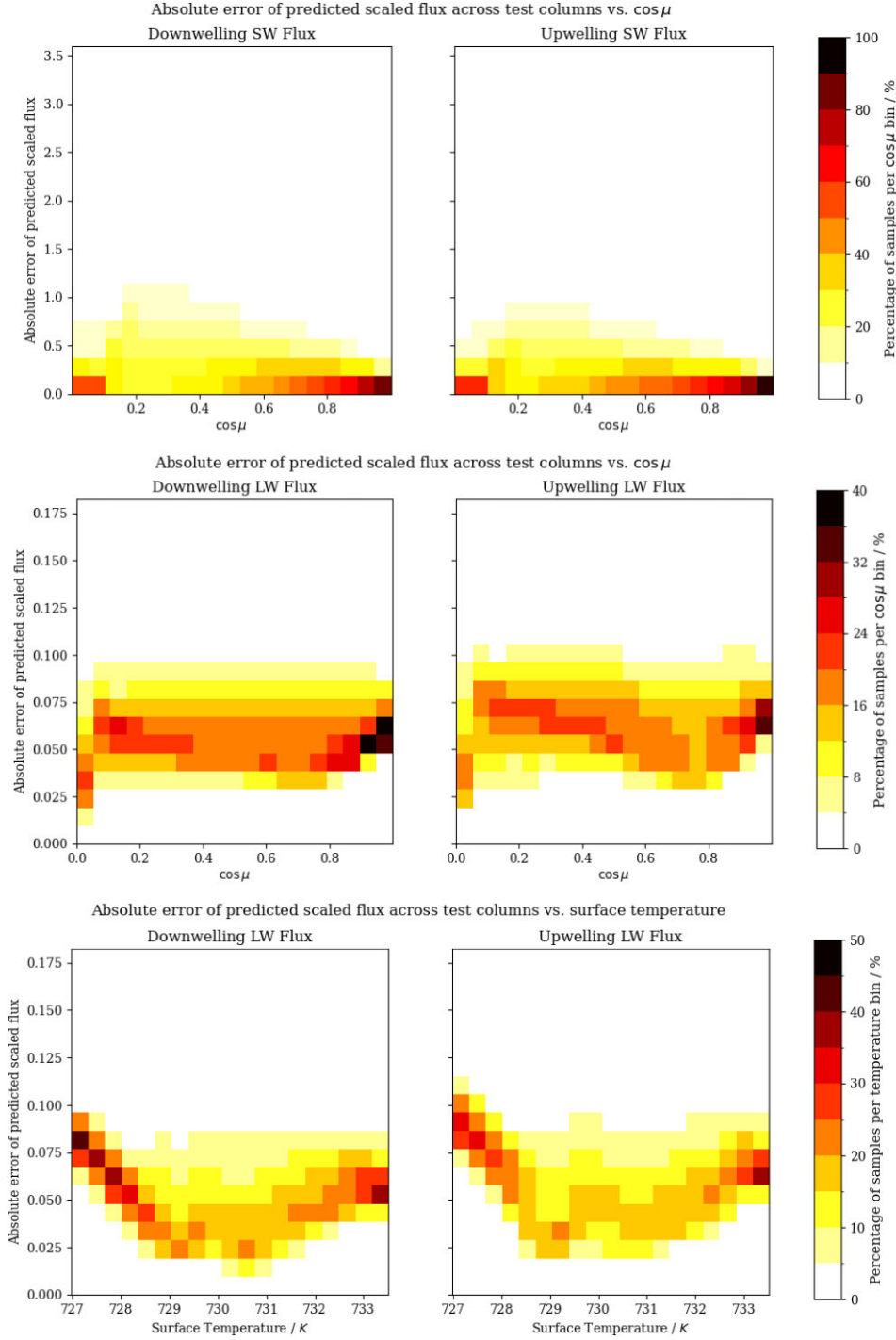


Figure 5. The above plots display the variation in $\widehat{\text{CAE}}_{i,k}$ (as defined in equation 17) across the long-wave and short-wave schemas, as functions of the cosine of the solar zenith angle ($\cos \mu$) and of surface temperature T_0 (for the long-wave schema only). The extent of the plots on the y-axis covers the full range of the error quantity on the y-axis of each plot, with white spaces corresponding to values falling in the lowest value bin. Top row: The above figure displays the error distribution of test samples per $\cos \mu$ bin, across both short-wave target variables. Test samples have been binned into 20 bins of $\cos \mu$, into 20 bins of $\widehat{\text{CAE}}_{i,k}$. Bins have been normalized to percentages by the total number of test columns in each $\cos \mu$ bin. Middle and bottom rows: The above figure displays the error distribution of test samples per $\cos \mu$ bin, across both long-wave target variables as a function of $\cos \mu$ (middle row) and T_0 (bottom row). $\cos \mu$, T_0 , and $\widehat{\text{CAE}}_{i,k}$ have been divided linearly into 20 bins; the plots display the proportion of test samples falling into each error bin relative to total samples in a given $\cos \mu$ or T_0 bin.

each of which is approximately 117 Earth days. Fig. 6 displays the temperature of the simulation using OASIS-RT averaged over the last simulated Venus day in the left plot, and the percentage difference in this quantity between the two simulations in the right

plot. Beneath the bottom of the cloud deck, the percentage difference between the time-averaged temperature profiles produced by the two simulations is below 1.5 percent; in the interval of the cloud deck, below 2.7 percent; and above the cloud deck, below 4.0

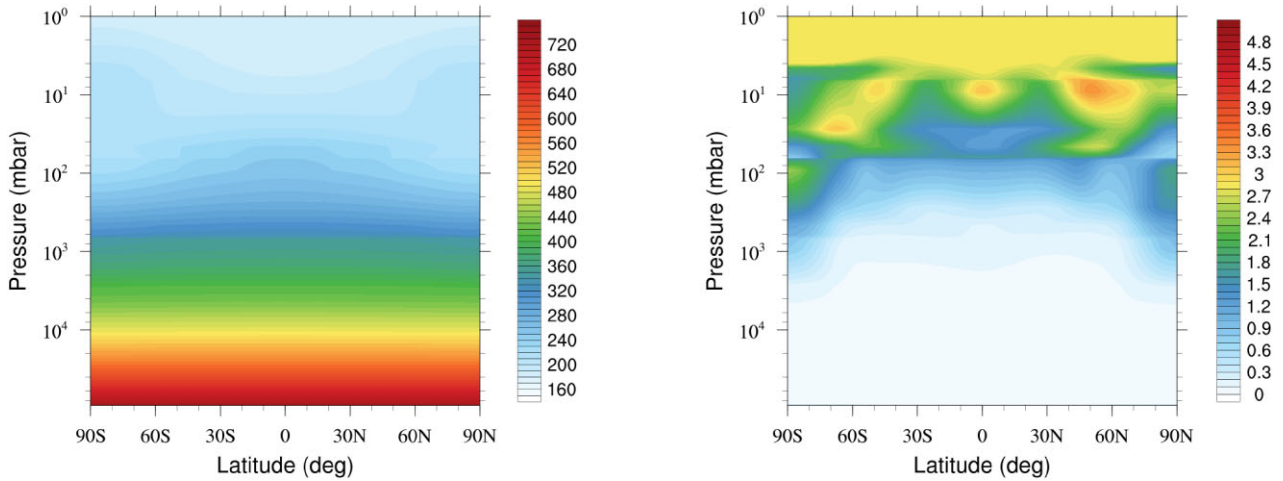


Figure 6. Left: The above figure displays the simulated atmospheric temperature profile (in units of K) of Venus averaged over the final Venus solar day of the simulation. The simulation was run using OASIS and OASIS-RT for 5 Venus solar days (each Venus solar day is equivalent to approximately 117 Earth days). Right: The above figure displays the percentage difference in simulated atmospheric temperature profiles of Venus averaged across the final Venus day of the simulation, for two simulations using (1) OASIS-RT and (2) the surrogate models presented in this work, to model the RT. **Both:** In both plots, the main cloud deck extends from approximately 10^2 to 2×10^3 mbar.

per cent. These deviations in the final temperature profiles fall within the range of uncertainties in the measurements. This is reasonable due to two main factors: first, there will be inherent discrepancies between simulated and real temperatures arising from assumptions made in the physical model, and secondly, deviations will also naturally arise between the physical model and real temperature profiles due to the spatial resolution of the simulation. Also, we expected differences to arise due to the frequency at which the radiative fluxes are updated. In the case of the surrogate model, it is possible to update them at every physical time-step. Fig. B1 displays the contrast of temperature profiles of simulations using OASIS-RT and the surrogate models, whereby the frequency of updating radiative fluxes in the latter matches that of the former (every 2880 steps for the short-wave regime and every 320 steps for the long-wave regime). There is negligible discrepancy between the final temperature profile of Venus as simulated using the surrogate-integrated schema for both frequencies of executing the RT update. A large difference between the plots was not expected because, in both cases, the RT updates are performed at intervals shorter than the radiative time-scale. This means that similar results would occur if the simulation were executed using full RT updates at every time-step, as compared to the simulation executed using RT updates at every 2880/320 time-steps for SW/LW, respectively. Despite this small difference in the resulting temperature profile corresponding to the different update frequencies, our future simulations will avoid using large step updates since these compromise the accuracy of the atmospheric physics, computed by the dynamical core, and would also compromise the future implementation of radiatively active clouds in the 3D simulations.

4.3 Simulation runtime

To benchmark the speed-up achieved using the surrogate RT schema, we run four simulations of the Venus atmosphere. All 4 of these simulations span 1000 time-steps, with each time-step corresponding to an increment of 15 s. Simulations were run on one NVIDIA Tesla V100 GPU. The simulations were as follows:

- (i) Simulation with no RT update at any time-step: 95 s runtime.

- (ii) Simulation with RT executed in the first time-step, and then updated every 2880 time-steps (equivalent to a time-step of 12 h) for the short-wave regime, and every 320 time-steps (equivalent to a time-step of 1 h 20 min) for the long-wave regime: 195 s.

- (iii) Simulation with RT updated every time-step: 27 046 s.

- (iv) Simulation using the surrogate models to update the RT every time-step: 184 s.

Simulation (ii) uses the typical RT update frequency employed for using OASIS to simulate Venus: this satisfies a trade-off between sufficient temporal resolution of the RT, which must be modelled with a resolution smaller than the radiative time constant τ_R of the atmosphere ($\tau_R \approx 43$ h at the altitude in the atmosphere where it takes its smallest value),³ and overall simulation runtime. For clarity, in simulation (ii), the short-wave RT computation is executed once and the long-wave RT computation is executed three times. As mentioned in Section 4.2, less frequent RT updates limit the accuracy of modelling the atmospheric physics, and so a higher frequency of RT updates is preferred. Simulation (iii), which updates the RT at every 15 s time-step, illustrates how increasing the frequency of RT updates enormously increases the simulation runtime; here, by a factor of 92 compared to simulation (ii) with less frequent RT updates, and by a factor of 285 compared to simulation (i) with no RT update at all.

Comparing simulation (iv) to simulation (ii) with infrequent RT updates, we see a speed-up of ~ 6 per cent while achieving a higher temporal resolution of RT. Comparing simulation (iv) to simulation (iii), we see a factor of $147\times$ speed-up of the entire simulation runtime.

In addition to the reduction in simulation runtime, simulation (iv) is much more memory-efficient than simulations (i) and (ii), as the surrogate RT models do not require storing opacity cross-sections from the gas absorption or clouds. Furthermore, with the potential to update the radiative fluxes on every physical time-step now computationally feasible, our new approach allows for

³Radiative time constants for the Venus atmosphere can be seen in table 2 of Pollack & Young (1975).

the inclusion of dynamical cloud feedback at a small computational cost. This addresses one of the main limitations of current 3D Venus atmospheric models.

4.4 Limitations

The surrogate models produced in this work do not take explicit input of the structure of cloud and gas absorber constituents of the atmosphere being modelled, except for the input of gas density, $\rho_{i,j}$. This means that the learning objective for surrogate models in this work is to approximate the mapping from input thermodynamic column variables to output columnar flux profiles, *conditioned* on a specific cloud and absorber structure. Consequently, this means that the models produced in this work are only applicable for planets corresponding to the planetary parameters and cloud and absorber structure specific to Venus. This is a limitation in terms of *generalizability* of the surrogate-integrated GCM to other types of atmosphere.

5 CONCLUSIONS

This work introduces a surrogate model approach to replacing numerical simulations of short-wave and long-wave computations in a two-stream RT model, aimed at accelerating the GCM, OASIS. The results show a significant GCM speed-up by a factor of 147 GPU performance, with surrogate models for both long-wave and short-wave regimes achieving test set accuracies of approximately 99 per cent. Additionally, this approach replicates the temperature profile of the original Venus simulations averaged across a Venus solar day with differences of 4 per cent after 5 Venus solar days of simulation.

This work is significant in that it enables

- (i) $\sim 150\times$ faster simulations of planets with massive atmospheres that require complex RT schemes, such as the Venus atmosphere.
- (ii) Longer simulations with a much higher spatial resolution ($\sim 10\times$).
- (iii) Improved representation of the temperature evolution of short-term physical phenomena in the atmosphere. These can be atmospheric waves with time-scales shorter than the period at which the radiative fluxes are updated in the simulation. In the case of our Venus simulations, we can measure the temperature change of atmospheric waves with time-scales < 12 h.
- (iv) A model free of model tuning to optimize performance, such as the frequency of how the radiative fluxes are updated.
- (v) The inclusion in 3D simulations of cloud dynamical feedback or higher order, more complex radiative schemes with a small extra computational cost.

This achievement of faster and/or higher spatial resolution atmospheric simulations will facilitate better insight into the nature of the atmosphere of Venus, as well as benchmarking the utility and applicability of such modelling techniques for use in exoplanet science.

ACKNOWLEDGEMENTS

We thank Dr Ahmed Al-Refaie, Nikita Pond, and Max Hart for helpful conversations on integrating TENSORFLOW models within a C/C++ framework.

The authors acknowledge the use of the High Performance Computing facilities of University College London to carry out

this work, specifically the Hypatia cluster. This work used computing equipment funded by the Research Capital Investment Fund (RCIF) provided by UKRI, and partially funded by the UCL Cosmoparticle Initiative. This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 758892/ExoAI), and from the Science and Technology Facilities Council (STFC; grant no. ST/W00254X/1 and grant no. ST/W50788X/1). JMM acknowledges support from the Horizon Europe Guarantee Fund, grant EP/Z00330X/1.

DATA AVAILABILITY

The code for this project will be made publicly available at <https://github.com/ttahseen/oasis-rt-surrogate> on acceptance of this paper.

REFERENCES

- Abadi M. et al., 2016, preprint (arXiv:1603.04467)
 Agarap A. F., 2019, preprint (arXiv:1803.08375)
 Changeat Q., Al-Refaie A., 2020, *ApJ*, 898, 155
 Cho K., van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y., 2014, preprint (arXiv:1406.1078)
 Chubb K. L., Min M., 2022, *A&A*, 665, A2
 Crisp D., 1986, *Icarus*, 67, 484
 Deitrick R., Mendonça J. a. M., Schroffenegger U., Grimm S. L., Tsai S.-M., Heng K., 2020, *ApJS*, 248, 30
 Eymet V., Fournier R., Dufresne J.-L., Lebonnois S., Hourdin F., Bullock M. A., 2009, *J. Geophys. Res.: Planets*, 114
 Feng Y. K., Line M. R., Fortney J. J., 2020, *AJ*, 160, 137
 Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
 Géron A., 2019, *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*, 2nd edn. O'Reilly Media, Inc., Sebastopol, California, United States
 Goodfellow I., Bengio Y., Courville A., 2016, *Deep Learning*. MIT Press, Cambridge, Massachusetts, United States, <http://www.deeplearningbook.org> (accessed 20 Oct 2024)
 Himes M. D. et al., 2022, *Planet. Sci. J.*, 3, 91
 Himes M. D., Harrington J., Baydin A. G., 2023, *Planet. Sci. J.*, 4, 74
 Irwin P. G. J., Parmentier V., Taylor J., Barstow J., Aigrain S., Lee E. K. H., Garland R., 2020, *MNRAS*, 493, 106
 Joseph J. H., Wiscombe W. J., Weinman J. A., 1976, *J. Atmos. Sci.*, 33, 2452
 Kingma D. P., Ba J., 2017, preprint (arXiv:1412.6980)
 Knollenberg R. G., Hunten D. M., 1980, *J. Geophys. Res.: Space Phys.*, 85, 8039
 Lacis A. A., Oinas V., 1991, *J. Geophys. Res.: Atmos.*, 96, 9027
 Lebonnois S., Sugimoto N., Gilli G., 2016, *Icarus*, 278, 38
 Lee C., Richardson M. I., 2012, *Icarus*, 221, 1173
 Liu Q., Weng F., 2006, *J. Atmos. Sci.*, 63, 3459
 Mendonça J. a. M., Buchhave L. A., 2020, *MNRAS*, 496, 3512
 Mendonça J. M., Read P. L., 2016, *Planet. Space Sci.*, 134, 1
 Mendonça J., Read P., Wilson C., Lee C., 2014, *Planet. Space Sci.*, 105
 Mendonça J. M., Read P. L., Wilson C. F., Lee C., 2015, *Planet. Space Sci.*, 105, 80
 Mendonça J. a. M., Grimm S. L., Grosheintz L., Heng K., 2016, *ApJ*, 829, 115
 Mukkavilli S. K. et al., 2023, preprint (arXiv:2309.10808)
 Nixon M. C., Madhusudhan N., 2022, *ApJ*, 935, 73
 Pluriel W., Zingales T., Leconte J., Parmentier V., 2020, *A&A*, 636, A66
 Pollack J. B., Young R., 1975, *Icarus*, Vol. 34, Calculations of the Radiative and Dynamical State of the Venus Atmosphere. p. 28–45
 Prince S. J. D., 2023, *Understanding Deep Learning*. MIT Press, Cambridge, Massachusetts, United States
 Sagan C., 1962, in Deutsch A. J., Klemperer W. B., eds, *Space Age Astronomy*. Academic Press, New York, p. 430

- Schneider A. D., Mollière P., Louppe G., Carone L., Jørgensen U. G., Decin L., Helling C., 2024, *A&A*, 682, A79
- Tinetti G., Eccleston P., Lueftinger T., Salvignol J.-C., Fahmy S., Alves de Oliveira C., 2022, in *16th Europlanet Science Congress 2022*. Vol. 16, Ariel: Enabling planetary science across light-years, Granada, Spain, EPSC2022–1114, <https://www.epsc2022.eu/>
- Tomasko M. G., Dooze L. R., Smith P. H., Odell A. P., 1980, *J. Geophys. Res.: Space Phys.*, 85, 8167
- Ukkonen P., 2022, *J. Adv. Model. Earth Syst.*, 14, e2021MS002875
- Unlu E. B., Forestano R. T., Matchev K. T., Matcheva K., 2023, preprint (arXiv:2310.10521)
- Yao Y., Zhong X., Zheng Y., Wang Z., 2023, *J. Adv. Model. Earth Syst.*, 15, e2022MS003445

APPENDIX A: MODEL ARCHITECTURE

A1 Feedforward neural networks

A *neural network* is a type of function f that, given an input feature vector \mathcal{X} , can be implemented as combination of matrix transformations, $\{\mathcal{M}_k \in \mathbb{R}^{m_k \times n_k}\}$, and non-linear transformations, $\{\alpha_k\}$, where each α_k is a dimension-preserving mapping, i.e. $\alpha_k : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$. A simple feedforward neural network implements the transformation⁴

$$\mathcal{Y} = \left(\bigcirc_{k=0}^i (\alpha_k \circ \mathcal{M}_k) \right) (\mathcal{X}). \quad (\text{A1})$$

The matrices $\{\mathcal{M}_k\}$ are fitted as an optimization task of a chosen objective function \mathcal{L} given a set of examples $\mathcal{S} = \{\hat{\mathcal{Y}}, \hat{\mathcal{X}}\}$ to constrain the network to approximate some given mapping \tilde{f} captured by \mathcal{S} . Neural networks as described above are universal function approximators, and so constitute an appropriate function space within which to seek an approximate function f for the target function \tilde{f} . The matrices $\{\mathcal{M}_k\}$ are often referred to as network *weights* or network *parameters*; the dimension of each \mathcal{M}_k is free in one dimension [except for \mathcal{M}_i for which the dimension is fully constrained by the dimension of \mathcal{Y} and the dimension of $\left(\bigcirc_{k=0}^{i-1} (\alpha_k \circ \mathcal{M}_k) \right) (\mathcal{X})$]. This free dimension m_k of each \mathcal{M}_k we can

ascribe as the number of *neurons* of the *layer* \mathcal{M}_k , to be consistent with machine learning terminology. For further details about the mathematics and implementation of neural networks, see chapters 3 and 4 of Prince (2023).

A2 Recurrent neural networks

RNNs are function spaces defined using similar concepts to simple feedforward neural networks, designed to incorporate the structural dependence of sequential input feature vectors. Given a set of input feature vectors $\{X_k\} \forall k \in [0, N]$, with some relation between consecutive feature vectors X_k and X_{k+1} , a simple RNN operates as follows:

$$r_{k+1} = \alpha_0(\mathcal{M}_{\text{in}} \mathcal{X}_{k+1} + \mathcal{M} r_k), \quad (\text{A2})$$

$$\mathcal{Y}_{k+1} = \alpha_1(\mathcal{M}_{\text{out}} r_{k+1}), \quad (\text{A3})$$

where r_k is defined as the *hidden state* of the network for the k th input. The above formulation of the simple recurrent network illustrates how information from preceding input feature vectors contributes to the function output corresponding to the k th input.

In practice, the RNN implementation is more complicated than the two equations above, but these equations capture the core operation of the RNN. For an in-depth treatment of the mathematics and implementation pragmatics of RNNs, see Géron (2019).

APPENDIX B: RESULTS

Below, we display the contrast of temperature profiles of simulations using OASIS-RT and the surrogate models, whereby the frequency of updating radiative fluxes in the latter matches that of the former (every 2880 steps for the short-wave regime and every 320 steps for the long-wave regime). There is negligible discrepancy between the final temperature profile of Venus as simulated using the surrogate-integrated schema for both frequencies of executing the RT update.

⁴Below, $\bigcirc_{k=0}^i (\alpha_k \circ \mathcal{M}_k)$ represents the composition of functions $(\alpha_k \circ \mathcal{M}_k)$ from $k = 0$ to $k = i$.

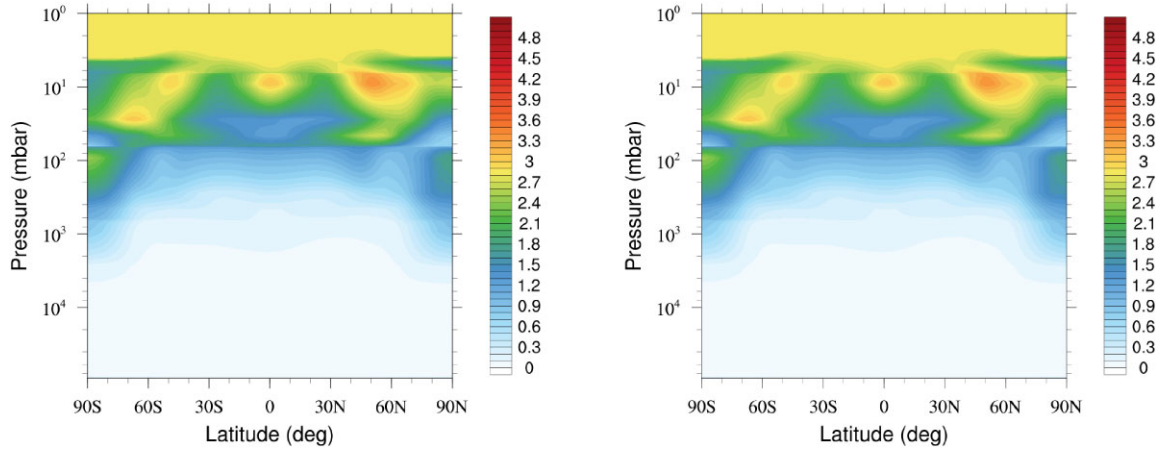


Figure B1. Both: The above figures display the percentage difference in simulated atmospheric temperature profiles of Venus averaged across the final Venus day of the simulation, for two simulations using (1) OASIS-RT and (2) the surrogate models presented in this work, to model the RT. Left: This plot was generated by updating the short-wave radiative fluxes at every 2880 steps, and the long-wave radiative fluxes at every 320 steps, for the simulation with the surrogate models. Right: This plot was generated by updating the short-wave radiative fluxes and long-wave radiative fluxes at every time-step, for the simulation with the surrogate models.

APPENDIX C: MODEL PARAMETERS

The tables below display the number of model parameters per layer of the two surrogate models produced in this work.

C1 Short-wave surrogate model

C2 Long-wave surrogate model

Table C1. This table displays the number of parameters across model layers for the short-wave surrogate model, as well as the output shapes of each layer. n denotes the number of atmospheric columns passed as input to the model; layers correspond to those displayed in Fig. 2.

Layer	Output shape	Number of parameters
Main inputs	$[(n, 49, 3)]$	0
Auxiliary inputs	$[(n, 5)]$	0
GRU _↓	$[(n, 49, 128), (n, 128)]$	51 072
Dense ₁	$[(n, 128)]$	17 152
GRU _↑	$[(n, 50, 128)]$	99 072
Dense _{out}	$[(n, 50, 2)]$	514
Total number of model parameters:		167 810

Table C2. This table displays the number of parameters across model layers for the long-wave surrogate model, as well as the output shapes of each layer. n denotes the number of atmospheric columns passed as input to the model; layers correspond to those displayed in Fig. 2.

Layer	Output shape	Number of parameters
Main inputs	$[(n, 49, 3)]$	0
Auxiliary inputs	$[(n, 4)]$	0
GRU _↓	$[(n, 49, 32), (n, 32)]$	3552
Dense ₁	$[(n, 32)]$	1184
GRU _↑	$[(n, 50, 32)]$	6336
Dense _{out}	$[(n, 50, 2)]$	130
Total number of model parameters:		11 202

APPENDIX D: FURTHER AGGREGATE STATISTICS OF MODEL TEST SET PERFORMANCE

The figures included in this section display the $\widehat{\text{MAE}}_{j,k}$ (defined in equation 19) of the entire test set for the long-wave (Fig. D1) and short-wave (Fig. D2) schemas. These figures are included for completeness, and are analogous to Figs 3(A)–(C) and 4(A)–(C), respectively, but for the entire test sets instead of the [5, 95] percentile interval subset of the test sets.

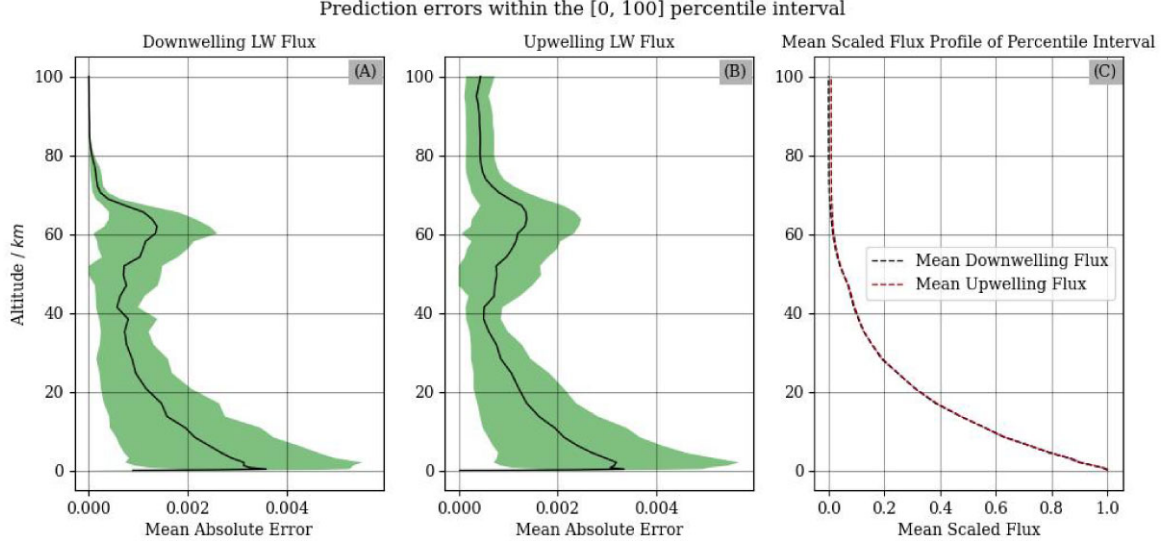


Figure D1. The above plots display the $\widehat{\text{MAE}}_{j,k}$ (defined in equation 19) of the long-wave surrogate model predictions. Plot C displays the target scaled flux profiles averaged across the test set. Plots A and B display the MAE of the scaled predictions and targets across the test set.

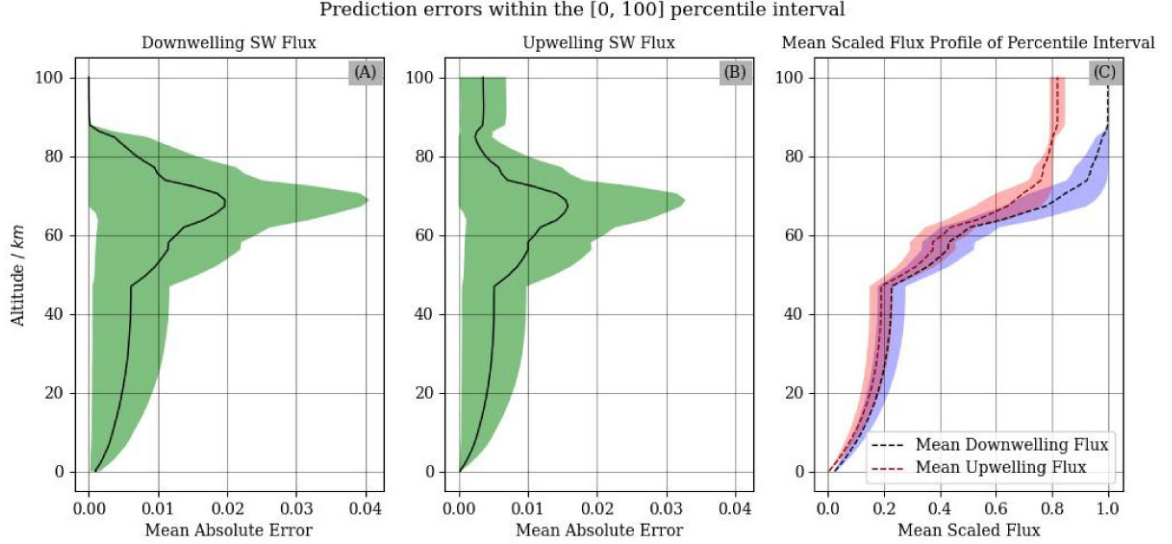


Figure D2. The above plots display the $\widehat{\text{MAE}}_{j,k}$ (defined in equation 19) of the short-wave surrogate model predictions. Plot C displays the target scaled flux profiles averaged across the test set. Plots A and B display the MAE of the scaled predictions and targets across the test set.

APPENDIX E: DISTRIBUTION OF SAMPLES

The plots in this section display the distribution of the entire data set across surface temperature and cosine of the solar zenith angle.

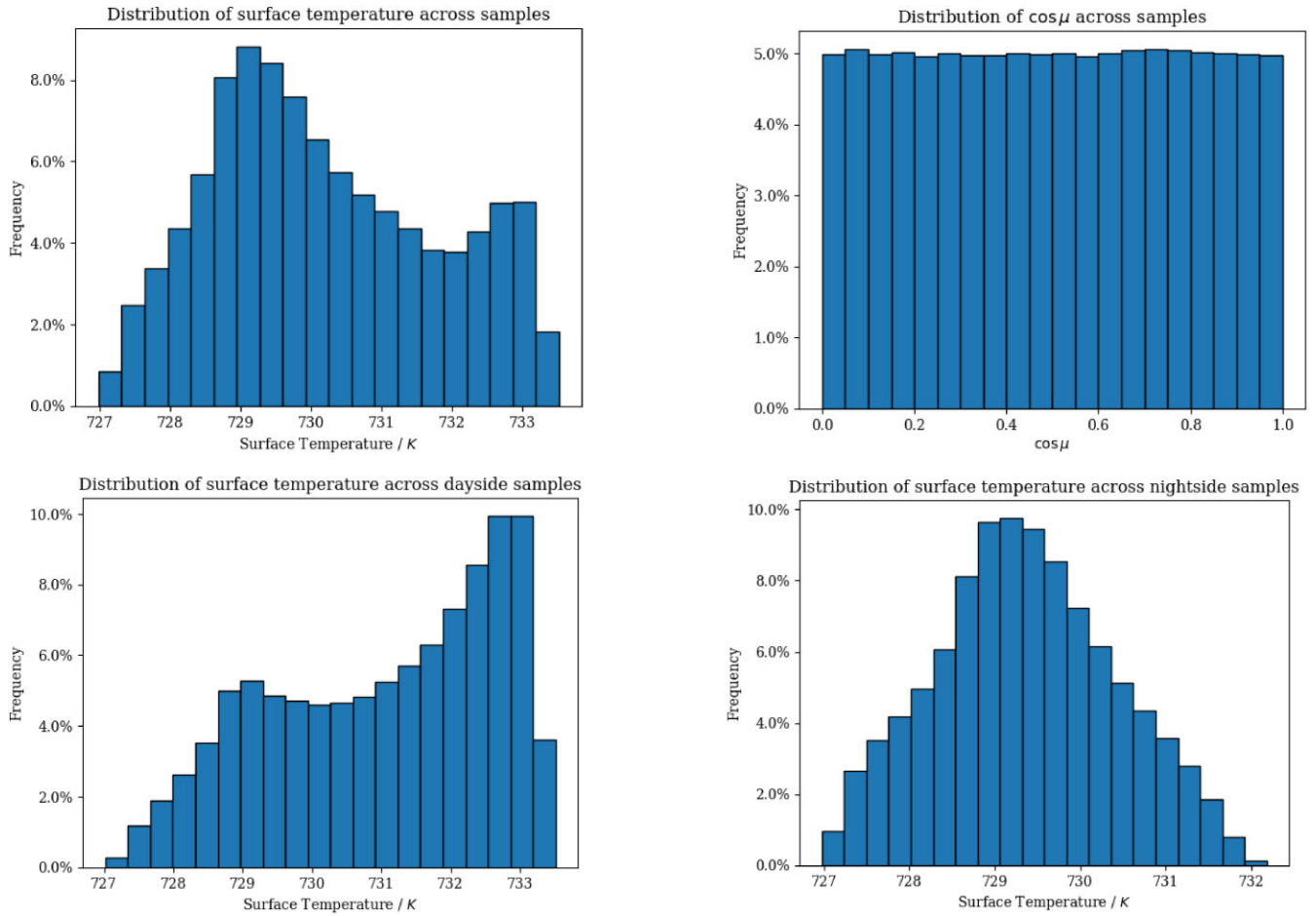


Figure E1. Top row: The figure displays the distributions of surface temperature (left tile) and cosine of the solar zenith angle, $\cos \mu$ (right tile), across the data set used for this work. Bottom row: The figure displays the distributions of surface temperature across samples falling within the dayside (left tile) and nightside (right tile).

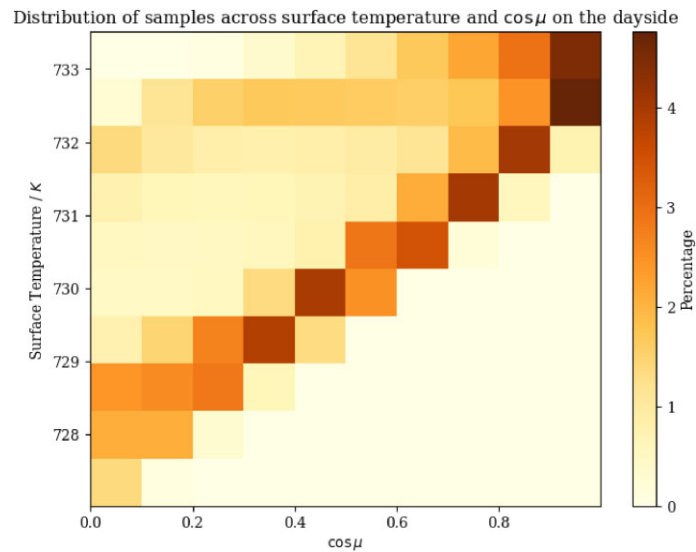


Figure E2. The figure displays the co-variation of surface temperature and $\cos \mu$ across dayside samples within the data set.

APPENDIX F: DISTRIBUTION OF ERRORS WITH SCALAR PARAMETERS

The plots in this section are analogous to those displayed in Fig. 5 and display the variation in $\widehat{\text{CAE}}_{i,k}$ (as defined in equation 17) of short-

wave (Fig. F1) and long-wave model predictions (Fig. F2) versus scalar model variables. These figures are included for completeness.

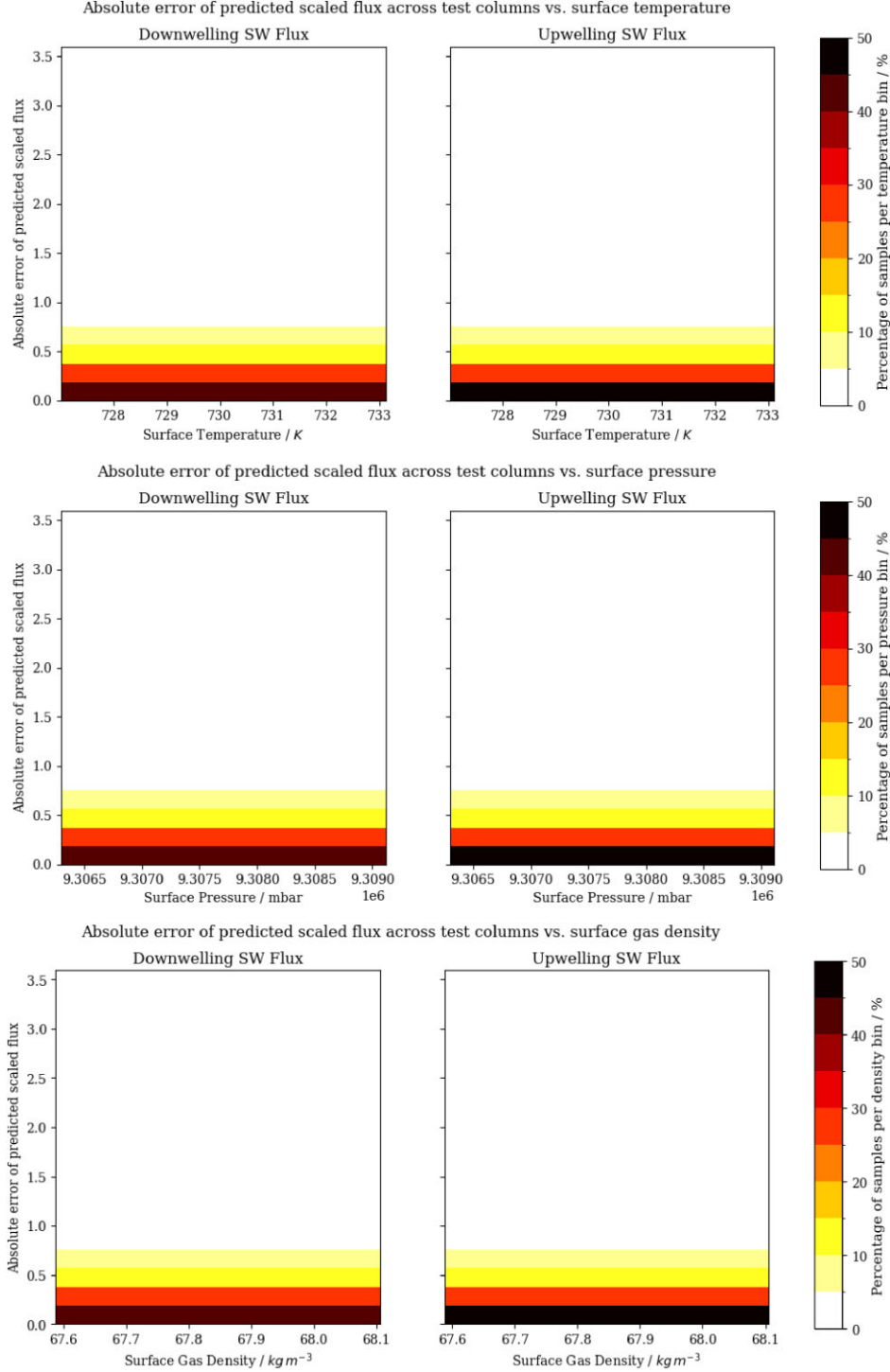


Figure F1. The figure displays variation in $\widehat{\text{CAE}}_{i,k}$ (as defined in equation 17) of short-wave model predictions as a function of surface temperature (top row), surface pressure (middle row), and surface gas density (bottom row). The extent of the plots on the y-axis covers the full range of the quantity on the y-axis of each plot, with white spaces corresponding to values falling in the lowest value bin. Bins have been normalized to percentages by the total number of test columns in each x -axis bin.

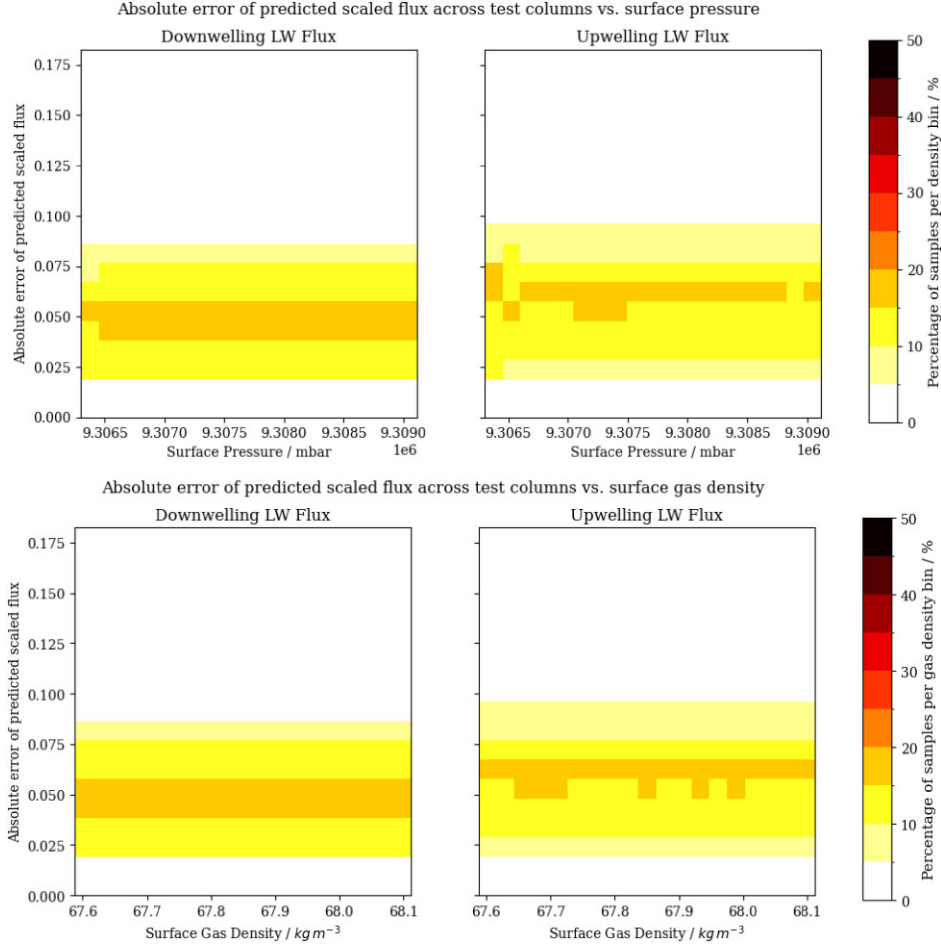


Figure F2. The figure displays variation in $\widehat{\text{CAE}}_{i,k}$ (as defined in equation 17) of long-wave model predictions as a function of surface pressure (top row) and surface gas density (bottom row). The extent of the plots on the y-axis covers the full range of the error quantity on the y-axis of each plot, with white spaces corresponding to values falling in the lowest value bin. Bins have been normalized to percentages by the total number of test columns in each x-axis bin.

APPENDIX G: VARIATION OF ERROR VERSUS $\cos \mu$ IN THE SHORT-WAVE REGIME

To investigate the amount of variability in targets versus $\cos \mu$, we define the following quantity:

$$\sigma_{i,k} = \sum_{j=0}^{n_{\text{levels}}-1} \sigma_{i,j,k}, \quad (\text{G1})$$

and plot this quantity for the test set, displayed in Fig. G1. The plots displayed in Fig. G1 illustrate a lower level of variation in the flux–altitude profiles with higher values of $\cos \mu$, and for the bin with the lowest values of $\cos \mu$ (i.e. at the terminator).

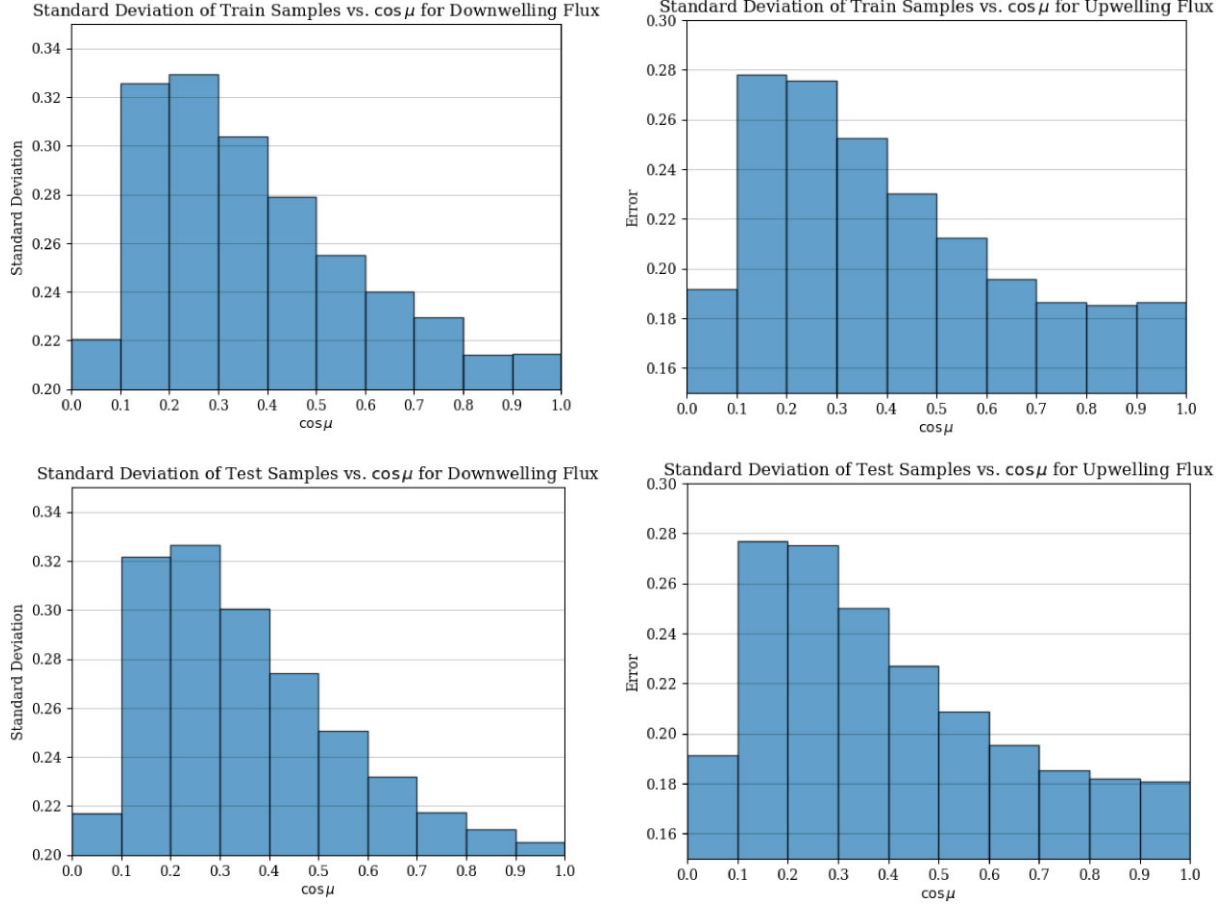


Figure G1. The above figure shows the distribution of column-aggregated error $\sigma_{i,k}$ (computed according to equation G1) per $\cos \mu$ bin per short-wave target for both the train set (top row) and test set (bottom row). These figures illustrate a lower level of variation in the flux–altitude profiles with higher values of $\cos \mu$, and for the bin with the lowest values of $\cos \mu$ (i.e. at the terminator).

This paper has been typeset from a \LaTeX file prepared by the author.