Characterizing the Atmospheric Dynamics of HD 209458b-like Hot Jupiters Using AI-driven Image Recognition/Categorization

F. Sainsbury-Martinez[1,2] , P. Tremblin[1] , M. Mancip[1], S. Donfack[1,3], E. Honore[1], and M. Bourenane[1]
[1] Université Paris-Saclay, UVSQ, CNRS, CEA, Maison de la Simulation, F-91191 Gif-sur-Yvette, France
[2] School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK
[3] IDRIS, CNRS, F-91403 Orsay, France

Abstract

In order to understand the results of recent observations of exoplanets, models have become increasingly complex. Unfortunately, this increases both the computational cost and output size of said models. We intend to explore if AI image recognition can alleviate this burden. We used DYNAMICO to run a series of HD 209458-like models with different orbital radii. Training data for a number of features of interest was selected from the initial outputs of these models. This was used to train a pair of multi-categorization convolutional neural networks (CNNs), which we applied to our outer-atmosphere-equilibrated models. The features detected by our CNNs revealed that our models fall into two regimes: models with shorter orbital radii exhibit significant global mixing that shapes the dynamics of the entire atmosphere, whereas models with longer orbital-radii exhibit negligible mixing except at mid-pressures. Here the initial nondetection of any trained features revealed a surprise: a nightside hot spot. Analysis suggests that this occurs when rotational influence is sufficiently weak that divergent flows from the dayside to the nightside dominate over rotational-driven transport, such as the equatorial jet. We suggest that image classification may play an important role in future, computational, atmospheric studies. However special care must be paid to the data feed into the model, from the color map, to training the CNN on features with enough breadth and complexity that the CNN can learn to detect them. However, by using preliminary studies and prior models, this should be more than achievable for future exascale calculations, allowing for a significant reduction in future workloads and computational resources.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (487); Exoplanet atmospheric structure (2310)

Supporting material: animations

1. Introduction

In recent years, exoplanetary atmospheres, particularly hot Jupiters, have become increasingly complex while also requiring increasingly long durations of time to reach equilibrium (as deeper regions of the atmosphere are considered).

Briefly, this increase in model complexity (and hence computational resource requirements) has come about as a result of a desire to more accurately model atmospheres (specifically their chemical composition and radiative dynamics) in order to better match/recover/understand the high spectral resolution observations made possible by Hubble (e.g., Charbonneau et al. 2000; Ranjan et al. 2011), JWST (e.g., Böker et al. 2022; Ferruit et al. 2022; Jakobsen et al. 2022; Pontoppidan et al. 2022), and other next-generation observing missions. As a result, while earlier (and longer timescale) studies ignored chemistry and/or modeled radiative dynamics in an equilibrium sense (e.g., Newtonian cooling; Showman et al. 2008; Rauscher & Menou 2010; Mayne et al. 2014a), this has changed recently thanks to next-generation global circulation models (GCMs). These models now include explicit nonequilibrium chemistry, including cloud formation and condensation, and multi-banded, correlated-k, radiative transport schemes (see, e.g., Amundsen et al. 2016; Lee et al. 2021;

Deitrick et al. 2022) that can simulate more accurate atmospheric feedback of evolving atmospheric dynamics. As a result, even when using high-performance, next-generation, GCMs (such as LFRic; Adams et al. 2019; Maynard et al. 2020), both the time required to run the model as well as the total output data generated has increased exponentially. This increase in data requirements is further exasperated by recent studies (such as Sainsbury-Martinez et al. 2019, 2021, 2023) that have shown that accurately modeling both the outer atmosphere dynamics, as well as global observable features, also requires the model to include a (close to) equilibrium deep atmosphere (which is computationally expensive due to the long dynamical timescales of deep atmospheric circulations).

While some of this burden can be lifted via runtime data preprocessing (e.g., the XIOS[4] library, which can not only interpolate complex GCM grids onto a simple long-lat grid at runtime, but also perform temporal and spatial averages), this still leaves a lot of data that needs to be analyzed, a time-consuming endeavor that only becomes more so when paired with the need to monitor ongoing models in order to either confirm their stability or determine when they have equilibrated.

Fortunately, recent developments in deep learning (DL—a subset of general machine-learning methods that allows for input data and trained features to be encoded at various levels of abstraction) driven image classification provide a potential solution. Through the proper application of training data from

---

[4] http://forge.ipsl.jussieu.fr/ioserver

GCM studies, it might be possible to train a DL model to detect anticipated atmospheric features and dynamics, thus enabling a somewhat automated form of concurrent- or post-processing of simulation data. Note that we focus our investigation on image-classification models due to the maturity of algorithms in this area, driven by research into, for example, facial recognition, on-device image recognition, and of course driverless cars.

In this paper, we will explore how DL-driven classification can aid in our understanding of exoplanetary atmospheres. More specifically, we will explore image classification for a series of HD 209458b-like models based upon those first performed in Sainsbury-Martinez et al. (2019), albeit with varying orbital radii (and hence surface irradiations, and synchronous, rotation rates) in order to expand upon the atmospheric features available for characterization.

The structure of this work is as follows: in Section 2 we introduce both the atmospheric models analyzed here (which we calculate using the 3D GCM DYNAMICO; Dubos et al. 2015) as well as the DL model used for the post-processing image classification (which itself is based on a multi-categorization convolutional neural network (CNN) implemented in Tensor-Flow; Abadi et al. 2015; TensorFlow Developers 2023). We follow this, in Section 3, with our results: first we introduce some of the key features of exemplary HD 209458b-like models, including the atmospheric features selected to train the DL models. Note that that we have made the decision to use computer vision to analyze our results since (a) this allows us to use the visualization of both models and our own analysis, and (b) computer vision DL algorithms are highly mature. Then, after applying the trained neural networks to our atmospheric models, we explore their performance at tagging (i.e., identifying) atmospheric features of interest. This includes a discussion of how the failure of the neural networks to detect any features at some pressures of our slowly rotating model is in and of itself an interesting result that led to the discovery of an uncommon/unusual atmosphere feature: a vertically coherent hot spot on the nightside of our tidally locked hot Jupiter. Finally, we also discuss what features are easy/difficult for the network to detect, as well as how the format and breadth of the data being analyzed impacts the final classifications—including factors ranging from the completeness of the training data set, to the color maps used for each plot, which play an import role in transmitting information from the model to the CNN, with the choice of color maps playing a particularly outsized role in the detection of gradients in model data. We finish, in Section 4, with concluding remarks, discussing the implications of our results and potential plans for future studies that will explore the atmospheric dynamics of variably rotating hot Jupiters in more detail.

2. Methods

In order to investigate how image classification can aid in the concurrent- and post-processing of exoplanetary atmospheric models (particularly hot Jupiters), we must introduce both the atmospheric models we intend to analyze (Section 2.1) as well as the machine-learning model we will use to perform said analysis (Section 2.2).

2.1. Hot Jupiter Models

Following Sainsbury-Martinez et al. (2019, 2021) we use the GCM DYNAMICO to perform a series of 3D atmospheric models at different orbital radii (i.e., with different stellar-

Table 1
Parameters for HD 209458b-like Simulations

Quantity (Units)	Description	Hot HJ	Cool HJ
dt (s)	Time step	120	
N_z	No. of pressure levels	33	
d	No. of subdivisions	30	
$N°$	Angular resolution	2°4	
P_{top} (bar)	Pressure at top	7×10^{-3}	
P_{bottom} (bar)	Pressure at bottom	200	
g (m s^{-1})	Gravity	8.0	
R_{HJ} (m)	Planetary radius	10^8	
a (au)	Orbital radius	0.021	0.192
Ω (s^{-1})	Angular rotation rate	7×10^{-5}	2.54×10^{-6}
c_p (j kg^{-1} K^{-1})	Specific heat	13226.5	
$\mathcal{R}$ (j kg^{-1} K^{-1})	Ideal gas constant	3779.0	
T_{init} (K)	Init. adiabatic T @ 10b	1400	2600

irradiation profiles, and tidally locked, planetary rotation rates). Here we give a brief introduction to DYNAMICO, and its Newtonian cooling approach to radiative transport (Section 2.1.1) before introducing the exact model setups considered here (Table 1), including how the outer atmosphere, equilibrium, Newtonian Cooling profiles were calculated (Section 2.1.2).

2.1.1. DYNAMICO

DYNAMICO is a highly computationally efficient GCM that solves the primitive equations of meteorology (see Vallis 2006 for a review and Dubos & Voitus 2014 for a more detailed discussion of the approach taken in DYNAMICO) on a spherical, icosahedral, grid (Dubos et al. 2015). It remains under development as a next-generation dynamical core for Earth and planetary climate studies at the Laboratoire de Météorologie Dynamique and is publicly available.[5]

In brief, DYNAMICO takes an energy-conserving Hamiltonian approach to solving the primitive equations of meteorology (see Mayne et al. 2019; Showman et al. 2020 for a discussion of the validity and limits of this approach).

Rather than the traditional latitude–longitude horizontal grid (which presents numerical issues near the poles due to singularities in the coordinate system; Williamson 2007), DYNAMICO uses a staggered horizontal–icosahedral grid for which the total number of horizontal cells, N, is defined by the number of subdivisions, d, of each edge of the main spherical icosahedral:[6]

$$N = 10d^2 + 2. \tag{1}$$

In all the models considered here, we set the number of subdivisions to 30, which results in a total horizontal resolution of 9002 cells. This corresponds to an angular resolution of approximately 2.4°. Additionally, at runtime, the output data is passed to XIOS which converts the horizontal–icosahedral grid onto a regular lat-long grid, with a resolution of 90×180 (i.e., 2°) in the lat/long directions, respectively, in order to simplify analysis.

---

[5] DYNAMICO is available at http://forge.ipsl.jussieu.fr/dynamico/wiki.
[6] Specifically, to generate the grid we start with a sphere that consists of 20 spherical triangles (sharing 12 vertex, i.e., grid, points) and then we subdivide each side of each triangle d times using the new points to generate a new grid of spherical triangles with N total vertices. These vertices then form the icosahedral grid.

Vertically, DYNAMICO uses a pressure coordinate system whose levels can be defined by the user at runtime. In our models, this means 33 pressure levels that are linearly spaced in $\log(P)$ space between 10^{-3} and 200 bar.

Finally, the boundaries of our simulations are closed and stress-free with zero-energy transfer (i.e., the only means of energy injection and removal are the horizontal numerical dissipation, i.e., hyperviscosity, and the Newtonian cooling thermal relaxation scheme—see Section 2.1.2).

We introduce the aforementioned horizontal numerical dissipation in order to stabilize the system against the accumulation of grid-scale numerical noise. This takes the form of a horizontal hyperdiffusion filter with a fixed hyperviscosity and a dissipation timescale at the grid scale, $\tau_{\rm dissip}$, which acts to adjust the strength of the filtering (i.e., the longer the dissipation timescale, the weaker the dissipation strength).

For all the models presented here, we set a horizontal dissipation timescale of $\tau_{\rm dissip} = 2500$ following the arguments of Sainsbury-Martinez et al. (2021)—a series of test cases with both faster and slower numerical dissipation were considered, and other than in the most extreme rotation cases (where we found that rapid dissipation lead to model instabilities—hence our choice to set $\tau_{\rm dissip} = 2500$ rather than $\tau_{\rm dissip} = 1250$) the results were found to be essentially $\tau_{\rm dissip}$ independent.

Note that this hyperviscosity is not a direct equivalent of the physical viscosity of the planetary atmosphere, but rather can be viewed as a form of increased artificial dissipation that both enhances the stability of the model and somewhat accounts for sub-grid-scale dynamics. This approach, known as the large eddy approximation, has long been standard practice in both the stellar (e.g., Miesch 2005) and planetary (e.g., Cullen & Brown 2009) atmospheric modeling communities.

Finally, since DYNAMICO (like many other GCMs) does not include a dynamic time step, the time step for each model had to be manually set. For the models considered here, we simply followed prior studies and set the time step to 120 s. Note that test cases with shorter time steps were explored, with little to no difference found in the observed dynamics.

2.1.2. HD 209458b-like Model Atmospheres

In our HD 209458b-like atmospheric models, we do not directly model either the incident thermal radiation on the dayside or the thermal emission on the nightside of the exoplanet. This is due to the high computational cost of modeling radiative dynamics directly with current-generation GCMs and the preliminary science status of next-generation GCMs (e.g., LFRic). Instead, we use a simple thermal relaxation scheme to model these radiative effects, with a spatially varying equilibrium temperature profile, $T_{\rm eq}$, and a radiative relaxation timescale, $\tau_{\rm rad}$, that increases with pressure throughout the outer atmosphere. Specifically, we model the radiation by adding a source term to the temperature evolution equation, which takes the form

$$\frac{\partial T(P, \theta, \phi)}{\partial t} = -\frac{T(P, \theta, \phi) - T_{\rm eq}(P, \theta, \phi)}{\tau_{\rm rad}(P)}. \qquad (2)$$

This method is known as Newtonian cooling and has long been applied within the 3D GCM exoplanetary community (i.e., Guillot & Showman 2002; Showman et al. 2008; Rauscher & Menou 2010; Showman & Polvani 2011; Guerlet et al. 2014;

Mayne et al. 2014a, or Mayne et al. 2014b) when a more complete/complex treatment of the radiative dynamics is unfeasible.

However, in order to use this approach we must first find/set $T_{\rm eq}(P, \theta, \phi)$ for every orbital radii (i.e., every stellar insolation rate). At first glance, the simplest solution would be to use a 1D atmospheric model (like ATMO; Tremblin et al. 2015; Drummond et al. 2016) to calculate dayside and nightside profiles for each hot Jupiter model at each orbital radii. However, as discussed in Sainsbury-Martinez et al. (2021), this technique has a number of downsides, resulting in overly strong outer atmosphere dynamics due to an exaggerated day/night temperature difference (which itself results from the 1D models lacking horizontal advection which cools the dayside and heats the nightside). Similarly, we cannot use the solution proposed in Sainsbury-Martinez et al. (2021) since these models do not represent real exoplanets and hence we lack observations of the advected/observed day/night temperature contrast.

As such we must find an alternate method to define $T_{\rm eq}(P, \theta, \phi)$. Fortunately, since we are interested in the general dynamics of the atmosphere at different orbital radii rather than specific features in comparison to observations, we can take a slightly parameterized approach to derive an approximate $T_{\rm eq}(P, \theta, \phi)$ profile from 1D atmospheric models. This approach can be understood as follows: using ATMO we run a series of 1D models of HD 209458b at 12 different latitudes between the substellar point (which is the point of the atmosphere of the tidally locked planet that is closest to the host star) and the anti-stellar point (i.e., the point on the cold nightside furthest from the host star)—these models used HD 209458-like stellar-irradiation profiles based upon the work of Castelli & Kurucz (2003). We then try to match which 1D models best match the known HD 209458b temperature–pressure profile at different points in the atmosphere. Specifically, and in order to recreate the $T_{\rm eq}(P, \theta, \phi)$ profile from Sainsbury-Martinez et al. (2019), we try to match the dayside, equilibrium, and nightside temperature in the outer atmosphere (i.e., reproducing the observed day/night temperature contrast at 10^{-2} bar) and the radiative-advective boundary temperature in the deep atmosphere (i.e., the temperature at 10 bar, where the day-night temperature difference is assumed to vanish.) The result is that the temperature at 10 bar is best fit with a 1D profile with an irradiation angle of 45° (i.e., halfway between the substellar point and the terminator) and the outer atmosphere day/equilibrium/nightside temperatures (at 10^{-2} bar) are best fit by models with irradiation angles of 25°/78°/84°, respectively. As a consequence of this, and in order to estimate/calculate $T_{\rm eq}(P, \theta, \phi)$ for our HD 209458b-like models at different orbital radii, specifically radii of 0.021 and 0.192 au, we next run the aforementioned 1D models for each of the orbital radii of interest. With this, assuming that the same points in the 3D and 1D models continue to coincide, we can now extract the estimated day/equilibrium/nightside and convergence temperatures, and use them to create parameterized $T_{\rm eq}(P, \theta, \phi)$ profiles, as shown in Figure 1 and Table 2. At the same time, we can also calculate the radiative timescale by perturbing one of the temperature–pressure profiles (specifically the profile with a 45° irradiation angle) at every pressure level and measuring the time taken for the profile to settle (see Showman et al. 2008 for details about this approach). We plot the profile calculated by ATMO, as well as its linear in $\log(P)$
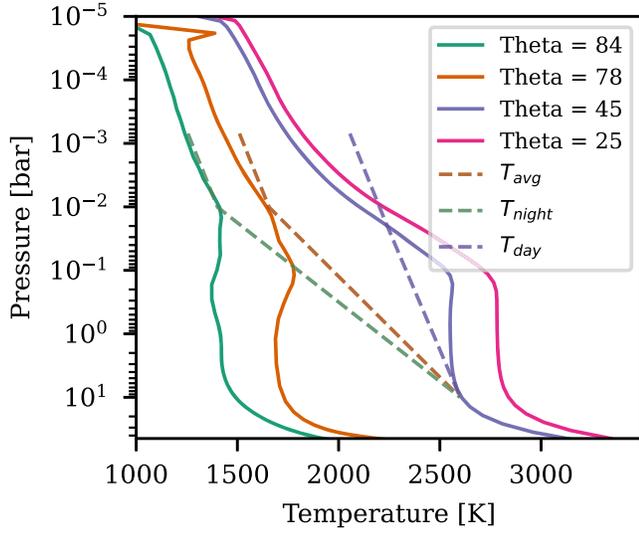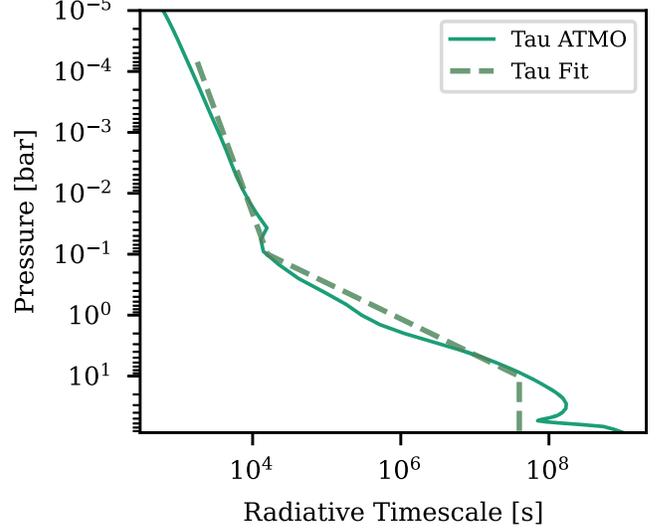
a) 'Hot': 1D and equilibrium T-P profiles
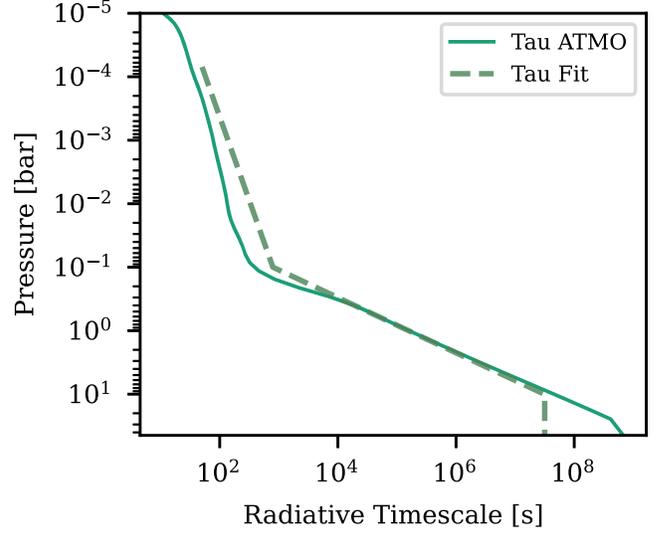
b) 'Hot': radiative timescales
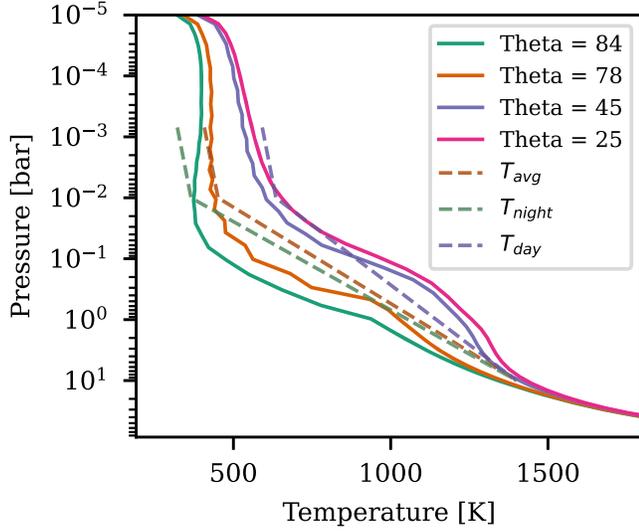
c) 'Cool': 1D and equilibrium T-P profiles

d) 'Cool': radiative timescales

Figure 1. Equilibrium temperature–pressure (left) and radiative timescale (right) profiles for the Newtonian cooling schemes used in our exemplary hot (top) and cool (bottom) HD 209458b-like atmospheric models. Note that in the above plots, the background solid lines correspond to the 1D ATMO models upon which the equilibrium profiles (dashed lines) are based/extrapolated.

Table 2
Equilibrium Temperature Profiles for Our Hot and Cool HD 209458b-like Atmospheric Models

a (au)	ΔT_0 (K)	$T_{\rm eq}$ Profile Interp Points $\left(\frac{T_{\rm eq}}{1\,{\rm K}}, \frac{P}{1\,{\rm bar}}\right)$		
0.021	800	$(1300, 10^{-6})$	$(1800, 10^{-2})$	$(2600, 10)$
0.192	260	$(360, 10^{-6})$	$(500, 10^{-2})$	$(1400, 10)$

Table 3
Newtonian Cooling Radiative Timescale Profiles for Our Hot and Cool HD 209458b-like Atmospheric Models

a (au)	$\tau_{\rm rad}$ Profile Interp Points $\left(\log\left(\frac{\tau}{1\,{\rm s}}\right), \frac{P}{1\,{\rm bar}}\right)$		
0.021	$(1.0, 10^{-6})$	$(2.9, 10^{-2})$	$(7.5, 10)$
0.192	$(2.7, 10^{-6})$	$(4.2, 0.1)$	$(7.6, 10)$

parameterization, in Figures 1(b) and (d), with the values at the interpolation points given in Table 3.

Note that in the aforementioned fits, $T_{\rm eq}(P, \theta, \phi)$, has been split into two pressure-dependent components. This was done for compatibility reasons with DYANMICO's implication of Newtonian cooling. The first is simply the 1D equilibrium profile, $T_{\rm eq-1D}(P)$, which we define using a series of linear in $\log(P)$ space interpolations (with interpolation points given in

Table 2). The second is the day/night temperature difference $\Delta T(P)$, which takes the form

$$\Delta T(P) = \begin{cases} \Delta T_0 & \text{if } P < 10^{-2} \\ \Delta T_0 \log(P/10^{-2}) & \text{if } 10^{-2} < P < 10 \\ 0 & \text{if } P > 10. \end{cases} \quad (3)$$
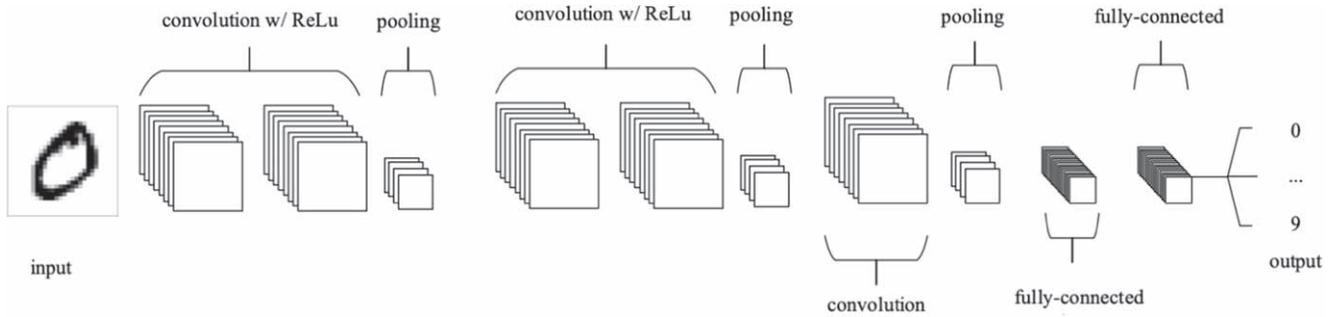
Figure 2. An example CNN layer structure used for modern optical character recognition. Note the use of stacked, and hence nonlinear, convolution layers that are separated by dimensionality-reducing pooling layers, and that eventually feed into fully connected neural network layers that perform the final image classification. Reproduced from O'Shea & Nash (2015).

Taken together, $T_{eq}(P, \theta, \phi)$, is then given by a position-dependent combination of the two profiles:

$$T_{eq}(P, \theta, \phi) = T_{eq-1D}(P) - \frac{\Delta T(P)}{2} + \Delta T(P)\cos(\theta)\max[0, \cos(\phi - \pi)]. \quad (4)$$

2.2. DL Setup

In order to autonomously analyze the outputs of our HD 209458b-like atmospheric models, and hence enable the runtime detection of key/interesting atmospheric features, we decided to make use of a CNN (specifically, the Keras CNN that is implemented as part of TensorFlow). The approach we take here is based upon the work of Lagerquist et al. (2019), who used a DL model to identify large-scale (i.e., synoptic-scale) fronts in meteorological data, modify so as to be able to detect multiple different events (tags) from each image. A more in-depth discussion of the structure of the lightweight CNN considered here is given in the Appendix, where we provide a broader overview.

Briefly, a CNN is a type of DL algorithm (which themselves are a subset of machine-learning data analysis tools) that is particularly well suited to image recognition, and whose design was inspired by how neurons in the visual cortex behave (Lecun et al. 1998). The key feature that separates a CNN from other image classification algorithms is, as the name would suggest, the inclusion of convolution layers. These are neural network layers whose primary purpose is the extraction and isolation of both low-level and high-level features from input data, resulting in a reduction in the size of the data set and thus allowing for the final, reduced, image classification to be performed by more conventional, that is to say fully connected, neural network layers. This reduction in the size of the data set to be worked on is crucial because for the type of images we are interested in classifying, the resolution means that each node (which are the equivalent of neurons) in a fully connected neural network layer would contain far too many weights (i.e., learnable parameters) to fit in even the largest GPUs memory. Consider, for example, a fully connected neural network layer analyzing a 128×128 color image, to maintain complete connection, each *individual* node would have to contain 49,152 weights, ballooning the memory and computational footprint of the network.

Briefly, this decrease in data dimensionality takes place via alternating convolution layers (which do the actual feature analysis and detection) with pooling layers in between (which reduce the dimensionality of the data). Note that, as the dimensionality of the data is reduced, the complexity of the upcoming convolutional layers is typically increased. See Figure 2 for an example of how the layers in a typical OCR CNN are arranged, Figure 15 in the Appendix for a flowchart showing how the layers in our CNN are arranged, and Fukushima & Miyake (1982), LeCun et al. (1989), Lecun et al. (1998), Krizhevsky et al. (2012), O'Shea & Nash (2015), and Albawi et al. (2017) for a more in-depth discussion of both the origins of CNNs, as well their specialized neural network layers.

But how does the CNN model know what features to extract? Via training. Training, specifically supervised training, is the process by which a *blank* neural network is provided a data set in which the expected tags (i.e., classifications) for each image are known—in our case the total training data set consisted of at least 50 images for each tag, although in some cases those images were artificially generated, via an over-sampling approach, in order to expand the available data set (see Section 3.3 for details of how and why this was done). Note that, due to the configuration of our models, the bottom 10 ($P > \sim 25$ bar) and top 10 ($P < \sim 10^{-2}$ bar) layers were not considered when creating the training set, the former because the deep atmosphere had not equilibrated at the time, and the latter because we already had a wealth of features representing the dynamics of the very outer atmosphere (an unadvected dayside hot spot, which we refer to as locked). The network is then evolved, via a sequence of forward and backward data propagation that uses a gradient descent algorithm to optimize a stochastic differentiable manifold (i.e., a surface of differential equations) to solve a very nonlinear problem, until it is able to reproduce the set of known tags with its own set of assigned tags. For all the models discussed here, this training was complete after approximately 30 epochs (i.e., iterations) with either no improvement or a regression in model accuracy (i.e., the ability of the model to reproduce the training data identified tags without being guided) if the model was evolved further—a phenomenon known as overfitting. For the CNNs discussed here our initial set of hand-selected/identified data was split into separate training and validation pools, with 80% of the tagged images being used to train the CNN and 20% of images used for validation. The results of this validation procedure can be seen in Figure 3, which reveals that the trained CNNs are able to recover approximately 90% of expected features in the validation data set after 30 epochs of evolution, after which accuracy regressed.

After this rather computationally expensive training process, we were left with a pair of multi-classification CNNs, one for
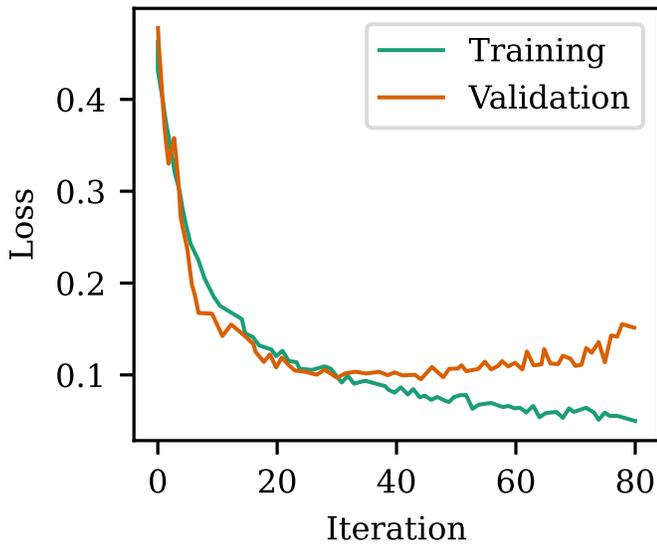
Figure 3. Convergence tests for our multi-categorization CNN, i.e., thermal features CNN, showing the convergence of the loss function during both training (green) and validation (orange). The training is considered complete/optimized after 30 iterations since this is the point at which overfitting starts to occur, as represented by a regression in validation accuracy (i.e., an increase in loss).

thermal features and one for wind features, that could rapidly analyze the results of new/evolved simulations. We discuss this analysis, as well as potential limitations/pitfalls (and how to overcome them) below.

3. Results

In order to explore the use of machine learning and image classification for the autonomous detection and characterization of hot Jupiter atmospheres, we ran a large series of synchronously rotating, HD 209458b-like, hot Jupiter atmospheric models with orbital radii between 0.012 and 0.334 au. This orbital-radii range was selected to span from near the inner edge of the distribution of known exoplanets (where irradiation is strong and synchronous rotation is rapid) out toward the upper limit of the highly irradiated and synchronously rotating regime (which occurs when tidal effects are weak enough that the synchronization time approaches the system age). Within this range, orbital radii were selected to be a whole fraction/multiple of the orbital radius of HD 209458b.

Early outputs of these variable orbital-radii models were then manually tagged with atmospheric features of interest (see Sections 3.2/3.3), after which this set of tagged data was used to train the CNNs. Finally, after the simulations had been run further and the outer atmospheres ($P < 1$ bar) had equilibrated, the full time series of temporally averaged (in order to remove small-scale fluctuations) horizontal wind and temperature maps were fed into the CNNs and a series of multi-categorization maps were generated—here we focus on this process, and its results, for two HD 209458b-like models in different rotation/insolation regimes, one *hot*, with a short orbital radius and hence strong surface irradiation, and one *cool*, with a longer orbital radius and hence weak surface irradiation. We also briefly explore the ability of our trained CNN to detect atmospheric features for the adiabatically initialized model of Sainsbury-Martinez et al. (2019)—a model that is expected to contain examples of all the atmospheric features of interest. Finally, we also discuss the importance of selecting proper

training/validation data for a CNN, including a discussion of how the initial categorization of our hot Jupiter atmospheres led to the discovery of an uncommon/unusual, but robust, atmospheric feature on the nightsides of our cool regime HD 209458b-like atmospheric models.

3.1. HD 209458b-like Atmospheric Models

For the sake of brevity, we have chosen to focus our analysis and discussion on two HD 209458b-like models, which, together, exhibit all the key features found in the broader set of models explored. Specifically, we have chosen one model in each of the primary dynamical regimes observed: a short orbital radius, $a = 0.021$ au, model (labeled hot) in which the dynamics are strongly influenced by both the strong stellar irradiation and significant rotational effects (i.e., strong Coriolis forcing), and a long-orbital radius, $a = 0.192$ au, model (labeled cool) in which the stellar insolation is significantly weakened, and the dynamics are much less rotationally influenced, leading to a weaker equatorial jet and a more divergence-driven day/night wind.

The difference in dynamics between these two primary regimes of interest can easily be identified when exploring the zonal-mean (i.e., longitudinal-mean) zonal wind and meridional (i.e., vertical and latitudinal) stream function (where the meridional steam function is a measure of mass circulations on the meridional/latitudinal plane). We plot both of these quantities in Figure 4 for our two exemplary HD 209458b-like atmospheric models.

Starting with the zonal wind (left) we find that, at steady state, our hot model (top) maintains a strong, deep, super-rotating (i.e., easterly) equatorial jet that is braced by significantly weaker, midlatitude, westerly counterflows. This is reminiscent of a mix of the wind structure found when modeling HD 209458b and SDSS1411b with DYNAMICO (Sainsbury-Martinez et al. 2019, 2021) and matches the leading theory for the driving mechanism of super-rotating jets in hot Jupiters: the pumping of easterly angular momentum from midlatitudes to the equator by standing Rossby and Kelvin waves, resulting in easterly acceleration at the equator and westerly acceleration at midlatitudes (Showman & Polvani 2011; Tsai et al. 2014).

Moving onto the slower rotating, and more weakly radiatively driven, cool model (bottom), the zonal-mean zonal wind profile reveals that while a super-rotating *jet* appears to have once again formed, it is both significantly more vertically constrained and slower than the jet found in either the hot model or prior HD 209458b models. We will explore why these differences occur in more detail below (see also Figure 4), but briefly it can be linked to the relative strength of divergence-driven (i.e., day-night) in comparison to rotationally driven (i.e., Rossby and Kevin wave-driven) flows—as divergent/rotational flows become stronger/weaker, respectively, the zonal-mean zonal jet weakens since divergent flows have little to no east/west preference, leading to them broadly canceling out in the zonal mean, and rotational winds are key to driving a super-rotating jet due to their role in angular momentum transport.

In turn, these very different zonal winds drive very different meridional circulations. In our hot model, the meridional circulation stream function reveals a pair of narrow clockwise (northern hemisphere—negative latitudes) and anticlockwise (southern hemisphere) circulation cells which drive a downflow at the equator, and in conjunction with weaker high-latitude
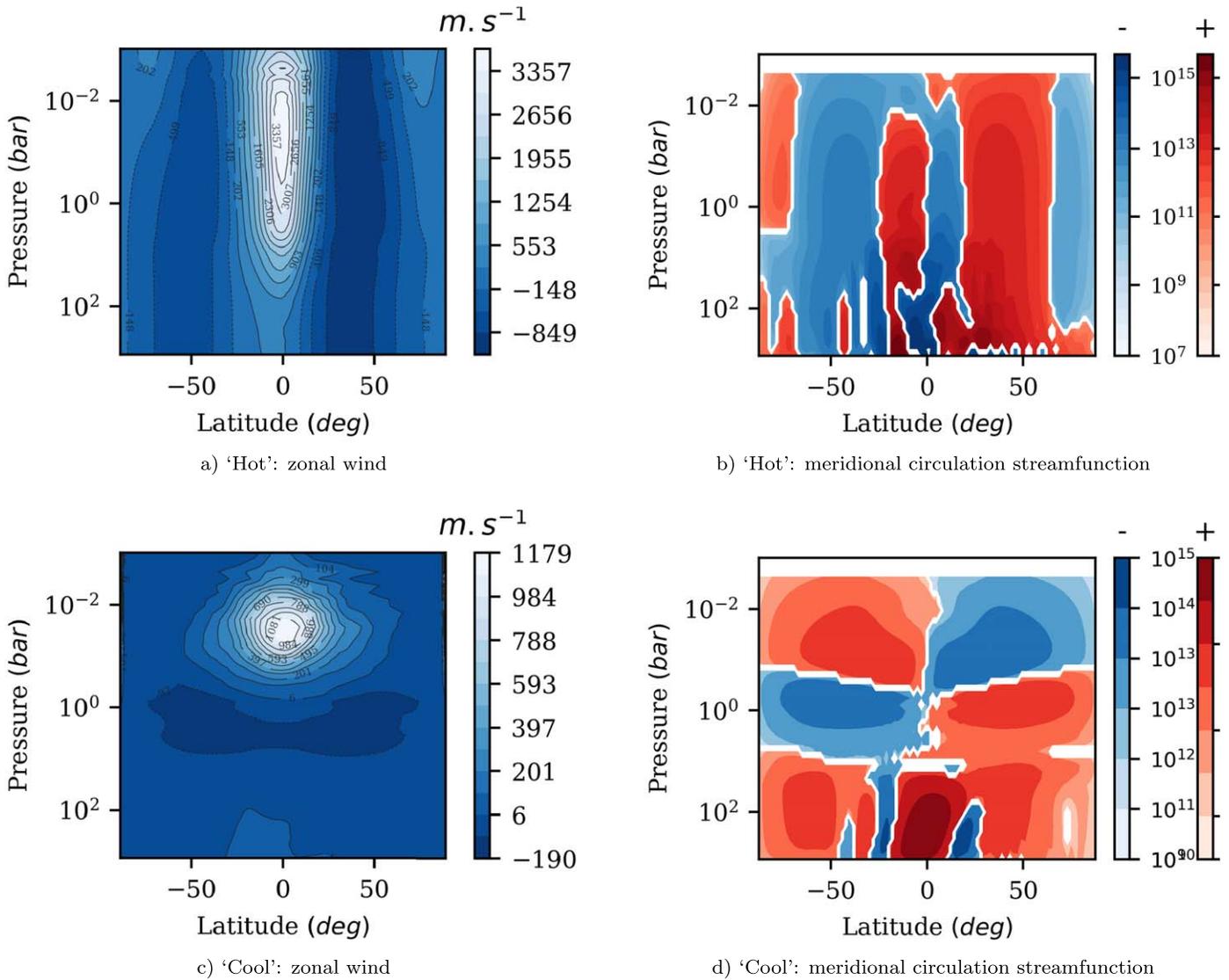
a) 'Hot': zonal wind

b) 'Hot': meridional circulation streamfunction

c) 'Cool': zonal wind

d) 'Cool': meridional circulation streamfunction

Figure 4. Longitudinally and temporally averaged zonal wind (left) and meridional circulation stream function (right) for both our hot (top) and cool (bottom) HD 209458b-like atmospheric models. In the zonal wind maps, positive quantities correspond to eastward flows, while in the meridional circulation profiles clockwise circulations are shown in red and anticlockwise in blue. Additionally, the meridional circulation stream function is plotted on a log scale.

circulation cells, upflows at low to midlatitudes—specifically those latitudes at which the zonal wind transitions from being primarily easterly to westerly. In turn, as discussed in Sainsbury-Martinez et al. (2019), this circulation drives a significant vertical heat flux leading to the heating of the deep, advective, adiabat, and hence radius inflation (see Figure 5(a)) due to the increased internal entropy (Tremblin et al. 2017; Sainsbury-Martinez et al. 2019).

On the other hand, in our cool model, the circulation is rather different. Not only are the circulation cells much wider latitudinally, with each cell occupying an entire hemisphere, but the circulation direction of the cells changes with pressure. In the outer atmosphere, we have a clockwise/anticlockwise circulation cell in the northern/southern hemispheres, respectively, which drives a downflow at the equator (with balancing upflows at the poles). Whereas at higher pressures (down to 10 bar), we find that the sense of the circulations has changed, resulting in a slight upflow at the equator. The pressure at which this change in circulation regime occurs matches the pressure at which the zonal-mean zonal jet vanishes, which

reinforces the conclusion of Tremblin et al. (2017) and Sainsbury-Martinez et al. (2019) that flows associated with the equatorial jet are responsible for driving vertical flows that advect mass/temperature/enthalpy. This also explains why we see little to no heating of the deep, advective, adiabat (Figure 5(b)), and hence would expect to observe little to no radius inflation for a planet at a similar orbital radius and stellar insolation.

3.1.1. A Note about Deep Atmosphere Equilibrium

It is important to note that, unlike Sainsbury-Martinez et al. (2019) and Sainsbury-Martinez et al. (2021), the deep atmosphere circulation profiles considered here reveal significant time variability. This is simply a result of the deep atmosphere only being initialized close to its adiabatic steady state and the high computational cost of running the models long enough for the dynamically slow deep atmospheres to completely thermally equilibrate—although this should not affect our results here as our focus in on the use of AI for analysis, and not the steady-state dynamics in the deep
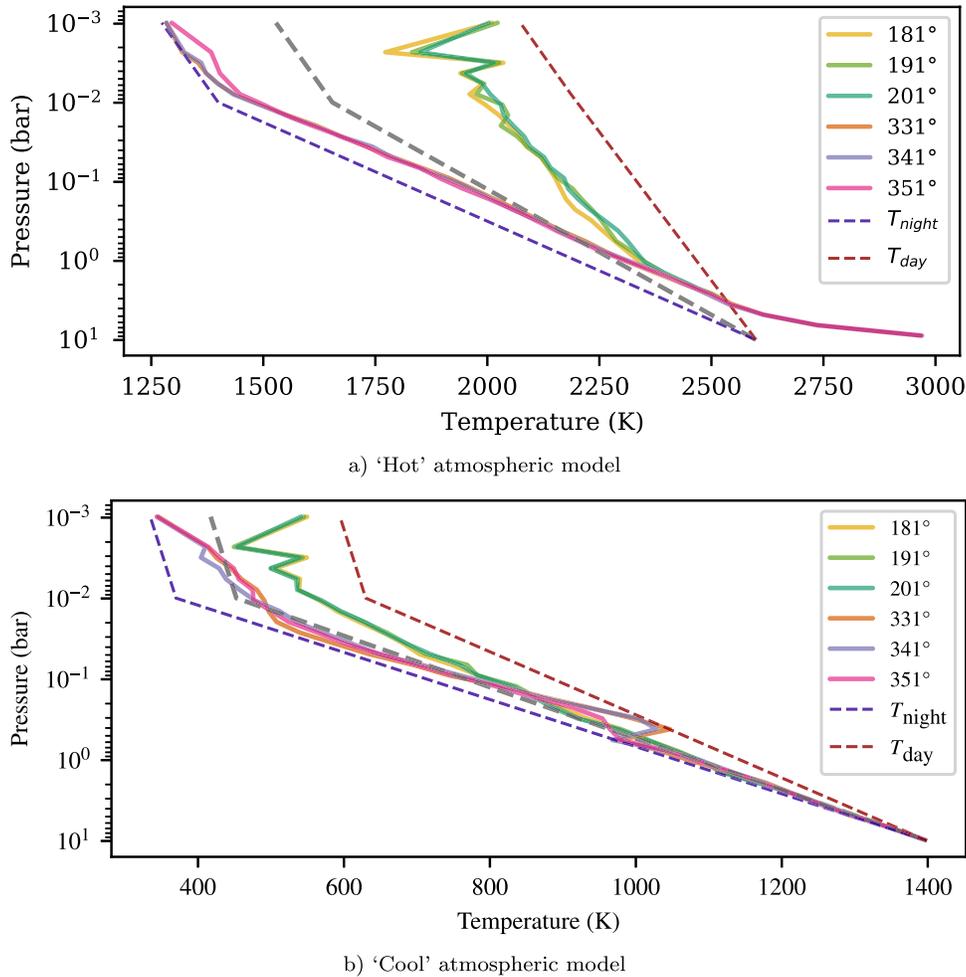
a) 'Hot' atmospheric model

b) 'Cool' atmospheric model

Figure 5. Latitudinally and temporally averaged (at the equator) T–P profiles for our hot (top) and cool (bottom) HD 209458b-like atmospheric models. Each plot includes profiles from six different longitudes, ranging from the substellar point (whose equilibrium profile is shown in red) eastward to the anti-stellar point (whose equilibrium profile is shown in blue) on the nightside. Note that we have excluded the deep atmospheres ($P > 10$ bar) from these plots since it has not fully equilibrated.

atmosphere, the detailed analysis of which we leave to future studies.

However, while attempts have been made to solve the problem of the high computational cost of resolving the steady-state deep atmospheres, these solutions come with their own problems. For example, Schneider et al. (2022) attempted to find the steady-state deep atmosphere of the ultrahot Jupiter WASP-76b by calculating the observed cooling at 650 bar in their model, including its rate of change, and then using this to extrapolate toward the steady-state T–P profile everywhere in the atmosphere (that is to say at pressures lower than 650 bar), which they find to be *cold*, implying that potential temperature advection is not driving any deep heating. However, this approach has a number of caveats which mean that extrapolation is generally a bad idea: to start, the pressure dependence of radiative heating, and radiative dynamics more generally, means that the 650 bar cooling rate is unlikely to be representative of the cooling rate at lower pressures. Specifically, at lower pressures, one needs to assess whether the atmosphere has reached equilibrium with the radiative dynamics or with the advection of potential temperature. An assessment cannot be made via the dynamics at 650 bar, especially since even this region has

not yet reached equilibrium. Furthermore, the model was initialized significantly hotter than the expected, inflated, steady state, resulting in an enhanced deep cooling rate driven by the deep atmosphere's need to expel excess energy. In turn, this deep cooling will drive very different deep circulations (as seen by Sainsbury-Martinez et al. (2019) when they modified the equilibrium temperature of the outer atmosphere of a previously equilibrated model), which takes time to evolve/settle back to deep heating once the atmosphere has cooled—in fact, evidence of this shift to deep heating can be seen in Figure 2 of Schneider et al. (2022): the extrapolated model shows signs of heating that is slowly pushing deeper with time, driven by potential temperature advection, and which will likely lead to a steady state somewhere between the two models at equilibrium. As such, we advise against estimating the deep atmosphere's temperature–pressure profile via extrapolation, and instead, suggest that future studies focus on next-generation GCMs, which will be efficient enough to model the equilibrium state of the deep atmosphere within a reasonable, computational, timescale. And which will benefit greatly from pairing with a concurrent, AI-analysis model.
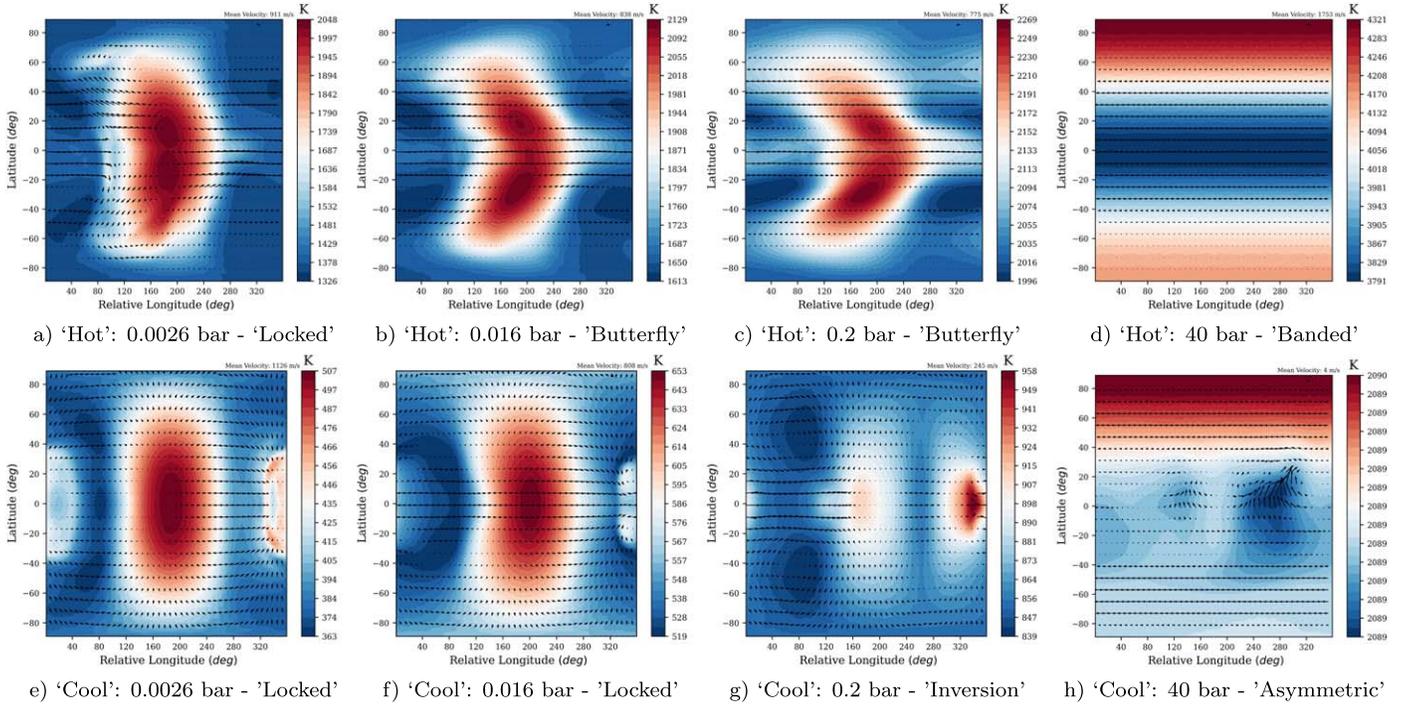
a) 'Hot': 0.0026 bar - 'Locked'   b) 'Hot': 0.016 bar - 'Butterfly'   c) 'Hot': 0.2 bar - 'Butterfly'   d) 'Hot': 40 bar - 'Banded'

e) 'Cool': 0.0026 bar - 'Locked'   f) 'Cool': 0.016 bar - 'Locked'   g) 'Cool': 0.2 bar - 'Inversion'   h) 'Cool': 40 bar - 'Asymmetric'

Figure 6. Temporally averaged zonal wind (arrows) and temperature profile (map) at four different pressure levels (0.0026 bar—left, 0.016 bar—middle left, 0.2 bar—middle right, and 10 bar—right) for both our hot (top) and cool (bottom) HD 209458b-like atmospheric models. Each profile has been labeled with the (class of) tag assigned to it by the CNN.

3.2. Initial Data Tags, Training, and Results

While interesting, the aforementioned zonal-mean dynamics are not what we intend to explore with our neural networks. Instead, we plan to look at the pressure-dependent horizontal wind and temperature profiles that combine to give these zonal-mean flows, and which contain many more interesting, and unique, features for the CNNs to detect.

Initially, we decided to train our CNNs to detect the presence of dayside hot spots in which the zonal winds have caused significant horizontal thermal advection (and whose shape is typically referred to as a butterfly in the hot Jupiter community —e.g., Figures 6(b) and (c)), longitudinally homogenized and latitudinally symmetric thermal bands (e.g., Figure 6(d)), dayside hot spots in which radiative effects dominate over advective dynamics (i.e., an irradiative hot spot which has not been significantly advected by horizontal winds—e.g., Figures 6(e) and (a) to a lesser extent), latitudinally asymmetric thermal structures (see, e.g., Figure 6(h)), and finally, the presence of an equatorial zonal jet (see the arrows that trace the horizontal wind in, for example, Figures 6(b) and (c)), although, as discussed in Section 3.4.1, this final categorization, and hence the wind-based CNN more generally, did not pan out for a number of reasons.

This training was performed using early outputs from our complete simulation data set, with orbital radii of between 0.012 and 0.334 au, which was hand labeled/tagged such as to produce at least 50 examples of each feature of interest.

Once the training was complete, but before we explored how rotation impacts the detected atmospheric features, we first investigated how our trained neural networks would behave when applied to a model that has manually been confirmed to contain examples of all of the current atmospheric features of interest, and from which no training or validation data was extracted.

Specifically, we consider the adiabatically initialized HD 209458b model of Sainsbury-Martinez et al. (2019). As shown in Figure 7, and as anticipated, the thermal CNN successfully identifies all of the expected atmospheric features, while on the other hand the zonal jet, and the zonal wind more generally, proved to be highly intractable (Section 3.4.1). The identified features include an outer atmosphere that is dominated by the radiatively driven (tidally locked) dayside hot spot, a mid-atmosphere that shows signs of an advected hot spot (i.e., a thermal butterfly), and a deep atmosphere that shows a mix of asymmetric and banded thermal structures, likely as a result of the ongoing deep heating in the model as it warms from its slightly cooler than steady-state adiabatic start.

Note that, by testing against a model from which no training or validation data was extracted, we are able to test the portability/generality of our trained CNNs. However, it is important to note that the model we consider was also calculated using DYANMICO, using a very similar setup for HD 209458b. For future studies, we suggest that a more diverse range of models should be considered when testing generalizability. This could include testing the CNNs on the outputs of models run with different GCMs or on the outputs of models calculated using the same GCM, but for rather different objects, such as the brown dwarf models of Sainsbury-Martinez et al. (2021).

Having confirmed that our CNNs can recover all the thermal atmospheric features for which they were trained, but when applied to a model from which no training data was taken, we next move on to exploring how rotation influences the detected dynamics. To do this, we consider our two exemplary models, which we ran until their outer atmospheres had reached equilibrium, i.e., for ~40 Earth yr of simulation time. We then applied the trained CNNs to the full data set, using a temporal averaging window of between 100 and 150 outputs in order to
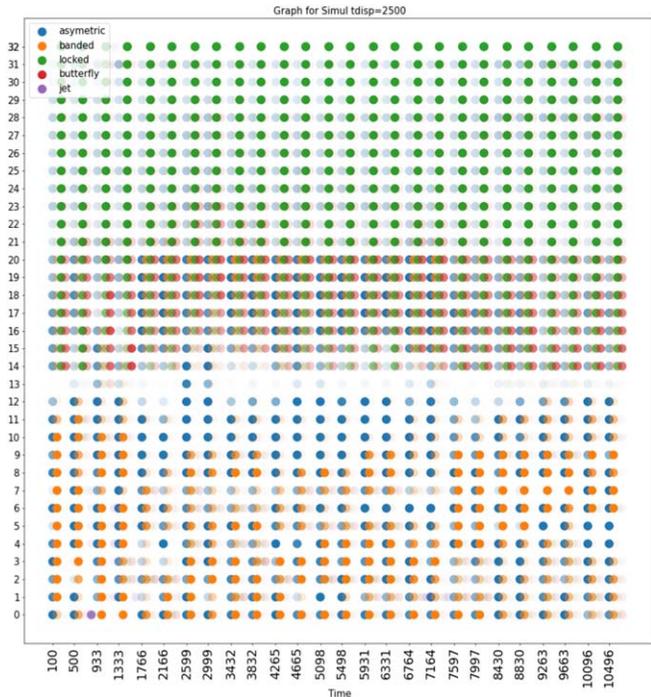
Figure 7. AI multi-categorization map for the HD 209458b atmospheric model of Sainsbury-Martinez et al. (2019). Here we show the categories detected (with the opacity of each detection point representing the strength of the detection) at each pressure level (where an increase in pressure level corresponds to a decrease in actual pressure as we move to higher altitudes) against time (where $t = 0$ corresponds to solid-body, adiabatic, initialization, and each latter point corresponds to the center of the temporal mean). The categories in question correspond to the detection of a north–south asymmetric thermal structure in blue, a banded (i.e., horizontally homogenized) thermal structure in orange, a radiative-dominated thermal structure (driven by tidal locking) in green, a butterfly thermal structure (i.e., eastward equatorial advection flanked by slight, off-equator, westward advection) in red, and an equatorial (wind) jet in purple. Note that this model was run at a lower resolution than the other models considered here, which has slightly impacted the ability of the CNN to discriminate between different atmospheric features.

Figure 8. AI multi-categorization map for our hot HD 209458b-like atmospheric model. Here we show the categories detected (with the opacity of each detection point representing the strength of the detection) at each pressure level (where an increase in pressure level corresponds to a decrease in actual pressure as we move to higher altitudes) against time (where $t = 0$ corresponds to solid-body, adiabatic, initialization, and each latter point corresponds to the center of the temporal mean). The categories in question correspond to the detection of a north–south asymmetric thermal structure in blue, a banded (i.e., horizontally homogenized) thermal structure in orange, a radiative-dominated thermal structure (driven by tidal-locking) in green, a butterfly thermal structure (i.e., eastward equatorial advection flanked by slight, off-equator, westward advection) in red, a vertical thermal inversion in purple (increasing T with P) and brown (decreasing T with P), and an equatorial (wind) jet in pink.

reduce the effects of small-scale oscillations on the final characterization. Note that we used broader averaging windows for the HD 209458b-like model shown in Figure 7 due to the longer time series available for that model.

As was the case for the zonal-mean dynamics, we find that the horizontal dynamics, and hence the detected atmospheric features, change significantly with rotation rate. Specifically, we find little crossover in the features detected in our exemplary hot and cool models:

Starting with the hot regime, whose detected characteristics are shown in Figure 8, we find that the dynamics are dominated by three features: at very low pressures (which are slightly higher near initialization, before advection kicks in) the thermal structure is dominated by a (tidally) locked dayside hot spot. However, this changes as we move toward higher pressures, with the advective timescale becoming comparable to (and eventually faster than) the radiative timescale, leading to significant horizontal advection and hence a transition to a thermal butterfly. Note that this detection of a butterfly tag does eventually vanish at later times, but as we discuss in Section 3.4.1, this is a problem with our initial training data set not properly accounting for the influence of very rapid rotation on horizontal thermal advection. Finally, the deep atmosphere is generally dominated by latitudinally symmetric,
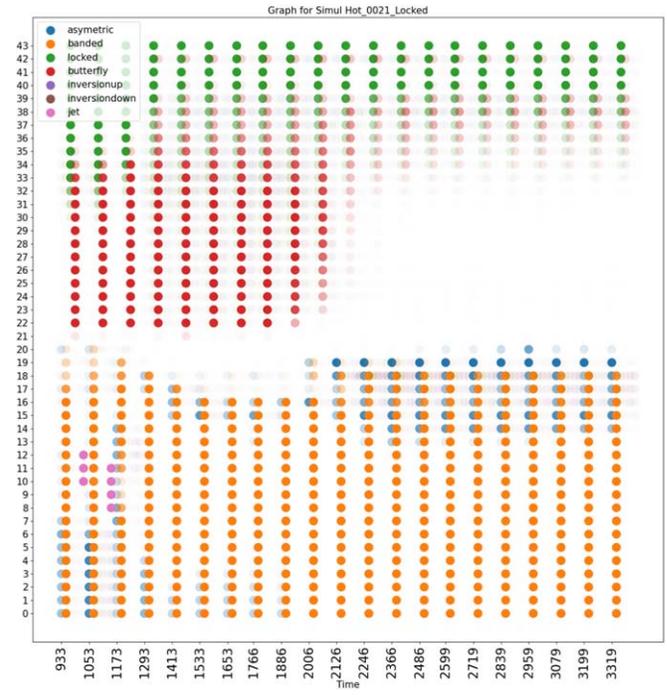
and longitudinally homogeneous, thermal bands, which indicate that it is generally well homogenized longitudinally, likely due to the zonal jet. Note, however, that we do occasionally find that the CNN assigns the asymmetric tag to the outer deep atmosphere—this is due to ongoing heating of the lowest pressure regions of the deep atmosphere.

On the other hand, analysis of our cool regime model, as shown in Figures 9(a) and (b), reveals a rather different set of identified dynamics. Here we again find that the outer atmosphere is dominated by a tidally locked, dayside, hot spot, but unlike in our hot model, this does not transition into a thermal butterfly with increasing pressure. Instead, we find that (a) the locked profile extends significantly deeper than in the hot model, except near initialization where the radiative-forcing timescale dominates the dynamics, and (b) this detection transitions into a region of nondetection within the mid-atmosphere (i.e., between pressure levels 20/21 (~1 bar) and 29 (~0.05 bar)). The former effect can likely be explained by the weaker role that zonal winds play in the atmospheric dynamics of cool regime models (Figure 4(c) and Section 3.4), while we explore the later lack of detected feature in more detail in Section 3.3. Finally, in the deep atmosphere, we find that detected dynamics are generally dominated by weakly asymmetric thermal bands. Note that, unlike in the hot regime, these asymmetric structures are not being driven by deep heating, instead analysis (Section 3.4) suggests that it occurs
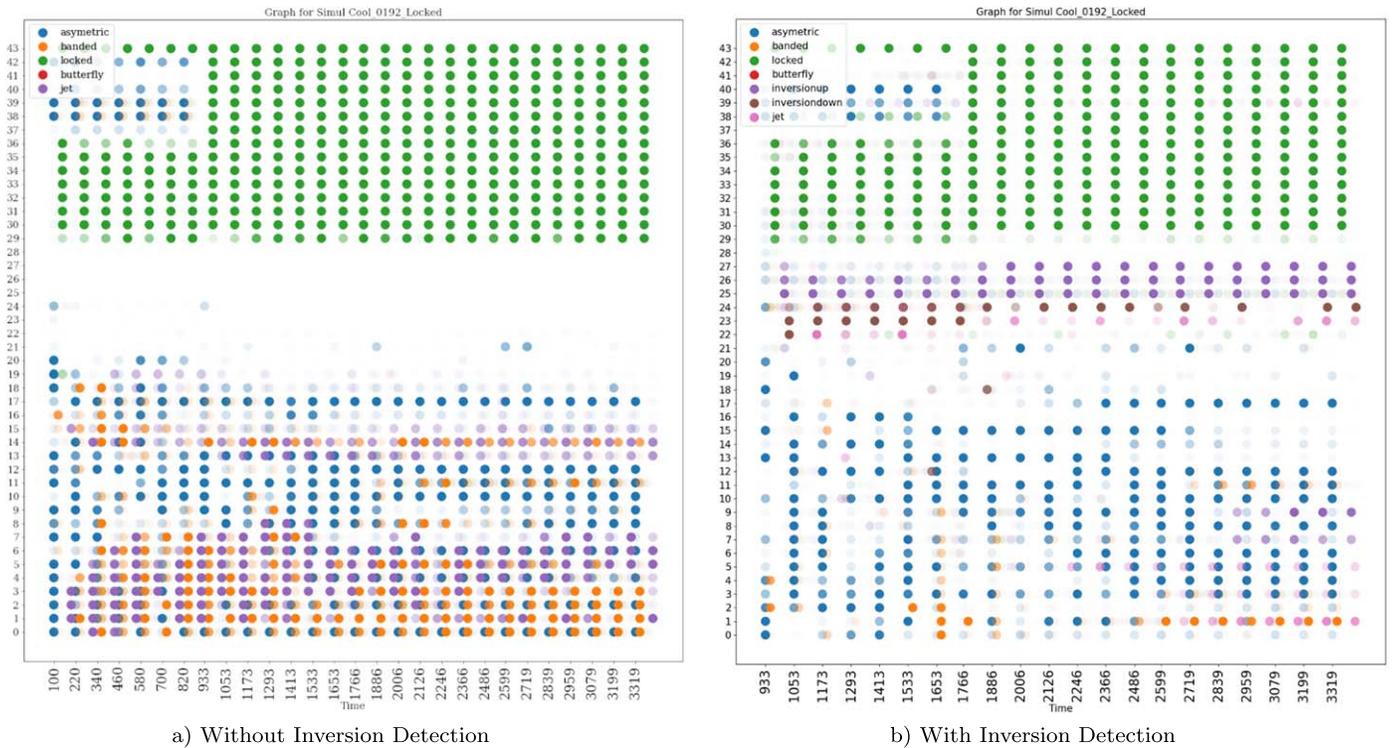
a) Without Inversion Detection   b) With Inversion Detection

Figure 9. AI multi-categorization map(s) for our cool HD 209458b-like atmospheric model. Here we show the categories detected (with the opacity of each detection point representing the strength of the detection) at each pressure level (where an increase in pressure level corresponds to a decrease in actual pressure as we move to higher altitudes) against time (where $t = 0$ corresponds to solid-body, adiabatic, initialization, and each latter point corresponds to the center of the temporal mean) for two different multi-categorization CNNs, one without thermal inversion detection (left) and one with (right). Note that the CNN that lacked thermal inversion detection also differs from the CNN used throughout the rest of this work in a number of other regards being an earlier iteration of the final model. This includes changes to the color map, resolution, and averaging period of the input data. The categories in question correspond to the detection of a north–south asymmetric thermal structure in blue, a banded (i.e., horizontally homogenized) thermal structure in orange, a radiative-dominated thermal structure (driven by tidal locking) in green, a butterfly thermal structure (i.e., eastward equatorial advection flanked by slight, off-equator, westward advection) in red, a vertical thermal inversion (in the figure on the right-hand side) in purple (increasing T with P) and brown (decreasing T with P), and an equatorial (wind) jet in purple (left-hand side) or pink (RHS).

because the deep atmosphere is rather quiescent (Figure 4(c)), with very weak vertical heat transport, leading to persistent, but weak, horizontal temperature gradients in the deep atmosphere.

Note that the differences between the detected features in Figures 9(a) and (b) are due to the preliminary nature of the models that lacked thermal inversion detection (see below—Section 3.3), with the preliminary models using a different color map and averaging scheme to the final CNN models presented in Section 2.2.

In Section 3.4, we explore why these differences in detected atmospheric features occur.

### 3.3. Updated Data Tags: A Nightside Thermal Inversion

In an effort to understand the region of nondetection identified for our cool atmospheric models (Figure 9(a)), we elected to explore this region of the atmosphere in more detail, and if appropriate, update our thermal CNN with what we find. As shown in Figures 6(f) and (g), our analysis reveals that, once we are deep enough for advective transport to start to dominate over radiative forcing (via Newtonian cooling), rather than a butterfly-like thermal structure on the dayside, as found in the classical hot Jupiter regime (e.g., Figure 6(c)) we instead find that a hot spot has formed on the cold nightside, slightly to the west of the anti-stellar point (see Figure 6(f)). This was of particular interest since its presence may have significant implications for both the atmospheric dynamics and observable features of more weakly irradiated Jupiters (so-called *warm* Jupiters). Furthermore, it

came as somewhat of a surprise since prior studies (e.g., Sainsbury-Martinez et al. 2019), other HD 209458b-like models (e.g., Figure 8), as well as our initial training tags suggested that weak (due to the slower rotation rate and weaker surface irradiation at higher orbital radii) butterfly-like features should have been detected at these pressure levels. One of the main potential impacts of this nightside hot spot is its associated thermal inversion, which can be clearly seen in longitudinally sliced temperature–pressure profiles of the cool model (Figure 5(b)). Briefly, as shown here, a thermal inversion occurs when the temperature profile switches from cooling as the pressure decreases to warming with decreasing pressure—this is similar to the stratosphere on Earth and can have significant effects on observed atmospheric dynamics and chemistry (see, for example, Hubeny et al. 2003; Fortney et al. 2008; Spiegel et al. 2009; Zahnle et al. 2009; Madhusudhan & Seager 2010; Mollière et al. 2015; Beatty et al. 2017; Lothringer et al. 2018; Gandhi & Madhusudhan 2019 for the discussion of thermal inversions in highly irradiated hot Jupiters).

Of course, since this feature was not anticipated, our initial training data did not include it, hence the blank regions on the multi-categorization maps. As such, and in order to better explore how widespread this feature is, we updated our training data set (and the underlying thermal CNN) to include a pair of additional tags designed to cross-correlate temperature structures on the nightside and hence detect hot spots: one indicating a nightside hot spot which cools as the pressure decreases
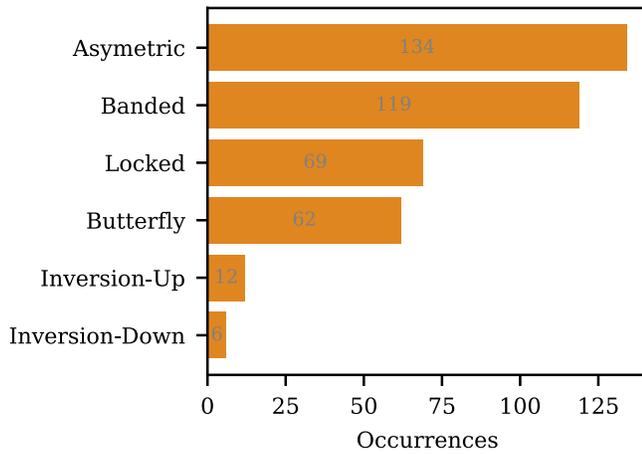
a) Distribution of Training Data Before Oversampling
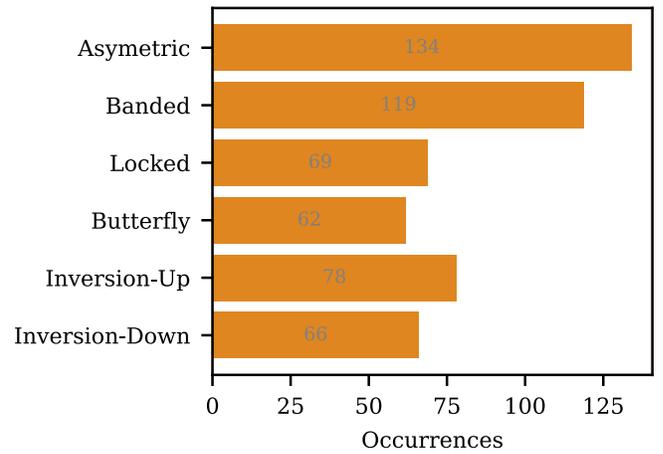
b) Distribution of Training Data After Oversampling

Figure 10. Distribution of (thermal) training/validation data before (top) and after (bottom) applying an oversampling technique to the inversion data set so as to generate artificial training data. The total count of items in each training data set is shown in gray. Note that we do not include the amount of training data used for the jet detection for two reasons: (a) unlike all the other features that were trained on the thermal structure, the jet tag was trained in the horizontal wind, and (b) the jet detection model did not make it into our final analysis due to problems with detecting highly symmetric structures (Section 3.4.1).

(*inversion down*) and one indicating the opposite (*inversion up*).

However, adding these new tags to the CNNs was not a simple matter since we had but a few examples of this phenomenon to use as training data (Figure 10(a)). To resolve this, and thus properly train our CNNs to detect nightside hot spots/thermal inversions, we turned to interpolative oversampling: that is to say, we used interpolation to create a series of artificial tagged images from our limited sample of hand-tagged examples. For the inversion-up tag, we generated three artificial tagged images per input image, while for the inversion-down tag, which was significantly more numerous in our hand-labeled data than the inversion-up tag, we merely generated a single artificial tagged image per input image. Consequently, we now had at least 50 examples of each new feature (Figure 10(b)), which could be used to update our multi-categorization thermal CNN to detect the new feature of interest.

The results of this updated analysis can be seen, for our cool regime model, in Figure 9(b), which reveals that the mid-atmosphere region of nondetection has been replaced with both inversion-up and inversion-down tags, which when taken together indicate the peak of the nightside hot spot (and also the pressure level at which the hottest point shifts from the dayside to the nightside). Further analysis of our models reveals that this feature is both robustly, but uniquely, present in the cool regime. For example, our hot regime model reveals no detected inversion tags, as expected from our visual analysis. But why does this feature only occur in our cool models? As we discuss below, it appears to be linked to the relative influence of rotation on the atmospheric dynamics and thermal energy transport.

3.4. How Atmospheric Dynamics Impact the Observed Features

As previously alluded to, the HD 209458b-like atmospheric models we consider here fall into two distinct regimes depending upon their, tidally locked, orbital radii (i.e., surface irradiation and rotation rate). At short orbital radii (i.e., the hot regime), the zonal-mean atmospheric dynamics are dominated by a strong, super-rotating, equatorial jet that extends deep into the atmosphere (Figure 4(a)), driving significant downflows (Figure 4(b)) that result in strong vertical mixing, and hence deep heating (and radius inflation; Tremblin et al. 2017; Sainsbury-Martinez et al. 2019). On the other hand, at longer orbital radii (i.e., the cool regime), this is no longer the case and instead, the zonal-mean zonal jet is significantly weaker and shallower (Figure 4(c)), and thus associated with significantly reduced vertical mixing (Figure 4(d)), which in turn drives little to no deep heating.

However, in order to fully understand the dynamics identified by our CNNs, we must look at more than just the zonal-mean dynamics. Specifically, we are interested in the differences in the horizontal wind between the hot and cool regimes, and how these differences affect horizontal and vertical energy (enthalpy) transport, leading to the various atmospheric features identified by our CNNs, particularly the nightside hot spot found in the cool regime.

To that end, we next explore the Helmholtz decomposition of the zonal wind, a decomposition that has previously been used to study both the atmosphere of Earth (Dutton 1986) and hot Jupiters (Hammond & Lewis 2021). Briefly, a Helmholtz decomposition can be used to split the horizontal wind at each pressure level, $\boldsymbol{u} = (u, v)$, into divergent (i.e., *vorticity-free*) and rotational (i.e., *divergence-free*) components (Dutton 1986):

$$\boldsymbol{u} = \boldsymbol{u}_{\mathrm{div}} + \boldsymbol{u}_{\mathrm{rot}} \tag{5}$$

$$= -\boldsymbol{\nabla}\chi + \boldsymbol{k} \times \boldsymbol{\nabla}\psi, \tag{6}$$

where χ is the velocity potential function, ψ is the velocity stream function, and both can be linked to the divergence δ/vorticity w directly:

$$\nabla^2 \chi = \delta \tag{7}$$

$$\nabla^2 \psi = w. \tag{8}$$

Additionally, in order to further isolate the equatorial zonal jet from other wind dynamics, we further split the rotational component, $\boldsymbol{u}_{\mathrm{rot}}$ into a zonal-mean component $\boldsymbol{u}_{\mathrm{zonal}}$ and an eddy component $\boldsymbol{u}_{\mathrm{eddy}}$:

$$\boldsymbol{u}_{\mathrm{zonal}} = \langle \boldsymbol{u}_{\mathrm{rot}} \rangle \tag{9}$$
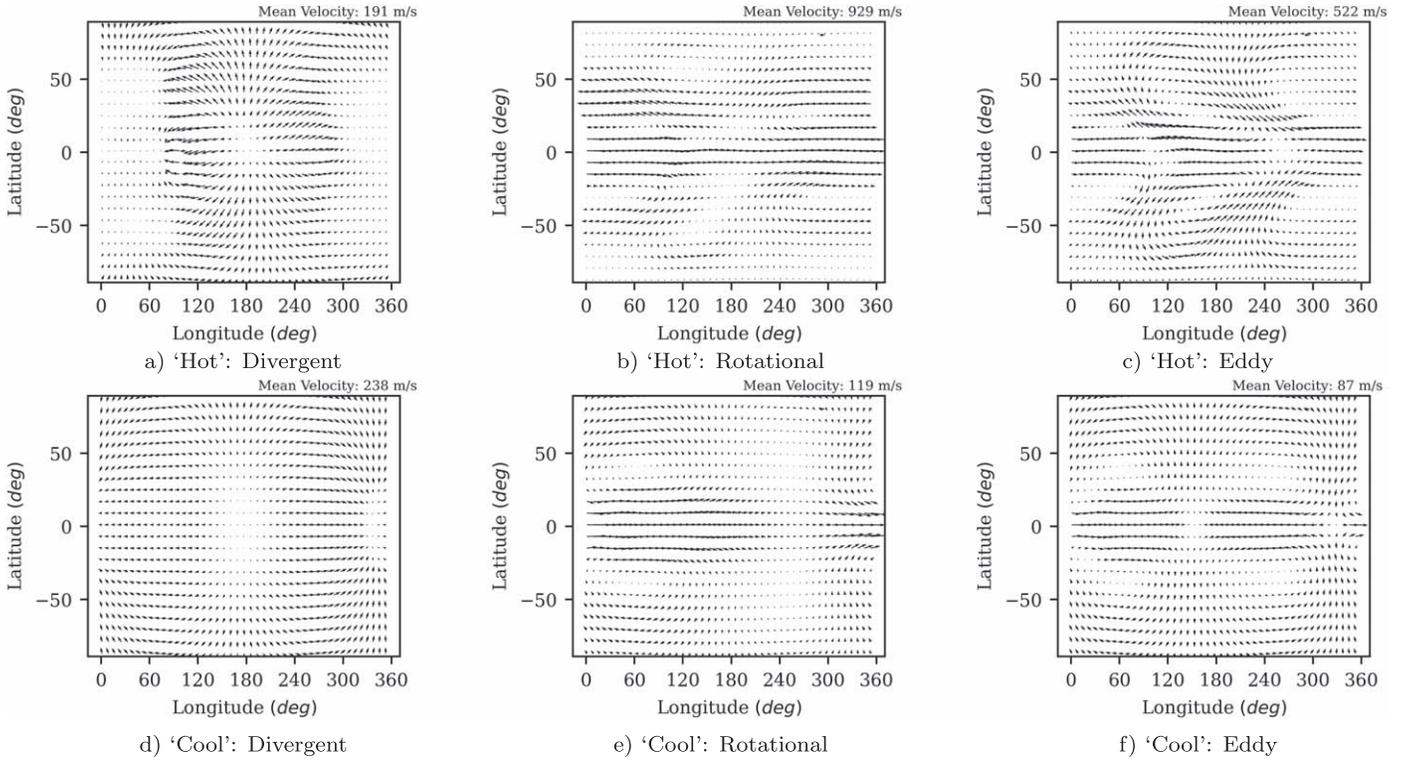
Figure 11. Helmholtz decomposition of the radially and temporally averaged horizontal wind for both our hot (top) and cool (bottom) HD 209458b-like atmospheric models. To the left, we plot the divergent component (u_{div}) of the Helmholtz decomposition, in the middle, the rotational component (u_{rot}), and on the right the eddy component ($u_{\mathrm{eddy}} = u_{\mathrm{rot}} - \langle u_{\mathrm{rot}} \rangle$) of the rotational component of the wind. A video version of this plot is available in Figures 12 and 13.

$$u_{\mathrm{eddy}} = u_{\mathrm{rot}} - u_{\mathrm{zonal}}, \qquad (10)$$

where $\langle \rangle$ indicates the zonal mean.

As for what these components represent: u_{div} represents flows that diverge from the hot spot on the dayside and converge on the cold nightside, forming a closed cycle when combined with the upwelling below the dayside hot spot and the downwelling on the nightside; u_{rot} represents dynamics driven by angular momentum transport via stationary Rossby and Kelvin waves—in typical hot Jupiters these standing waves transport angular momentum from midlatitudes to the equator, resulting in slight westward flows at midlatitudes and a super-rotating jet at the equator (Showman & Polvani 2011); and finally, as mentioned above, u_{eddy} and u_{zonal} are used to split u_{rot}, allowing for us to explore the transport by standing waves in cases where the presence of a super-rotating equatorial jet would completely dominate the dynamics.

In Figure 11, we plot u_{div} (left), u_{rot} (center), and u_{eddy} (right), radially averaged over the outer atmosphere, for both our hot (top) and cool (bottom) atmospheric models. We also, online and in Figures 12 and 13, give a 3D view of each component of the horizontal wind for our hot atmospheric model.

Starting with said hot atmospheric model, we can clearly see that by magnitude the rotational component ($|u_{\mathrm{rot}}| = 929$ m s^{-1}) significantly dominates over the divergent component ($|u_{\mathrm{div}}| = 191$ m s^{-1}). When combined with the difference between the rotational and eddy/zonal components of the wind, this suggests that the main driving force behind the horizontal dynamics seen here is, as expected, the presence of a strong equatorial jet. Further, as suggested by Showman & Polvani (2011), and confirmed by both u_{rot} and u_{eddy}, the super-rotating equatorial jet appears to be driven by standing

Rossby and Kelvin waves. More specifically, we find a $m = 1$ standing wave pattern that has become significantly tilted from west to east by a combination of both a strong Coriolis effect at high latitudes (due to rapid rotation), and equatorial, eastward, angular momentum transport. Finally, the divergent component of the wind plays a much more minor role, only transporting energy from the dayside hot spot toward the terminators and poles, with a slight, rotationally influenced, preference for easterly flows, and little to no transport on the cold nightside. This wind balance is typical of hot Jupiters, and leads to the primarily equatorial heat transport that we discuss below, heat transport, which is responsible for many of the atmospheric features detected by our thermal feature CNN, in particular, the advected butterfly.

On the other hand, the wind dynamics in the cool regime are rather different. Not only does the divergent component of the wind ($|u_{\mathrm{div}}| = 238$ m s^{-1}) dominate over the rotational component ($|u_{\mathrm{rot}}| = 119$ m s^{-1}), the difference between the rotational component and the eddy wind is small ($|u_{\mathrm{eddy}}| = 87$ m s^{-1}). Taken together, this suggests that, unlike in the hot, or even classical hot Jupiter, regimes, the primary driver of this regime's horizontal dynamics is a divergent flow of material from the dayside hot spot toward the colder nightside, which dominates over a much weaker equatorial jet. However, the mechanism by which this very weak equatorial jet forms remains the same: u_{rot} and u_{eddy} both reveal an $m = 1$ standing wave pattern, which is slightly shifted west of the substellar point, and which, due to the relatively weak influence that rotation has on the dynamics, is essentially untilted. Finally, is also interesting to note that the divergent component of the wind converges on the nightside just west of the anti-stellar point, nearly exactly where the hot spot can be found in
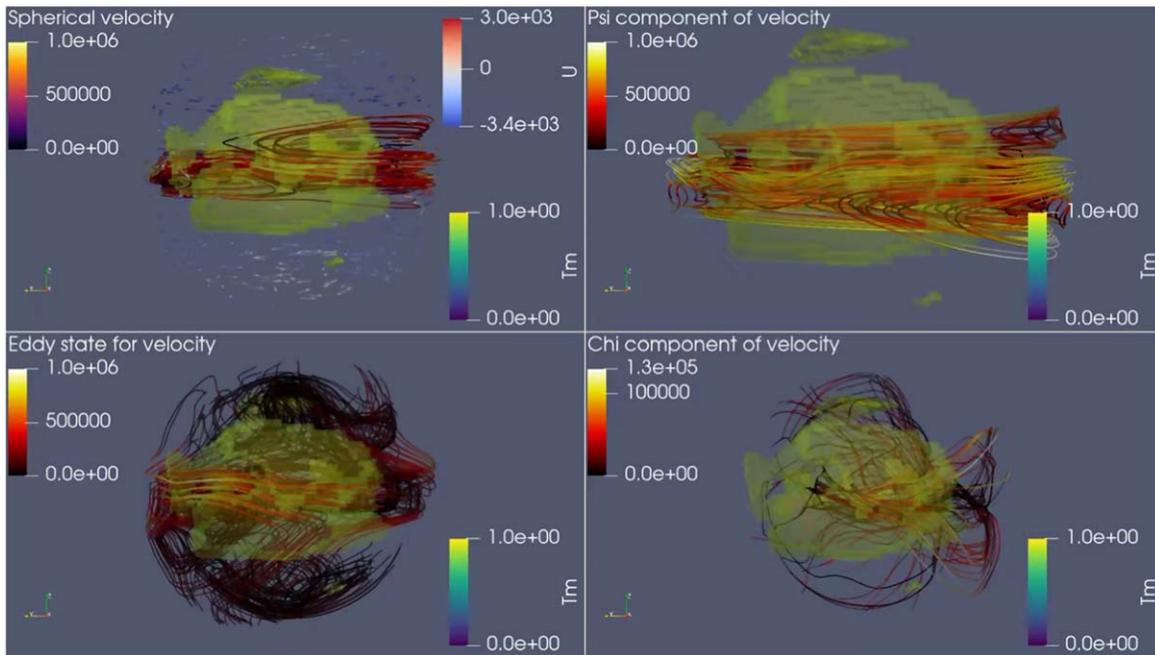
Figure 12. Visualizations of the atmospheric winds and temperature anomalies at an orbital radius of 0.021 au. The main feature of these visualizations are the winds, which we plot using a combination of traced streamlines and arrows. More specifically we plot the horizontal wind in the top left corner, as well as the Helmholtz decomposition of the wind in the remaining panels. By using a Helmholtz decomposition we are able to look at individual contributions to the wind dynamics in more detail. Here the decomposition reveals that the dynamics are primarily driven by the rotational component of the flow (top-right), which itself is primarily driven by Rossby and Kelvin waves pumping angular momentum from midlatitudes to the equator (as reflected by the $m = 1$—two cells per hemisphere—circulation cells). However, rotation (via the Coriolis effect) has significantly affected the waves (and hence the resulting flows), resulting in (a) tilting of the circulation cells, (b) weaker cross-polar flows, and (c) significant asymmetry in both strength and longitudinal scale between easterly and westerly flows from the substellar point. We also find that, due to the Coriolis effect suppressing off-equator flows, that divergent flows (bottom right) only play a minor role in the overall dynamics. Moving on to the temperature profile, this has been normalized such as to emphasize mid-atmosphere temperature anomalies, such as the hot spot on the dayside, which is driven by the stellar irradiation and can be linked to the divergent winds that drive a cooling downflow on the nightside and a heating up-swelling on the dayside. Note: In the visualizations shown and discussed above, the thickness of the atmosphere has been artificially enhanced in order to highlight the atmosphere dynamics at play.

(An animation of this figure is available.)

Figures 6(e)–(g)—as we discuss below, this is not a coincidence.

Given that our two models are in very different dynamical regimes, with very different horizontal wind structures, we next explore how these differences are reflected in the horizontal and vertical energy transport. More specifically, in Figure 14, we explore the zonal, latitudinal, and vertical enthalpy flux transport $E(u, v, w) = \rho * cp * T * \boldsymbol{u}(u, v, w)$ (where ρ is the density, T is the temperature, and cp is the specific heat) at select pressures, which were chosen in order to emphasize the differences in outer atmosphere energetics/dynamics.

Starting in the hot regime, we find that at all but the lowest of pressures where the radiative timescale is very short (and hence advection is suppressed), the zonal-enthalpy transport (e.g., Figures 14(a) and (b)) is dominated by strong eastward advection at the equator and significant westward advection off-equator. This advection can explain the thermal structure identified in the outer atmosphere by the CNNs: zonal advection leads to a shift in the dayside hot spot toward the east at the equator, and toward the west at midlatitudes, leading to the well-known butterfly-like structure (Figures 6(b) and (c)) identified by the CNNs (Figure 8).

Moving on to the latitudinal enthalpy advection (Figure 14(c)), we find that it is strongly correlated with the eddy component of the Helmholtz wind decomposition. For example, by comparing this transport with Figure 11(c), we see that the poleward and equatorward transport align well with the tilted standing wave pattern. This includes a peak in latitudinal

enthalpy transport that occurs near the equator and west of the substellar point, and which corresponds to a similar convergence/divergence point found in the eddy wind component. This correlation is reinforced by the relative magnitude of the latitudinal advection in comparison to the zonal transport—much like the eddy wind being much slower on average than the rotational wind (which includes the zonal jet), the latitudinal heat transport is much weaker than the zonal advection.

Interestingly, this difference in energy transport remains as we move into the deep atmosphere. Here, while we do see some slight signs of an asymmetric thermal structure around 10 bar, deeper than this, the strong vertical heat transport and significant longitudinal mixing has resulted in a deep atmosphere in which *zonal* temperature differences have almost completely vanished (Figure 6(d)), leaving bands of temperature that vary latitudinally, with an enhancement in temperature near the poles due to the off-equator downflows (Figure 14(d)). Note: It is possible, and likely (Sainsbury-Martinez et al. 2019), that with enough time, the deep atmosphere will eventually mix further, reducing the latitudinal temperature differences and resulting in a deep atmosphere that is mostly horizontally homogenized, and hence may be identified by the CNNs as asymmetric if any small-scale, residual, temperature variations are present.

An example of this vertical enthalpy transport, in the equilibrated outer atmosphere, can be seen in Figure 14(d), where red/blue flows indicate upward/downward transport,
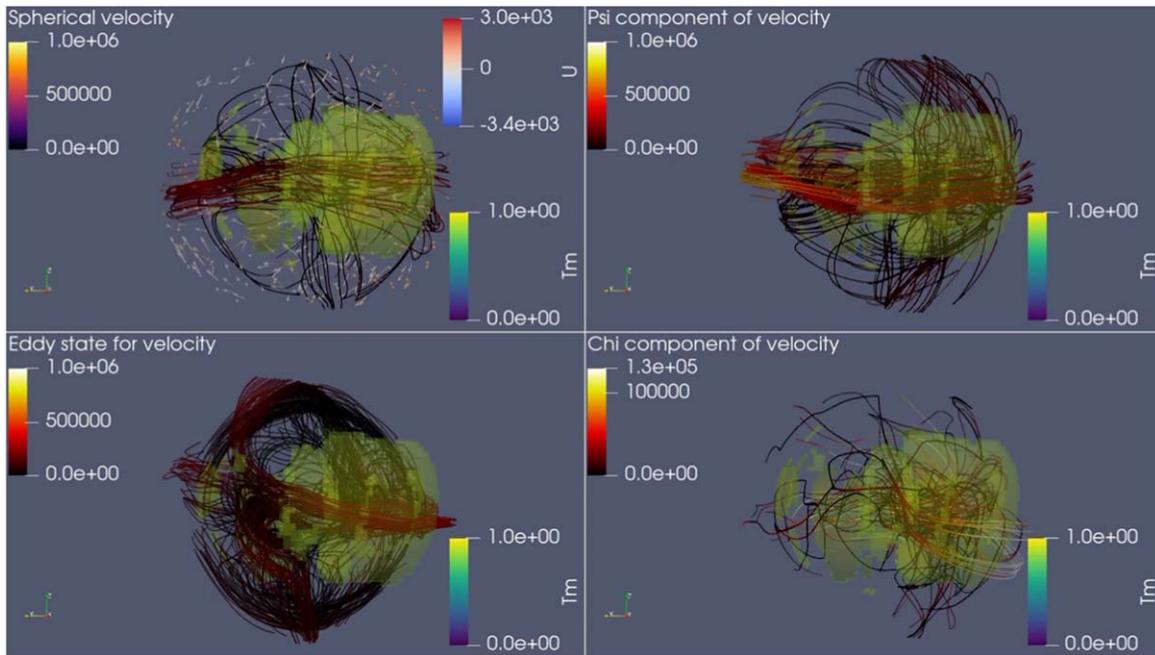
Figure 13. Visualizations of the atmospheric winds and temperature anomalies for a hot Jupiter at an orbital radius of 0.192 au. The main feature of these visualizations are the winds, which we plot using a combination of traced streamlines and arrows. More specifically we plot the horizontal wind in the top left corner, as well as the Helmholtz decomposition of the wind in the remaining panels. By using a Helmholtz decomposition we are able to look at individual contributions to the wind dynamics in more detail. Here the decomposition reveals that the dynamics are primarily driven by the rotational component of the flow (top right), which itself is primarily driven by Rossby and Kelvin waves pumping angular momentum from midlatitudes to the equator (as reflected by the $m = 1$—two cells per hemisphere— circulation cells). Note that, unlike at shorter orbital radii, rotational effects only play a minor role here and so we find both symmetric circulation cells and significant cross-polar flows. We also find that, due to the weaker stellar irradiation at this orbital distance, that divergent flows (bottom right) only play a minor role in the overall dynamics. Moving on to the temperature profile, this has been normalized such as to emphasize mid-atmosphere temperature anomalies, such as the hot spot on the dayside, which is driven by the stellar irradiation and can be linked to the divergent winds that drive a cooling downflow on the nightside and a heating up-swelling on the dayside, or the unexpected hot spot on the dark nightside that inspired this work. Note: in the visualizations shown and discussed above, the thickness of the atmosphere had been artificially enhanced in order to highlight the atmosphere dynamics at play.

(An animation of this figure is available.)

respectively. Here we see a slight upwelling on the dayside that can be linked to the hot spot, surrounded by downwelling near the terminators, on the nightside, and at higher latitudes. This transport extends deep into the atmosphere, growing stronger as the pressure (and hence also the density of the material being transported) increases, explaining the observed deep heating (Figure 5(a)).

We next turn to the cool regime, in which the wind dynamics, and hence also the enthalpy transport, differ significantly from the hot regime. This is best illustrated by the outer atmosphere zonal and latitudinal enthalpy flux, which we plot in Figures 14(e) and (g), respectively. Here we find that the mean zonal and latitudinal enthalpy fluxes are approximately equal, and are very strongly shaped by the divergent component of the wind (Figure 11(d))—with clear transport occurring from the hot dayside to the cooler nightside, both zonally and latitudinally across the poles, converging just west of the anti-stellar point, exactly where the deeper nightside hot spot, and thermal inversion, is found (e.g., Figures 6(e)–(g)). This divergence-driven transport also explains why we do not detect/observe the dayside butterfly that is typically associated with hot Jupiter atmospheres: due to the relatively low influence that rotation plays on the dynamics, the day-night energy transport is primarily, isotropically, divergent, rather than the highly anisotropic (equatorial) transport found for the hot regime. As such the temperature structure in the outer atmosphere remains largely *locked* (and maybe a little broadened) until enough energy has been transported by the

divergent flows to form a nightside hot spot (hence changing the tag applied by the CNN—see Figure 9(b)).

Moving deeper into the atmosphere, we start to see the impact that the nightside hot spot has on the mid-atmosphere horizontal enthalpy transport. Here we find that the zonal-enthalpy flux (Figure 14(f)), much like the rotational component of the zonal wind (Figure 11(e)), is dominated by an eastward flow from the nightside hot spot toward the relatively cool dayside. This suggests that the divergent flows are dominant in the outer atmosphere, where the day-night forcing is strong, and *rotational* flows are dominant in the mid-atmosphere, where the forcing has switched to being driven by the nightside hot spot, reinforced by the eddy winds (Figure 11(f)).

Finally, we come to the vertical enthalpy transport (see, for example, Figure 14(h)), which is weaker than the vertical energy transport found in the hot regime, as well as being more vertically confined. More specifically, we find that the main region of strong downward enthalpy transport is focused on the nightside around the hot spot, and like the hot spot itself, this downward transport does not extend into the deep atmosphere. Instead, we find that vertical mixing is weak in the deep atmosphere, helping to explain the lack of observed deep heating (Figure 5(b)) and hence radius inflation. In fact, mixing is generally weak in all directions in the deep atmosphere, although it is slightly stronger in the zonal direction than vertically or latitudinally. This helps to explain the slight asymmetry seen in the deep atmosphere (and identified by the
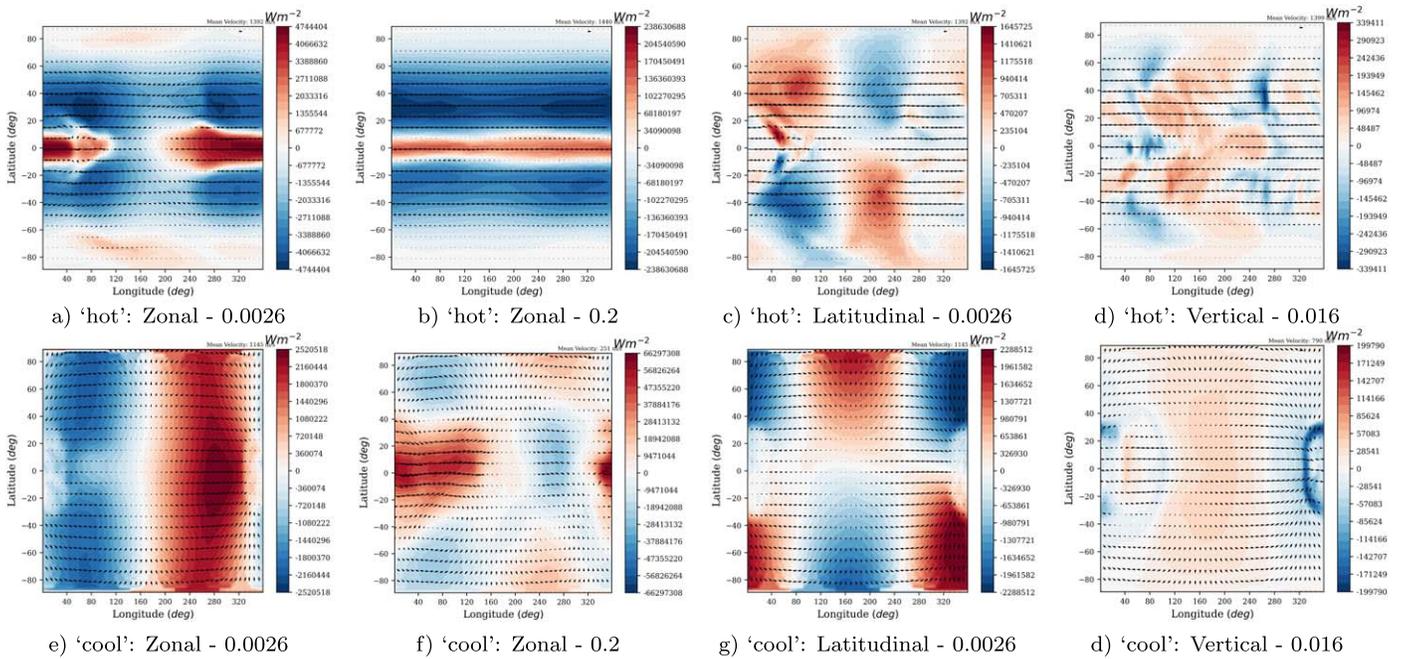
Figure 14. A selection of maps showing the temporally averaged zonal, meridional, or vertical enthalpy transport at selected pressure levels for our exemplary hot (top) and cool (bottom) HD 209458b-like atmospheric models. Here, positive (red) fluxes represent eastward/polar/outward flows for the zonal/meridional/vertical enthalpy flux maps, respectively. Note: We include two maps for the zonal-enthalpy transport, at different pressure levels, in order to emphasize how the zonal advection changes with both height and orbital radius.

thermal CNN). Small amounts of thermal energy are transported to the deep atmosphere by vertical mixing (in an essentially random way since the flow is so weak), leading to slight temperature variations that become smoothed out longitudinally but not latitudinally—hence leaving us with a weak asymmetric thermal structure, as can be seen in Figure 6(h). Note that, unlike in the hot regime above, the particularly slow dynamical timescales of the deep atmosphere found here mean that we do not expect complete horizontal homogenization to occur on any reasonable timescale.

3.4.1. Difficulties, Warnings, and Advice for Using CNNs to Analyze Atmospheric Dynamics

While the above results show some promising outcomes of pairing DL neural networks with (atmospheric) simulations, this process is not without its own problems and limitations, which we discuss here.

One particular important factor that can impact the ability of CNNs to detect atmospheric features is the choice of color map used to visualize the data since this can significantly impact the ability of the neural network to detect the edges and gradients that are key to encoding detectable features during the training process. As part of this work, we tested training the thermal, and wind, CNN with every available color map included as part of Matplotlib, using training data that was, other than the color map, identically tagged. While the results were similar for many of the color maps chosen, a few stood out both positively and negatively. Note that the effectiveness of each set of variable color map networks was evaluated using their confusion matrices, allowing for a direct comparison of the ability of each set of models to reproduce the set of known tags.

For example, the color map jet, which used to be the default map used in Matplotlib, and which remains a staple of astrophysics research to this day, resulted in particularly poor detection and characterization of atmospheric features. This occurs for exactly the same reason why it was replaced as the default color map in Matplotlib: due to rapid color and brightness changes (i.e., the color map is not perceptually uniform), the jet tends to both exaggerate and suppress small changes/gradients in the underlying data—while this can be an advantage when quickly inspecting data visually (but even this is in doubt), it also significantly hampers the ability of the multi-categorization CNN to detect and isolate atmospheric features, either through mistraining, because the feature is hidden, or even because the color map has warped the morphology of the feature by exaggerating/suppressing gradients.

Interestingly, even the modern replacements, such as Viridis, Inferno, or Plasma, which were designed to solve this nonperceptional uniformity problem are also poor choices for use with a CNN. This can be linked to their use of a large array of colors, which again can act to disguise, morph, and exaggerate/suppress features, as proved by the best color maps for use with image-recognition algorithms: monochromatic color maps.

Simply put, by using a monochromatic color map (such as grays, reds, Gist_gray, etc.), simulations with the same atmospheric feature (but with different magnitudes and horizontal gradients, due to, for example, differences in surface irradiation strength) are more likely to produce outputs that are visually near identical, making it easier for a CNN to both learn and detect said features. Furthermore, the use of monochromatic color maps can slightly reduce the memory footprint of the CNN by reducing the initial data set size by a factor of 3, from RGB to grayscale, thus, theoretically, enabling either faster data processing while the simulations run, or larger and more complex CNNs (with either more complex layers or more layers generally) to be developed in the same memory footprint. However, as detailed in the Appendix, the

CNNs we consider here are already fairly lightweight when compared to those used for, say, facial recognition, and so the possibilities for computational savings are minimal—for example, the *high-resolution*, full RGB, network only take 4 minutes to train on four K80s (which themselves are relatively old). Furthermore, testing with training our thermal CNN using normalized raw data rather than images produced with Matplotlib resulted in only a small reduction, 0.3% for our thermal CNN, in the size of the neural network. This occurred at the cost of introducing significant problems with the optimizer leading to significant oscillations during the validation phase of training the network. While fixing such issues should be possible, it would require modifications to the structure of the network and individual layers, complicating issues for a non-AI expert all for a minimal reduction in computational cost and very limited potential improvement over a well-chosen color map.

However, this all comes at the cost of making the data harder to interpret visually/manually. The compromise, in terms of generating data that is suitable for use with both CNNs and humans, albeit at the cost of not reducing the memory footprint, is to use a diverging color map, i.e., a color map with only two primary colors that diverge from a neutral color, such as white, at either a fixed point (e.g., zero wind) or at the data mean, as used in, for example, Figure 6. These color maps result in similar accuracy to monochromatic color maps for the thermal CNN when trying to recover training data, while also generating plots with which humans can easily visualize atmospheric features/dynamics.

Note that, while the above results are fairly robust, we advise that future studies, particularly those that intend to use multi-categorization CNNs to analyze data unsupervised (i.e., during a simulations runtime), should fine-tune both the color maps as well as the data boundaries in order to ensure suitability before committing to the run.

In addition to ensuring that the data is prepared in such a way as to be suitable for use with a CNN, we also have to ensure that the features being searched for are also detectable and are properly trained for.

This proved to be an issue here. As discussed in Section 3.2, our original plan included using a separate CNN to explore the presence of zonal jets/horizontal winds in our simulation data. This proved to be highly intractable for a variety of reasons. To start, we initially tried to train the CNN by using the, overlaid, arrow quivers shown in Figure 6. However, this ended up being nearly impossible since, at the resolution of the CNNs' analysis, the arrows were essentially undetectable. Furthermore, even when we adjusted the plots to make the quivers more visible/ bolder, or increased the resolution of the initial layers of the network, small changes in the horizontal wind structure led to mis/failed detection of the jet—it appears that the localized nature of the quivers not only made the CNN very sensitive to changes in their structure but also meant that the model struggled to detect large-scale structure in the first place. It is possible that a significantly larger training data set, larger quivers, higher resolution images, or a different kernel size might solve this issue.

Rather than testing the aforementioned ideas, which we leave to future work, we instead tried to avoid the quiver issue completely by replacing them with horizontal wind maps, much like those used with the thermal CNN. This greatly improved the ability of our CNN to detect the zonal winds, but

it also revealed two additional limitations. First, the zonal wind is not as horizontally homogenized as the zonal-mean zonal wind would suggest, thus making it trickier for the CNN to detect unless very carefully trained with a wide variety of zonal wind structures, and second, even when carefully trained, the CNNs can have a difficult time detecting zonal jets due to their highly symmetric structure (similar difficulties are also faced when searching for the banded structure with the thermal CNN; however, it is less of a problem there as alternative structures, which may lead to misidentification, are not as prevalent due to the relative quiescence of the deep atmosphere). More specifically, our CNNs can find it more difficult to isolate a band of temperature/wind than a more complex structure, such as a nightside hot spot or a thermal butterfly. This is because said complex structures have more features (edges) for the neural network to latch onto and learn, increasing the complexity of the neural network and hence its ability to internalize features. As a consequence of all of the above, we have not focused on the detection of zonal winds in this work, and we suggest that future studies that wish to implement data analysis via neural networks should focus on more detailed and/or derived atmospheric features, ranging from thermal structures to complex wind dynamics, such as vortices or (e.g., by using a Helmholtz decomposition) standing waves (although here we again caution against using quivers due to their localized nature).

Another example of the possible difficulties faced when our thermal CNN searched for atmospheric features is the nondetection of the thermal butterfly structure at later times in our exemplary hot regime model (Figure 8). Comparisons between the training data, which was based upon early outputs of our full simulation set, and the steady-state outer atmosphere thermal profiles reveal the simple reason why this is the case: the strong equatorial jet and off-equator, low-latitude, counter-flows have resulted in a butterfly structure that is much more longitudinally extended/stretched than any profiles included in the initial training data. As such, our trained CNN was unable to properly identify the evolved/advected feature. This serves as an important caveat of CNNs and neural networks in general: they are only as effective as their training data, and are not able to identify new or highly evolved/warped features. However, as previously discussed, this lack of detection itself can also be an interesting result, identifying new, uncommon, or unexpected features. Yet this is of cold comfort when trying to use a CNN, or AI more generally, to concurrently process an ongoing simulation—here the best solution is a broad training set that contains multiple examples of all the features expected. For example, now that we have steady-state training data for a wide variety of HD 209458b-like atmospheric models at different orbital radii, and with equilibrium outer atmospheres, it should be possible to build CNNs to accurately, and quickly, analyze future simulations within, or near, this parameter space on the fly.

4. Discussion and Conclusion

In this paper, we have explored the role that AI-driven image classification, via the use of CNNs, can have in the concurrent- and post-processing of simulations of planetary atmospheres, specifically HD 209458b-like hot Jupiters at various orbital radii.

17

4.1. Model and AI (CNN) Setup

To that end, we started by running a series of HD 209458b-like atmospheric models with different orbital radii, and hence different surface irradiation and synchronous planetary rotation rates. The orbital radii considered here varied between 0.012 and 0.334 au, but for the sake of brevity, we chose to focus on models at two different orbital radii, with dynamics that are characteristic of their contemporaries. Specifically, we focused on one model with an orbital radius of 0.021 au, which we refer to as our hot model, and one with an orbital radii of 0.192 au, which we refer to as our cool model. These simulations are based on those first presented in Sainsbury-Martinez et al. (2019), but modified such that their outer atmosphere temperature forcing is derived from 1D models calculated using ATMO at every orbital radius of interest. From the first outputs of these simulations, we then selected and labeled a number of thermal and wind atmospheric features that we wished for our AI model to detect/characterize. These features included the presence of a tidally locked dayside hot spot, a horizontally advected dayside hot spot (better known as a butterfly-like thermal structure), a latitudinally asymmetric deep atmosphere, a deep atmosphere that is fully longitudinally homogenized, and in which latitudinal temperature variations remains small (we refer to as banded), and the presence of a super-rotating equatorial jet, although this latter tag/feature proved to be difficult to identify.

This hand-labeled data set was then fed into a pair of multi-categorization CNNs—i.e., a neural network that is particularly suited to image-recognition tasks (see Section 2.2 for a more detailed description) and detecting multiple features nonexclusively. Once trained, this neural network was then applied to the full time series outputs of our exemplary simulations (which had been run to a steady state in the outer atmosphere—i.e., at lower pressures).

4.2. Identified Features, or Lack Thereof, with CNNs

Applying the trained CNNs to our exemplary models revealed our first key result: at higher orbital radii, i.e., in the cool regime, the thermal CNN multi-classification map (which shows the identified features versus both pressure and time) contained a region with no identified atmospheric features. The resulting analysis of this model, as well as other cool regime models, revealed a mid-pressure atmosphere (i.e., $0.05\,\text{bar} \rightarrow 1\,\text{bar}$) that was behaving in a rather unusual way: the hottest region of the atmosphere had shifted from the irradiated dayside to the nightside, ending up just west of the anti-stellar point. Once this data was added to the thermal CNN (using interpolative oversampling to generate artificial training data, a necessity due to the small sample size available as this feature only occurs in the cool regime and over a limited pressure range), we were able to successfully identify this hot spot, and its associated thermal inversion in all models with orbital radii>0.11 au—i.e., all cool regime models.

This is just one example of the differences observed between models in the cool and hot regimes, differences that extend throughout the model atmospheres and which appear to be highly linked to the mixing/transport/circulation regime that the models fall into. Only our analysis of the low-resolution and adiabatically initialized HD 209458b model of Sainsbury-Martinez et al. (2019) revealed every original (i.e., excluding the nightside hot spot) feature, which was to be expected since

it was HD 209458b's dynamics that formed the basis for the original features selected for detection.

Starting in the cool regime, the identified features tended to correspond to weaker mixing and anisotropic horizontal energy transport—that is to say that the outer atmosphere dynamics remain highly radiatively forced (despite the relatively weak stellar irradiation), the mid-atmosphere is dominated by isotropic (i.e., divergent) energy transport from the dayside to the nightside, which eventually leads to the formation of a nightside hot spot and associated thermal inversion, and the deep atmosphere is highly quiescent with weak mixing and deep heating that allow for slight latitudinal temperature gradients to develop and be maintained, which the CNN identifies as an asymmetric thermal structure.

On the other hand, for models that fall into the hot regime, we found that the identified dynamics correspond to strong zonal energy transport and significant horizontal, and vertical, mixing. For example, we find that in the outer atmosphere, once the radiative timescale is long enough (i.e., not at very low pressures), the dayside hot spot becomes significantly horizontally advected by both the equatorial jet as well as the associated midlatitude counterflows, resulting in the well-known butterfly-like thermal structure on the dayside, with the exact shape depending upon the strength of the rotational influence, and hence the jet structure. This strong advection/mixing extends to the deep atmosphere, where not only do we find significant heating due to vertical potential temperature (enthalpy) transport, but also that the deep horizontal advection, which is strongest in the longitudinal direction, has resulted in strong zonal homogenization paired with a weak latitudinal temperature gradient—this is the temperature structure we refer to as banded.

4.3. Understanding the Different Dynamical Regimes

In order to try and understand the differences between these two dynamical regimes, and also how the unusual nightside hot spot and thermal inversion forms in the cool regime, we next explored the wind and energy transport (specifically enthalpy flux) in more detail.

Starting with the wind, we use a Helmholtz decomposition to split the horizontal wind into its divergent (i.e., vorticity-free), rotational (i.e., divergence-free), and eddy (i.e., perturbations to the rotational wind) components. This reveals that, as expected, the wind dynamics differ significantly between the cool and hot regimes.

In the hot regime, strong stellar irradiation and rapid surface rotation mean that the wind dynamics are dominated by the rotational component of the wind. This is in agreement with Showman & Polvani (2011), who suggest that the rotational component of the wind is correlated with standing Rossby and Kelvin waves that in turn drive a strong, highly advecting, equatorial jet: here the rotational component of the wind reveals both a strong equatorial jet as well as a significant, $m = 1$, standing wave pattern.

On the other hand, in the cool regime where the irradiation is weaker and the surface rotation is slower, the winds fall into a completely different dynamical regime: the divergent component of the wind dominates over the rotational component, although signs of the latter remain, exhibiting a weak, but stable, $m = 1$, standing wave pattern that fails to drive a significant equatorial jet. As for the dominant divergent wind, this flows, isotropically, from the substellar point to the

unirradiated nightside, converging just west of the anti-stellar point. i.e., exactly the same location as the mid-atmosphere nightside hot spot.

The horizontal enthalpy transport reveals that this is not a coincidence. For instance, our cool model reveals zonal and latitudinal enthalpy transport that is highly shaped by the divergent component of the wind, and hence, also converges just west of the anti-stellar point. This explains the formation of the nightside hot spot at mid-pressures: in the very outer atmosphere, radiative forcing is too strong for significant advection to occur; however, as we move deeper, the radiative timescale lengthens and day-night advection starts to occur (isotropically due to the wind structure), leading to the formation of the nightside hot spot. This hot spot then dominates the mid-atmosphere energy transport, leading to the unusual scenario that is night-day heat transport. However, relative to the transport seen in HD 209458b, or the hot regime, this transport is weak and does not extend into the deep atmosphere. This is reflected in the vertical enthalpy transport profile, which reveals that vertical advection is focused on maintaining the nightside hot spot, leading to little to no deep heating and a quiescent deep atmosphere. This link between the nightside hot spot and the divergent wind is further reinforced by more rapidly rotating cool regime models: as the influence that rotation has on the atmosphere increases, the location of both the hot spot, as well as the divergent wind convergence point shifts westward, likely as a result of the slight tilt introduced to the divergent wind by off-equator Coriolis forces.

A complementary result is found in the hot regime, although here, as discussed above, the enthalpy transport is controlled by the rotational wind, resulting in transport that is primarily driven by the zonal jet, with weaker off-equator transport linked to the $m = 1$ standing wave pattern. As such, we find significant horizontal advection near the equator, either eastward where the jet dominates, or westwards off-equator where the standing wave-driven transport is strongest. Taken together, this transport results in the formation of the synonymous butterfly-like thermal structure. At higher latitudes, we also find evidence for the influence that Coriolis forces have on enthalpy transport, with a significant westward tilt developing as we move toward more rapid rotation. As for the off-equator longitudinal and latitudinal enthalpy transport, this is highly correlated with the same Rossby and Kelvin standing wave pattern that drives the equatorial flows, including the significant westward tilt as we move to higher latitudes, a tilt that is highly dependent upon the planetary rotation rate. Note, however, that at very high rotation rates, off-equator heat transport is suppressed as the Coriolis force suppresses higher latitude winds. Finally, as suggested by Sainsbury-Martinez et al. (2019), the strong zonal wind, and hence zonal-enthalpy transport, that develops in this model results in significant vertical heat transport that extends from the outer atmosphere all the way to the bottom of the simulation domain, increasing the entropy of the internal adiabat, and hence driving significant radius inflation (as shown in Figure 5(a)).

4.4. Limitations and Advice for Future Pairings of CNNs with Atmospheric Models

While our above results show some promising outcomes of pairing neural networks with atmospheric simulations, they also reveal some of the limitations of this approach, as well as potential pitfalls and avenues for improvement.

To start, CNNs cannot detect any features for which they are not robustly trained. For example, our initial training and validation data set did not include the uncommon nightside hot spot found in the cool regime, and thus, when we fed the data into our networks, no tags were assigned to the mid-atmosphere. More subtly, while we did train our thermal CNN to detect butterfly-like features, at rapid rotation rates the strong zonal jet and latitudinally compressed midlatitude counterflows resulted in a butterfly-like thermal structure that was notably different from those included in our training set. Consequently, the thermal CNN was only able to assign the butterfly tag in our exemplary hot model near initialization, before advection had significantly changed its structure.

Both of these examples emphasize how important a robust training set is when using neural networks (and CNNs in particular) to analyze data. This is doubly true when trying to use a neural network to concurrently analyze and process a running model, since a result that is only revealed through further study of an unusual region of nonidentification occurs too late to be of use in deciding which runs to continue, which to discard, or which have reached equilibrium.

However, this does not mean that CNNs (and neural networks more generally) cannot be used for concurrent processing. One area for which they are particularly well suited is when paired with next-generation exascale super-computers. These next-generation machines will allow for incredibly high-resolution, and long-timescale, simulations, albeit at the cost of vast amounts of computational resources that will be in high demand (and which carry a high cost both financially and environmentally). As such it is critical to ensure that the allocated computational resources are being used efficiently, as well as minimizing researcher time required to analyze the vast outputs of these models. Fortunately, as part of the process of designing an exascale-calculation, it is typical to run a series of lower-cost (i.e., lower resolution or shorter timescale), preliminary, simulations in order to both get a sense of the value/significance of a very large-scale simulation, as well as the exact model parameters to use (for example, with DYNAMICO, it is import to recalibrate the diffusion timescale since the models hyperdiffusion is resolution dependent). These preliminary models might provide a source of training data that can be used to generate a set of CNNs to analyze the final, production, simulation. In a similar vein, with our series of outer atmosphere equilibrium HD 209458b-like models at different orbital radii, we now have enough data to train a robust thermal-data CNN to analyze any models we want to run at intermediary orbital radii. For example, we showed that the networks trained here on our *high*-resolution, HD 209458-like models, could be applied to analyze the lower-resolution HD 209458b models of Sainsbury-Martinez et al. (2019). Furthermore, the validity of such a technique could be further verified, in future studies, by analyzing the CNNs with a post-processing tool such as GradCAM (Selvaraju et al. 2017), which highlights important features in a sample image, allowing for confirmation that the CNNs predictions can be trusted.

4.5. Future Perspectives for Atmospheric Modeling

As we have alluded to above, the primary differentiator between the two regimes discussed here appears to be the relative influence that rotation has on the dynamics, specifically on the horizontal wind structure. Furthermore, understanding

these differences and their impacts may prove crucial to our understanding of hot Jupiters in longer orbits, orbits that are now becoming accessible to observations to next-generation telescopes, such as JWST or the Transiting Exoplanet Survey Satellite. One example of this is the unusual feature we detected in the atmospheres of our more slowly rotating hot Jupiters models, a nightside hot spot, which was first detected via our thermal CNN, thus reinforcing the value of pairing long-timescale, and computationally complex models with trained networks for both concurrent- and post-processing. If this feature proves to be robust, it may have implications for our understanding of hot Jupiter atmospheric chemistry. For example, a thermal inversion on the nightside may significantly impact the distribution of chemical compounds by acting as a cold trap in which denser materials condense, essentially raining out of the outer atmosphere, thus becoming depleted.

As a result, we have a number of suggestions for future studies. First, we strongly encourage any future studies with high-resolution and long-timescale simulations to consider pairing them with an AI model, if only to reduce the initial analysis burden, with the AI helping to identify regions of interest or uncommon dynamics. Second, we propose a more in-depth study isolating how rotation alone affects irradiated atmospheric dynamics, with a particular focus on day-night winds and energy transport in slowly irradiated Jupiters, i.e., on the unusual nightside hot spot. If this result still proves to be robust, we further suggest that this work be followed up with a next-generation GCM, which includes both nonequilibrium chemistry and robust radiative dynamics, to fully explore what possible impacts that a nightside hot spot may have on observable dynamics.

Appendix

Following and expanding upon the work of Lagerquist et al. (2019), we implement a pair of multi-categorization, light-weight, CNNs, one for thermal features and one for horizontal wind features. Here, as in the paper, we focus our discussion on the model designed for thermal feature detection.

This model follows the structure shown in Figure 15 (and Figure 2 of Lagerquist et al. 2019). After input, the image data is analyzed and reduced by a series of four convolution blocks of decreasing dimensionality and increasing complexity, as shown in Figure 15. Each convolution block consists of three layers: the 2D convolution layer itself, which has a kernel size of 5×5 and includes $16/32/64/64$ filters in the first/second/third/fourth blocks, respectively, a 2D max pooling layer that downsamples the data, halving the resolution in both dimensions (i.e., latitudinally and longitudinally), and finally, a dropout layer, which is only active during training, and which sets a fraction of the input arrays values to 0 (rescaling the remainder of the array), in order to reduce overfitting. During the convolution process, this fraction is set to 0.25, while in the fully connected layers, the fraction is increased to 0.5. Note that, in the convolution layers, the 5×5 kernel represents the matrix used to enhance features, i.e., detect edges. This kernel is swept across the entire input moving horizontally and vertically with a stride length of 1, hence reducing the dimensionality of the input by 4 in both the horizontal and vertical directions (since we do not include any padding). The exact form that this kernel matrix takes is a result of the training process undertaken, with the kernel being optimized to recover the feature of interest. Of course, since we are looking for multiple features and since these features vary spatially, running a single kernel per convolution layer would be highly inefficient. Instead, we consider and train multiple kernels per layer, with the number of kernels, referred to as the number of filters increasing as the dimensionality (and hence size) of the data set decreases, i.e., 16 filters in our first convolution block and 64 in the last.

Once the convolution blocks have processed the data, and the total dimensionality has been reduced to 7×18 with 64 filters (i.e., $7 \times 18 \times 64$), the data set is small enough that, after flattening (so that the dense layers connect all points), it can be fed into a series of *low-resolution* fully connected blocks that consist of fully connected dense layers, followed by overfitting reducing dropout layers. The final dense layer returns eight values corresponding to the probability of detection of each of the trained features. This process is repeated for every pressure level and every time-averaged point in order to generate the multi-categorization maps shown throughout this paper. It is important to note that, even after the dimensionality reduction associated with the convolution process, the first dense layer, which fully connects the 8064 points in the flattened output of the convolution process with only 128 output points, contains over a million weights, although this is small compared with the almost 25 million weights that would be required to fully connect the initial image with a similar sized output (a process that is likely to lead to spurious outputs due to the massive single-step decrease in dimensionality from 194,000 to 128 points!).

In addition to the above, it is important to note, for reproducibility, that (a) other than the final fully connected dense layer (which uses a sigmoid function to generate the final probabilities), all the neural network layers considered here include rectified linear unit activation, which is commonly used in CNNs, is believed to improve the efficiency of deep learning (Nair & Hinton 2010), and which essentially works by zeroing out any negative values in the output of the associated neural network layer. (b) The learning rate of the model was 0.001 and it made use of the Adam optimizer. (c) The initial resolution of the network discussed here is half that of the input, print quality, and image file; however, tests with higher initial
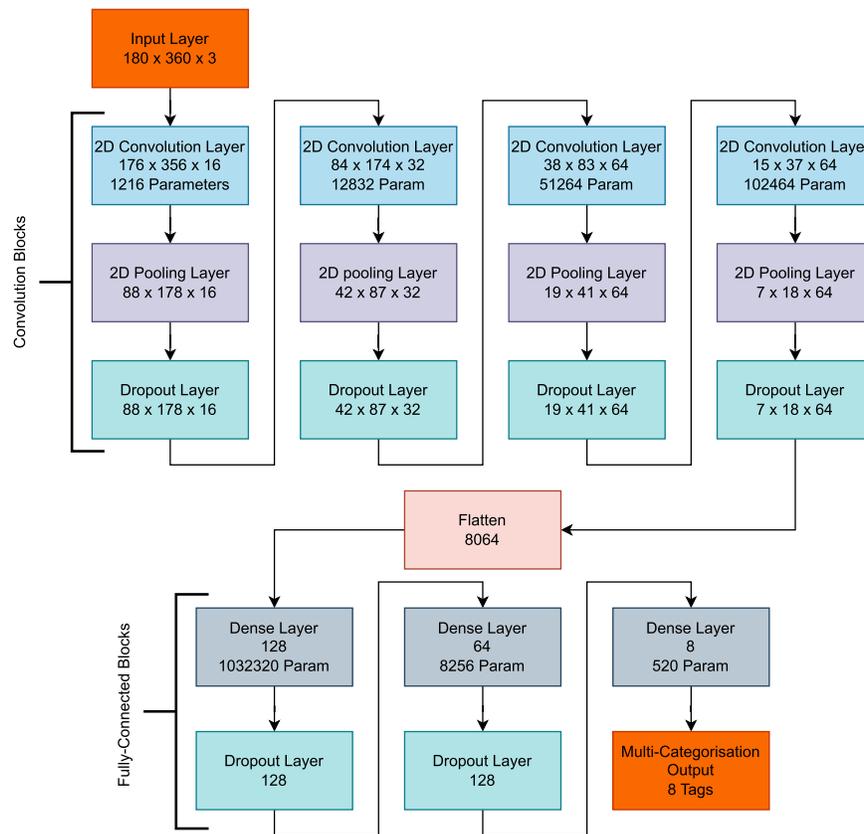
Figure 15. Flowchart showing the layers and blocks of a half resolution (initial image resolution is 180×360, half that produced by Matplotlib for print quality figures) image-recognition network designed to detect the six features of interest in the thermal structures of our HD 209458b-like atmospheric models. Note that in each box we provide the name of the layer, the output resolution of the layer, as well as the number of trainable parameters, when nonzero.

resolutions were performed, and while the accuracy slightly improved, the computational cost significantly ballooned. And (d) the training/validation data split was 80%/20% with pseudo-random assignment between the two categories (that is to say we made use of the random state parameter to ensure that the train/test split was consistent between models).

ORCID iDs

F. Sainsbury-Martinez ⬤ https://orcid.org/0000-0003-0304-7931
P. Tremblin ⬤ https://orcid.org/0000-0001-6172-3403

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, http://download.tensorflow.org/paper/whitepaper2015.pdf
Adams, S. V., Ford, R. W., Hambley, M., et al. 2019, J Parallel Distrib Comput, 132, 383
Albawi, S., Mohammed, T. A., & Al-Zawi, S. 2017, in IEEE Int. Conf. on Engineering and Technology (ICET) (Piscataway, NJ: IEEE), 1
Amundsen, D. S., Mayne, N. J., Baraffe, I., et al. 2016, A&A, 595, A36
Beatty, T. G., Madhusudhan, N., Tsiaras, A., et al. 2017, AJ, 154, 158
Böker, T., Arribas, S., Lützgendorf, N., et al. 2022, A&A, 661, A82
Castelli, F., & Kurucz, R. L. 2003, in IAU Symp. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray (San Francisco, CA: ASP), A20
Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJL, 529, L45
Cullen, M., & Brown, A. 2009, RSPTA, 367, 2947
Deitrick, R., Heng, K., Schroffenegger, U., et al. 2022, MNRAS, 512, 3759
Drummond, B., Tremblin, P., Baraffe, I., et al. 2016, A&A, 594, A69
Dubos, T., Dubey, S., Tort, M., et al. 2015, GMD, 8, 3131
Dubos, T., & Voitus, F. 2014, JAtS, 71, 4621
Dutton, J. 1986, The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion (New York: Dover)
Ferruit, P., Jakobsen, P., Giardino, G., et al. 2022, A&A, 661, A81
Fortney, J. J., Lodders, K., Marley, M. S., & Freedman, R. S. 2008, ApJ, 678, 1419
Fukushima, K., & Miyake, S. 1982, in Competition and Cooperation in Neural Nets, ed. S.-i. Amari & M. A. Arbib (Berlin: Springer), 267
Gandhi, S., & Madhusudhan, N. 2019, MNRAS, 485, 5817
Guerlet, S., Spiga, A., Sylvestre, M., et al. 2014, Icar, 238, 110
Guillot, T., & Showman, A. P. 2002, A&A, 385, 156
Hammond, M., & Lewis, N. T. 2021, PNAS, 118, e2022705118
Hubeny, I., Burrows, A., & Sudarsky, D. 2003, ApJ, 594, 1011
Jakobsen, P., Ferruit, P., Alves de Oliveira, C., et al. 2022, A&A, 661, A80
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems, ed. F. Pereira et al., Vol. 25 (Red Hook, NY: Curran Associates Inc.), 1106
Lagerquist, R., McGovern, A., & Gagne, D. J., II 2019, WtFor, 34, 1137
LeCun, Y., Boser, B., Denker, J. S., et al. 1989, Neural Comput, 1, 541
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proc. IEEE, 86, 2278
Lee, E. K. H., Parmentier, V., Hammond, M., et al. 2021, MNRAS, 506, 2695
Lothringer, J. D., Barman, T., & Koskinen, T. 2018, ApJ, 866, 27
Madhusudhan, N., & Seager, S. 2010, ApJ, 725, 261
Maynard, C., Melvin, T., & Müller, E. H. 2020, QJRMS, 146, 3917
Mayne, N. J., Baraffe, I., Acreman, D. M., et al. 2014a, GMD, 7, 3059
Mayne, N. J., Baraffe, I., Acreman, D. M., et al. 2014b, A&A, 561, A1
Mayne, N. J., Drummond, B., Debras, F., et al. 2019, ApJ, 871, 56
Miesch, M. S. 2005, LRSP, 2, 1
Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47
Nair, V., & Hinton, G. E. 2010, Proc. Int. Conf. Machine Learning 27 (ICML 2010), ed. J. Fürnkranz & T. Joachims, (Madison, WI: Omnipress), 807, https://icml.cc/Conferences/2010/papers/432.pdf

O'Shea, K., & Nash, R. 2015, arXiv:1511.08458

Pontoppidan, K., Blome, C., Braun, H., et al. 2022, ApJL, 936, L14

Ranjan, S., Charbonneau, D., Deming, D., et al. 2011, AAS/ESS Meeting, 2, 12.06

Rauscher, E., & Menou, K. 2010, ApJ, 714, 1334

Sainsbury-Martinez, F., Casewell, S. L., Lothringer, J. D., Phillips, M. W., & Tremblin, P. 2021, A&A, 656, A128

Sainsbury-Martinez, F., Tremblin, P., Schneider, A. D., et al. 2023, MNRAS, 524, 1316

Sainsbury-Martinez, F., Wang, P., Fromang, S., et al. 2019, A&A, 632, A114

Schneider, A. D., Carone, L., Decin, L., Jørgensen, U. G., & Helling, C. 2022, A&A, 666, L11

Selvaraju, R. R., Cogswell, M., Das, A., et al. 2017, in IEEE Int. Conf. on Computer Vision (ICCV) (Piscataway, NJ: IEEE), 618

Showman, A. P., Cooper, C. S., Fortney, J. J., & Marley, M. S. 2008, ApJ, 682, 559

Showman, A. P., & Polvani, L. M. 2011, ApJ, 738, 71

Showman, A. P., Tan, X., & Parmentier, V. 2020, SSRv, 216, 139

Spiegel, D. S., Silverio, K., & Burrows, A. 2009, ApJ, 699, 1487

TensorFlow Developers, 2023 TensorFlow, v2.12.1, Zenodo, doi:10.5281/zenodo.8118033

Tremblin, P., Amundsen, D. S., Mourier, P., et al. 2015, ApJL, 804, L17

Tremblin, P., Chabrier, G., Mayne, N. J., et al. 2017, ApJ, 841, 30

Tsai, S.-M., Dobbs-Dixon, I., & Gu, P.-G. 2014, ApJ, 793, 141

Vallis, G. K. 2006, Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation (2nd ed.; Cambridge: Cambridge Univ. Press), 946

Williamson, D. L. 2007, JMeSJ, 85B, 241

Zahnle, K., Marley, M. S., Freedman, R. S., Lodders, K., & Fortney, J. J. 2009, ApJL, 701, L20