THESIS

USING NEURAL NETWORKS TO LEARN THE FORCED RESPONSE OF THE JET-STREAM TO TROPOSPHERIC TEMPERATURE TENDENCIES

Submitted by

Charlotte Connolly

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2023

Master's Committee:

Advisor: Elizabeth Barnes

David Randall
Chuck Anderson

ABSTRACT

USING NEURAL NETWORKS TO LEARN THE FORCED RESPONSE OF THE JET-STREAM TO TROPOSPHERIC TEMPERATURE TENDENCIES

Two distinct features of anthropogenic climate change, warming in the tropical upper troposphere and warming at the Arctic surface, have competing effects on the midlatitude jet-stream's latitudinal position, often referred to as a "tug-of-war". Many previous studies have investigated the strength of the jet response to these thermal forcings, as well as many others, and have shown that the jet response is sensitive to model type, season, initial atmospheric conditions, and the shape and magnitude of the forcing. Here, we explore the potential for training a convolutional neural network (CNN) on internal variability alone to examine possible nonlinear jet responses to a variety of thermal forcings. Our approach thus makes use of the fluctuation-dissipation theorem, which relates the internal variability of a system to its forced response. We train a CNN on data from a long control run of the CESM dry dynamical core, thereby providing it with ample data to learn relationships between the temperature forcing and the jet movement over the coming days. Then, we use the CNN to explore the jet response to a wide range of tropospheric temperature tendencies. Despite being trained on the jet-stream response to internal variability alone, we show that the trained CNN is able to skillfully predict the nonlinear response of the jet-stream to sustained external forcing. The trained CNN provides a quick method for exploring the jet-stream sensitivity to a wide range of tropospheric temperature tendencies and, considering that this method can likely be applied to any model with a long control run, could lend itself useful for early stage experiment design.

TABLE OF CONTENTS

Chapter 1

Introduction

Questions about how the Earth system responds to external forcing have been fundamental in atmospheric science for decades. From early studies that investigate the sea surface temperature response to atmospheric wind (Elsberry and Raney, 1978) and the forcing responsible for the onset of El Niño events (Wyrtki, 1975), to more recent studies that investigate soil moisture's ability to regulate the atmosphere's response (Erdenebat and Sato, 2016) and the response of El Niño events (Yeh et al., 2018) to climate change. Learning about the Earth system by understanding its response to a forcing is a consistent framework used by atmospheric scientists. This thesis outlines a novel method that applies machine learning to examine the forced response of the climate system.

The fluctuation-dissipation theorem (FDT) relates the mean response of a nonlinear system to a small forcing (e.g. Kraichnan, 1959; Leith, 1975; Marconi et al., 2008). Using this theorem, we train a neural network on data that contains only internal variability and then use that network to predict the system's forced response. Training a neural network on only internal variability allows the method introduced in this thesis to transfer to other climate models that have a long control run, and is therefore proposed as a tool for early stage climate model experiment design. This method does not replace the need to run the forced climate model experiments, but instead creates the opportunity to explore a large number of forcing scenarios to help decide which experiments should be simulated in more expensive climate models.

In this work, we focus on investigating the response of the northern hemisphere eddy-driven jet-stream to enhanced warming in two distinct regions known to drive competing responses. These

regions of enhanced warming are located in the tropical upper troposphere and the Arctic surface (e.g. Held, 1993). When the regions of enhanced warming occur simultaneously, the competing forced responses on the jet is commonly referred to as the "tug-of-war" (e.g. Harvey et al., 2015; Stendel et al., 2021). The neural network we train relates a thermal forcing to a jet response by taking a temperature tendency as an input and predicting a jet shift. This setup allows us to investigate the jet response to a variety of thermal forcing including the two that comprise the tug-of-war. Here, we train a CNN on temperature tendencies from a Community Earth System Model (CESM) dry dynamical core control run and then use the trained CNN to explore the jet response to a temperature tendency forcing by predicting the change in the jet latitudinal location (i.e. jet shift).

Chapter 3 outlines the training data, CNN architecture, and two approaches used to evaluate the CNN's skill. The results from this study are found in Chapter 4 and include a comparison between CNN predicted jet shifts to true jet shifts from forced dry core runs. Chapter 5 provides and example of how the CNN we train in this study can be used to investigate jet sensitivities. Concluding remarks are found in Chapter 6 and possible future work in Chapter 7. Chapters 2 through 6 have been prepared for publication and plan to be submitted after additional co-author feedback. A tentative citation for this is provided below:

Connolly, C., Barnes, E. , Hassanzadah, P., Prichard M., 2023: Using Neural Networks to Learn the Forced Response of the Jet-Stream to Tropospheric Temperature Tendencies. *Artif. Intel. Earth System.*

Chapter 2

Training a Convolution Neural Network to learn a Forced Response

The Northern Hemisphere eddy-driven jet-stream drives much of the Northern Hemisphere midlatitude weather (e.g. Nakamura et al., 2004; Athanasiadis et al., 2010; Shaw et al., 2016; Madonna et al., 2017). Consequently, changes in the jet-stream position and strength can result in enormous societal impact by impacting heat waves, droughts, and flooding events (Schubert et al., 2011; Coumou and Rahmstorf, 2012; Bibi et al., 2020; Rousi et al., 2021, 2022), extreme weather across the midlatitudes (Mahlstein et al., 2012; Röthlisberger et al., 2016), hurricane tracks (Mattingly et al., 2015), and crop production (Kornhuber et al., 2019). Two robust features of anthropogenic climate change that impact the midlatitude jet, warming in the upper troposphere of the tropics and warming at the surface of the Arctic, have been shown to independently force opposite responses in the mean jet location (e.g. Held, 1993; Harvey et al., 2015; Stendel et al., 2021), and its response is largely attributed to changes in the pole to equator temperature gradient driven by the regional heating (Blackport and Screen, 2020; Stendel et al., 2021). Warming in the tropical upper troposphere drives a poleward shift in the mean jet location by increasing the upper tropospheric pole to equator temperature gradient, while, warming at the Arctic surface drives an equatorward shift in the mean jet location by decreasing the surface pole to equator temperature gradient (Butler et al., 2010; Screen et al., 2013; Chen et al., 2020; Stendel et al., 2021). The competing jet response stemming from these two thermal forcings is commonly referred to as the

"tug-of-war" on the jet-stream (Held, 1993). Current consensus across climate models is that the upper tropospheric warming wins-out over the Arctic surface warming, causing a net poleward shift of the jet (Yin, 2005; Swart and Fyfe, 2012; Barnes and Polvani, 2013; Harvey et al., 2015). However, there is still substantial disagreement over the magnitude of the jet response due to model disagreement on the strength and spatial extent of the regional heating anomalies (Grise and Polvani, 2016).

Warming in both the tropical upper troposphere and Arctic surface are consequences of anthropogenic greenhouse gas emissions, but are caused by distinct dynamical processes that determine the characteristics of the thermal anomalies. The tropical upper atmosphere warms more as a result of additional water vapor stored in the warmer tropical tropospheric air (i.e. a reduction in the moist adiabatic lapse rate; (Sherwood and Nidhi Nishant, 2015)). The enhanced Arctic warming, commonly referred to as Arctic Amplification, is occurring three times faster than elsewhere on the planet (Blunden and Arndt, 2012; Druckenmiller et al., 2021) and is driven by multiple processes that include changes in poleward energy transport (Hwang and Frierson, 2010; Graversen and Langen, 2019), surface ice-albedo feedbacks (Manabe and Stouffer, 1980; Dai et al., 2019), cloud feedbacks (Abbot and Tziperman, 2008) and lapse-rate feedbacks (Pithan and Mauritsen, 2014). To further increase the complexity of the processes driving the regional warming, the two thermal forcings likely do not act entirely independently. Research has shown that increased transient Rossby waves initiated in the tropics may drive increased heat transport into the Arctic and as a result drive further warming in the mid- and upper-troposphere of the Arctic (Lee, 2014; Dunn-Sigouin et al., 2021). Uncertainties in the processes that contribute to the magnitude and shape of warming in the tropical upper troposphere and Arctic surface (Blackport and Screen, 2020; Stendel et al., 2021), in turn, make it even more challenging to predict the magnitude of the jet response.

Despite the large body of work that investigates the response of the midlatitude jet under climate change, the topic continues to receive attention because the jet's response to climate change is difficult to isolate from internal variability, the observational record is short and sparse at the high latitudes, and because of modeling deficiencies with ice and clouds feedbacks (Tjernström et al., 2008; Kattsov et al., 2010; Cohen et al., 2014; Pithan et al., 2014; Vihma, 2014). The studies that have investigated the response of the jet to a thermal forcing have shown that the jet is sensitive to the shape, location, and magnitude of the thermal forcing (Butler et al., 2010), the season in which the forcing is imposed (McGraw and Barnes, 2016), the current state of the atmosphere (i.e. position of the jet-stream; (Gerber et al., 2008; Barnes et al., 2010; Kidston and Gerber, 2010; Garfinkel et al., 2013)), and the climate model used for the study (Meehl et al., 2007; Barnes and Polvani, 2013).

In an attempt to explore circulation sensitivities to a wide range of possible thermal forcings, Hassanzadeh (2016b) used a control run from the GFDL dry dynamical core (Manabe et al., 1974) and employed the fluctuation-dissipation theorem (FDT) to compute the linear response function of the circulation to gaussian-like thermal forcings. FDT relates the mean response of a nonlinear system to a small forcing by means of a linear operator that is created from the internal variability of the system (e.g. Kraichnan, 1959; Leith, 1975; Marconi et al., 2008). With the ability to explore a forced response from internal variability, FDT has been proposed as a method to quickly estimate circulation sensitivities in climate models (Fuchs et al., 2015) and serve as a useful tool for planning expensive climate model experiements (Leith, 1975). There have been encouraging results using FDT to explore the circulation response to thermal forcings in general circulation models (Gritsun and Branstator, 2007) as well as more complex coupled climate model (Mk3L; (Phipps, 2010)) to estimate the response to realistic sea surface temperature forcings (Fuchs et al., 2015).

In order for the linear operator of FDT to accurately represent the mean response to a forcing, the system must satisfy a number of conditions (Marconi et al., 2008). The first condition is that the system must be near equilibrium, because FDT assumes that small changes in the system's state (internal variability) has a recovery back to equilibrium that is similar to the system's response to a small perturbation (Kraichnan, 1959; Leith, 1975). The second is that the perturbation must be small enough so that the response is linear even though the system that the operator is created from need not necessarily be linear (Leith, 1975). Lastly, the system must be gaussian, though work has been done to make versions of FDT where the system can be quasi-gaussian (Majda et al., 2005). A system that satisfies these conditions can use FDT to compute the systems' response to a small forcing.

Instead of using FDT to relate a forcing to a response, this study uses a convolutional neural network (CNN) to learn the nonlinear relationship between a forcing and a response. Moreover, using a CNN in place of the linear operator removes the need to make some of the FDT assumptions (i.e. small forcing for a linear response). The trained CNN reads in a temperature tendency field and a single value that represents the initial jet location and predicts a jet response in the form of a jet shift (in units of degrees latitude). Training is performed on data from a long control run with the CESM dry dynamical core and once complete, the CNN is used to explore the jet sensitivity to a variety of temperature forcings. Throughout this study, we evaluate the CNN's ability to quantify the CESM dry dynamical core's jet sensitivity placing particular emphasis on the tug-of-war between the warming in the tropical upper troposphere and the Arctic surface. Training a neural network on internal variability alone and then using it to predict a forced response is, to our knowledge, a novel application of a neural network to climate analysis. Therefore, we assesses the strengths and weaknesses of this approach in multiple ways (see Chapter 4).

Chapter 3

Methods

We train a CNN to predict the jet-stream response to zonally averaged regional temperature perturbations. The goal is to investigate jet sensitivity to thermal features associated with anthropogenic climate change. The CNN, detailed in Chapter 3.2, is trained on a long control run from a dry dynamical core, which is documented to reproduce the majority of the northern hemisphere's jet response to heating perturbations along with simulating the correct sign of the jet shift (e.g. Mbengue and Schneider, 2013; McGraw and Barnes, 2016; Baker et al., 2017). Once trained, the CNN's skill is examined by comparing its predictions to additional baseline prediction methods and dry core experiments that include an imposed temperature forcing. Details on the dry dynamical core setup, the CNN architecture and training, additional baseline prediction methods, and additional dry core heating experiments are discussed in more detail in the following Chapters.

3.1 Training Data

We use output from the Community Earth System Model (CESM) Eulerian spectral-transform dry dynamical core (Evans et al., 2014). The model runs are completed with the Held Suarez configuration (Held and Suarez, 1994), such that friction exists at the surface and the temperature is relaxed to a prescribed hemispherically symmetric temperature field. The relaxation temperature field is set to equinoctial conditions and there is no absorption of solar energy by the atmosphere (i.e. there are no seasons or diurnal cycles). All runs are performed at T42 resolution with 30 vertical levels, 64 latitude bands, and 128 longitude bands. The simulation is run in the above

configuration for one million six hour time steps. The first 20,000 time steps (13.7 years) are thrown out to account for model spin up.

All data processing is performed to create efficient training data for a convolutional neural network (CNN; Chapter 3.2) to predict the zonally-averaged jet response to a range of tropospheric temperature perturbations. Two variables are used in this study, zonally averaged temperature and zonally averaged zonal wind speed. The zonally-averaged temperature data is used to calculate the temperature tendency field used as input to the CNN. The zonally-averaged zonal wind speed is used to calculate the location and shift of the jet, which is used as a CNN input and the CNN prediction respectively. We exclude data from 200 hPa and above, effectively removing the stratosphere which is not well resolved in this model without modification (Polvani, 2002), so focus can remain solely on the troposphere for both the forcing and the jet response. Given that this study focuses on hemispheric jets, we take advantage of the hemispheric symmetry in the dry core and use each hemisphere as a separate independent sample, doubling the amount of data from one million samples to two million. After zonally averaging, removing the stratosphere, and considering each hemisphere as a separate sample, the resulting size of the temperature field is 25 vertical levels by 32 latitude bands.

Backward differencing is used to calculate the temperature tendency by subtracting the temperature field of the current time step from the temperature field one time step into the past. This temperature tendency is then smoothed to remove higher frequency variability. Smoothing is completed by calculating a backward running mean of 240 time steps (60 days), returning a low frequency temperature tendency. Removing the high frequency variability allows the network to focus on learning the response to a forcing that more closely mimics continuous climate change forcing.

Smoothing the data before calculating the temperature tendency did not result in any changes in the CNN skill.

Following established methods (Woollings et al., 2010; McGraw and Barnes, 2016), the jet location is defined as the latitude of the maximum wind speed at a pressure level near the surface. Zonal wind speeds from the 850 hPa level are used here and are first smoothed with a 240 time step (60 days) backward running mean. Then, a second order polynomial is fit to the peak of the smoothed 850 hPa zonal wind profile and the jet location is defined as the latitude of the maximum wind speeds.

In this work, the jet response to a given temperature tendency is defined by how far the jet latitude changes from the time of input to 120 time steps later (i.e. the jet shift). A positive jet shift indicates a poleward shift in jet location and a negative jet shift indicates an equatorward shift in jet location relative to the jet's latitude at the time of prediction. Before the jet shifts are calculated, the processed temperature tendency and jet latitude are split into training, validation, and testing data to ensure no data leakage across the datasets. Splitting is completed by chunking the data. The training data is the first chunk, validation the second, and testing the last. Then, the jet shift is calculated within each dataset separately by subtracting a 240 time step backward running mean of jet stream locations from a 240 time step forward running mean of jet stream locations 120 time steps into the future. This processing results in 359,280 training samples, 199,280 validation samples, and 1,399,558 testing samples. Training the CNN required fewer samples than expected as we found that adding more samples to the training dataset did not improve the CNN skill, explaining why the testing dataset is much larger than the training and validation datasets.

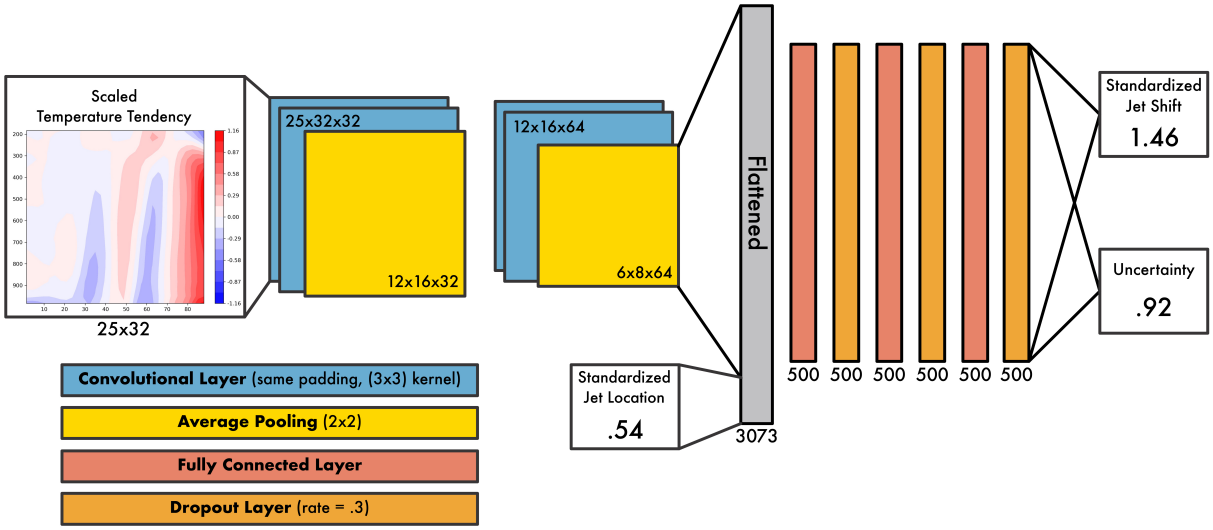

Figure 3.1: Schematic of the convolutional neural network with an example of an input and output.

3.2 Convolutional Neural Network

This study trains a convolutional neural network (CNN) to learn the jet's response to tropospheric temperature tendencies. CNN's are commonly used for image recognition and classification tasks as the convolutional layers can extract spatial features in the input image that help the network learn the correlations between the inputs and output (Fukushima, 1980; Yann et al., 1998; Zeiler and Fergus, 2014). Initially, we attempted to train a fully connected feedforward network (e.g. LeCun et al., 2015) but could not train a model with noteworthy skill. In concept, a fully connected feedforward network should be able to learn the same features extracted by the CNN but may require a larger network and more training data (Yann et al., 1998; Ingrosso and Goldt, 2022). For our study, the convolutional layers were important for creating a skillful network while also trying to minimize the amount of data required to train the CNN.

The CNN architecture is trained with two different inputs: a temperature tendency field (K day^{-1}) and an initial jet location (degrees latitude). The temperature tendency input is crucial

since we are motivated by investigating the relationship between regional temperature tendencies and the future jet location. The initial jet location supplies the CNN with essential information about the current state of the jet at the prediction time. The current state of the atmosphere is an important factor in the jet response to forcing (Gerber et al., 2008; Barnes et al., 2010; Kidston and Gerber, 2010; Garfinkel et al., 2013) and providing this information to the CNN increased the network's performance (not shown). Both the temperature tendency and jet location inputs are rescaled in order to be of similar magnitudes (order of 1) by multiplying the temperature tendency by a factor of 10 and standardizing the initial jet location using the mean and standard deviation of the jet location from the training data.

The network is comprised of four convolutional layers: two average pooling layers, three dense layers, and three dropout layers (Fig. 3.1). Convolutional and dense layers use the hyperbolic tangent activation function. Data are passed through the network as follows: the scaled temperature tendency goes directly into the first convolutional layer with 32 filters of size 3 x 3 and a stride of 1 followed by a second convolutional layer with the same attributes. The second convolutional layer is then connected to an average pooling layer with kernel size of 2 x 2. These three layers, two convolutional and a single average pooling, are repeated with the same attributes with the exception of 64 filters rather than 32 in the convolutional layers. The output from the second average pooling layer is flattened and the standardized initial jet location is concatenated to the end. This layer is then fed into the first dense layer with 500 nodes and then goes through a dropout layer with a dropout rate of 30%. The data passes through a combination of dense layers with 500 nodes followed by dropout layers with a dropout rate of 30% two more times. The data from the final dropout layer then passes into the output layer comprised of two nodes.

The CNN outputs two values, denoted μ and σ, which represent a mean and standard deviation of a gaussian distribution where μ denotes the predicted jet shift and σ represents its uncertainty. Predicting the parameters of a gaussian distribution is commonly used to quantify uncertainty for neural networks (Nix and Weigend, 1994, 1995), and Gordon and Barnes (2022) recently showed the utility of incorporating uncertainty into a regression neural network for climate science applications. The network learns to predict μ and σ through the implementation of the negative log-likelihood loss function:

$$L_i = -log(p_i) \tag{3.1}$$

where p is a value of the predicted gaussian distribution evaluated at the true jet shift for the i^{th} input sample. To ensure the network is calibrated we employ the probability integral transform (PIT) probability calibration scheme (Gneiting et al., 2007; Nipen and Stull, 2011; Barnes et al., 2022). The PIT histogram for this CNN can be found in Appendix B.

To train the CNN, we use the Adam stochastic gradient descent optimization algorithm (Kingma and Ba, 2014) with a learning rate of 10^{-7}, a batch size of 256, and a random seed of 300. We apply early stopping to halt the training process once the validation loss fails to decrease for 10 consecutive epochs. Once training is stopped, the model weights are restored to the version with the lowest validation loss. Early stopping reduces the risk of overfitting by monitoring the validation loss and reducing the chance the network overfits to the training data (Prechelt, 2012).

3.3 Baselines

We establish two baselines in this study to assess the performance of the CNN and demonstrate that the CNN has learned relationships between the jet response and the regional temperature tendencies. The first baseline is called *persistence*, similar to "persistence forecasting" (MacDonald, 1992), where future conditions are predicted to be identical to the current conditions. In our case, this translates to the jet's future location being the same as its location at the time of prediction (i.e. jet shift equal to zero). Comparing this baseline to the CNN ensures that the CNN is predicting jet shifts that are more accurate than predicting a jet shift of zero. The second baseline is called *average evolution* and describes the average movement of the jet based on its position at the time of prediction. For this baseline calculation, the training data is separated into 100 different bins according to the initial jet location, essentially grouping samples with similar initial jet locations together. The average jet shift for each bin is calculated, resulting in an average jet response that is solely dependent on the jet-stream's starting position. The average evolution baseline is not sensitive to the number of bins or their exact spacing (not shown). This baseline ensures that the CNN is not just predicting the average evolution of the jet based solely on the initial jet location but is also using the temperature tendency to make its prediction.

Every sample is thus associated with three jet shift predictions, one from the CNN and two from the additional persistence and average evolution baselines. Although comparing results between the CNN and the two baselines is useful for placing the CNN's predictions into context, we highlight that the baselines and the CNN make predictions based on different information. The baselines make predictions based solely on information about the initial location of the jet while

the CNN is provided additional information in the form of the temperature tendency. Thus, the CNN is able to explore the correlations between a temperature tendency and a jet response.

3.4 Heating Experiments

The main goal of this study is to investigate the jet-stream sensitivity to temperature forcing driven by anthropogenic climate warming. However, as we have designed it, the CNN only trains on data from a long control run (i.e. internal variability), and thus, only provides insights into the forced response if fluctuation-dissipation theorem holds (Kraichnan, 1959; Leith, 1975; Marconi et al., 2008). In other words, if the CNN can learn a forced response solely from internal variability. To investigate whether this assumption is valid, we run additional dry core simulations (referred to as *heating experiments*) with zonally symmetric imposed temperature tendencies (F) that take the form of a 2-D gaussian in the latitude/pressure plane (Equation 3.2):

$$F(\Theta, p) = q_o exp\left[\frac{(\mid \Theta \mid -\Theta_o)^2}{\Theta_w^2} - \frac{(\mid p \mid -p_o)^2}{p_w^2}\right] \tag{3.2}$$

where Θ_o and p_o are the horizontal and vertical centers respectively, Θ_w and p_w define the width and height, and the magnitude of the forcing is given by q_o. Gaussians that fall near the edge are cut off and therefore not complete 2-D gaussians. For all gaussians created in this study, $\Theta, \Theta_w, p, p_w, q_o$ are reported in Appendix A.

Eighteen heating experiments with either one or two temperature tendencies imposed are run out to equilibrium in order to quantify the true jet shift driven by different imposed temperature

forcings (see experiments #1-18 in Table A1). To perform a direct comparison between the CNN predicted jet responses and the dry core jet responses, the CNN is given the exact same temperature tendency forcing that is imposed in each of the dry core heating experiments. For the CNN's initial jet location input, the average jet location from the long control run is used (42.4°), which once standardized, becomes an input of zero. By comparing the true forced response from the dry core to the predicted forced response by the CNN we are able to investigate the CNN's ability to learn and predict the jet response to temperature forcing from training on internal variability alone.

Each heating experiment is initiated at the end of the long control run (time step one million; 684.9 years), and therefore, have the same initial conditions. The heating experiments are run for an additional 20,000 time steps with the first 4,000 removed to ensure that the model has reached its new equilibrium. The 850 hPa zonal winds are then used to compute the location of the jet (see Chapter 3.1). Finally, the true response of the jet-stream to an imposed temperature tendency is defined as the average jet location during the long control run subtracted from the jet location in the corresponding heating experiment.

Chapter 4

Results

4.1 Evaluation of CNN skill

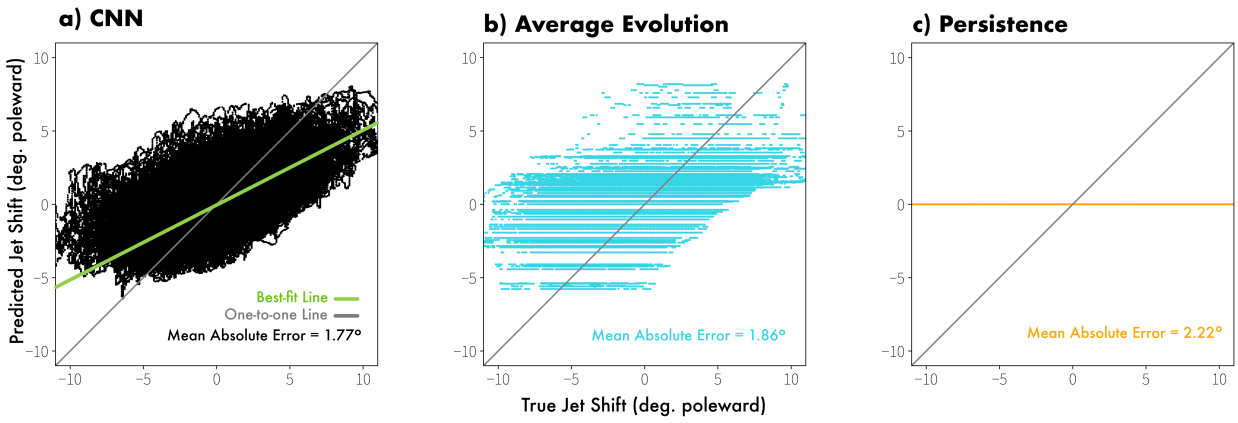

Figure 4.1: Scatter plots of the predicted jet response (y axis) versus the true jet response (x axis) for the (a) CNN, (b) average evolution baseline, and (c) persistence evolution baseline. The mean absolute error across all testing samples is shown in the bottom right corner of each panel. The grey line denotes the one-to-one line and represents a perfect prediction. The green line in panel (a) represents the best-fit line of the testing data.

We begin our discussion of the results with a focus on the deterministic predictions by the CNN (μ). The deterministic skill of the testing data reveals how well the CNN generalizes to unseen samples within the control simulation, and here, we quantify skill as the mean absolute error between the predicted jet shift and the true jet shift. Figure 4.1a shows the relationship between the predicted jet shift and the true jet shift where predictions with higher accuracy are closer to the grey diagonal line (one-to-one line). Using orthogonal distance regression (Boggs and Rogers,

1990; Virtanen et al., 2020), which takes into account error in both the x and y variables as well as the CNN-predicted uncertainties in y, we calculate the slope from the testing data to be 0.5 deg. poleward / (deg. poleward). This positive slope demonstrates the CNN has learned relationships between the jet shift and the inputs. However, the slope of the CNN predictions is less than that of the one-to-one line which implies that the CNN underestimates the magnitude of the largest jet shifts. This is likely a result of the imbalanced training data as it includes more samples with smaller jet shifts than larger ones (not shown). During training, the goal of the CNN is to minimize the negative log-likelihood loss function (see Equation 3.1), but with an unbalanced dataset the CNN may never predict the most extreme cases. A few solutions to this issue include: balancing the data set, using sample weights, or creating custom loss functions (Ma et al., 2013; Krawczyk, 2016). Applying these methods to our CNN either caused a severe decrease in skill or did not succeed in making the network predict the extremes (not shown). Nonetheless, as we will show next, the CNN outperforms the two benchmark baselines and is an effective tool for exploring the sensitivity of the jet to external forcing.

To demonstrate the CNN's ability to outperform the benchmark baselines, we compare the CNN's skill on the testing data to that of the two baselines: average evolution (Fig. 4.1b) and persistence (Fig. 4.1c). Recall that these benchmarks are created to place the CNN's prediction skill into context against other basic prediction methods. Persistence has the lowest performance of the three models with a mean absolute error of 2.22°. Average evolution has a mean absolute error of 1.86° and performs slightly worse than the CNN which has a mean absolute error of 1.77°. Unlike persistence, which can only ever predict a jet shift of zero, average evolution makes a prediction based off of the average relationship between the initial jet location and the jet shift of the training data, allowing it to capture the mean jet response. Regardless, average evolution

is limited to predict one of the 100 jet shifts based on the initial jet location, hence the stripes in Figure 4.1b. The discrete predictions are a product of the binning used to calculate the average evolution baseline (see Methods). As noted in Chapter 3.3, the error of the average evolution baseline is not sensitive to increasing the number of bins. Based on the mean absolute error alone, the CNN outperforms both the persistence and average evolution baselines for the testing data.

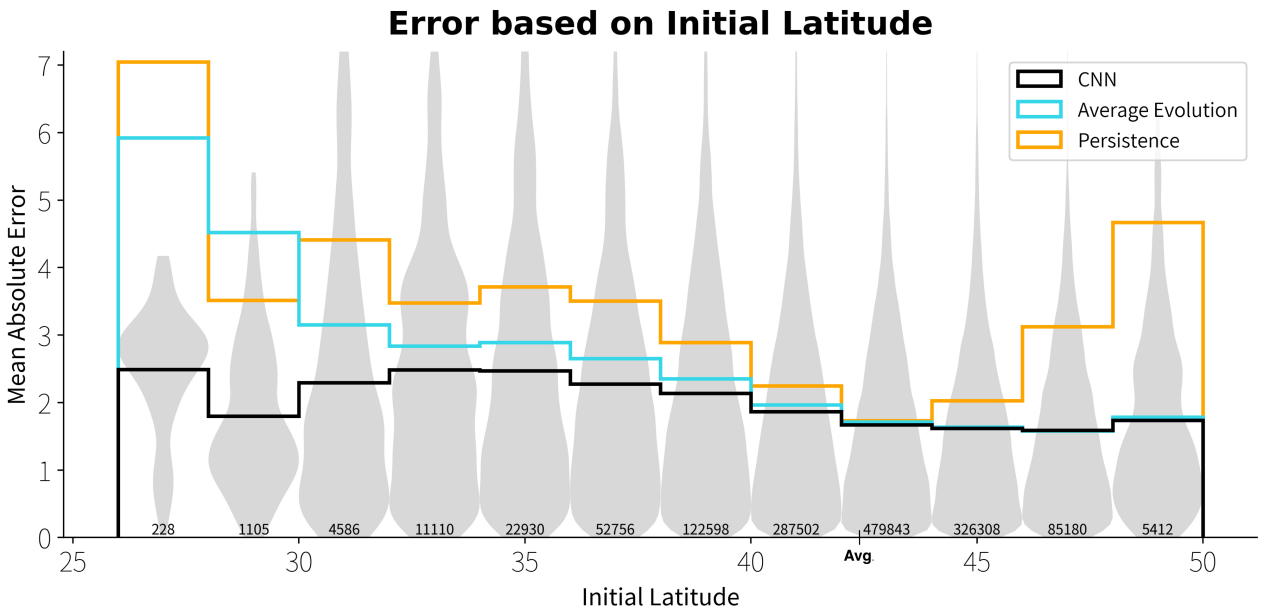

Figure 4.2: The mean absolute error from the three prediction methods, CNN (black line), average evolution (cyan line), and persistence (orange line) grouped by initial jet locations. Grey violin plots show the density curves of the CNN's error distribution (i.e. the data used to calculate the mean absolute error of the CNN in each group) where the width corresponds with the frequency of the data. Numbers at the bottom of each bar indicate the number of samples in each group and the average initial jet location from the training data is marked on the x axis (42.4°).

The mean absolute errors in Figure 4.1 represent the error over the entire testing set. For a more exhaustive description of the CNN's skill, the testing data is separated into groups based on the initial jet location. The mean absolute error for each initial jet location group is shown for the CNN and the baselines in Figure 4.2 and describes how the CNN's skill and the baselines' skill depend on the initial state of the jet. The grey violin plots behind each bar indicate the distribution

of the CNN's mean absolute error within that bin (smoothed with a kernel estimator using Scott's rule and 100 points, which are the default parameters of the Matplotlib library (Hunter, 2007)). The bins do not contain an equal amount of samples, therefore the numbers at the bottom of each bar denote the number of samples used to calculate the average error and the grey violin plots for that bin. For all initial jet location bins in Figure 4.2, the CNN outperforms the baselines as demonstrated by the CNN's error (black line) falling below the baseline errors (cyan and orange lines). When the initial jet location is equatorward of 42° (labeled as "Avg." along the x axis of Fig. 4.2), the CNN does considerably better than the baselines, but when the jet location is poleward of 42°, the CNN and average evolution achieve similar skill. That is, in the cases where the initial jet is near the pole, it appears that the CNN does not learn more than average evolution but instead learns this average behavior to make its prediction.

When the initial jet location is near this climatological average position (42.4°), the errors of the CNN and baselines converge (Fig. 4.2). About 30% of the samples in the training data have initial jet locations within 2° of the climatological average and 14% of these samples have a jet shift between -0.5° and 0.5°. Since so many samples near the climatological average have small jet shifts, and because the persistence baseline can only predict a jet shift of zero, the mean absolute error for persistence is at its lowest near the jet's climatological average position. The average evolution baseline converges to a near zero prediction when the initial jet location is near the climatological average, resulting in persistence and average evolution exhibiting similar errors. Although the persistence and average evolution baselines have an advantage near the climatological average, the CNN still outperforms both baselines, implying that the CNN is using the additional information provided by the input temperature tendencies.

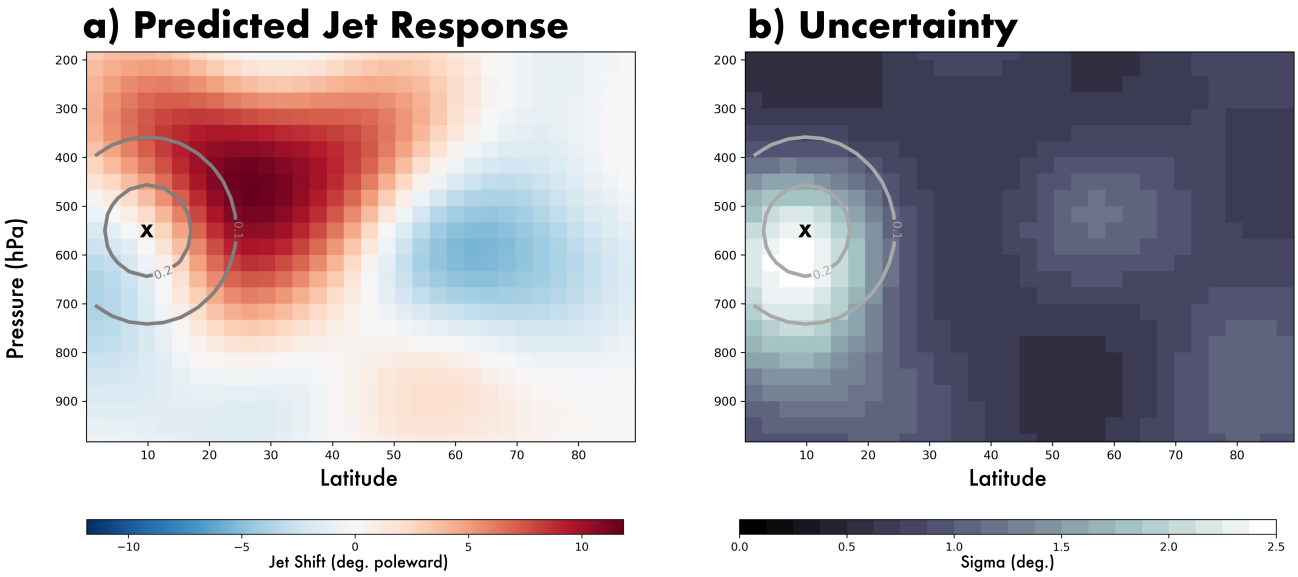

Figure 4.3: The CNN prediction from a gaussian temperature forcing with a magnitude of q_o = 0.25 K day $^{-1}$ that is moved around the latitude and pressure plane where the shading in (a) and (b) represent the jet shift and uncertainty respectively. An example of a gaussian is seen in the grey contours in both panels. The "x" shows the center of the gaussian as well as the predicted jet shift and predicted uncertainty associated with that gaussian. Gaussian parameters are found in Table A1 experiment #19.

Next we focus on evaluating the CNN's ability to predict a jet response from a temperature tendency that includes a forcing by including a 2-D gaussian temperature forcing (see Methods) into the temperature tendency input. These temperature tendency inputs contain a 2-D gaussian with a prescribed magnitude, size, and location (latitude and pressure). Outside of the gaussian, the temperature tendency field is filled with zeros. Although some of these temperature forcings have magnitudes larger than any temperature tendencies found in the internal variability training data, we will provide strong evidence to support the CNN's ability to extrapolate in the coming sections. CNN predictions made from a temperature forcing use an initial jet location defined by the average jet latitude over the entire long control run (42.4°). Therefore, differences in the predicted jet response between the different temperature forcings is a response to the temperature tendency alone and not the presumed initial state.

Figure 4.3 explores the jet sensitivity to the location of heating by holding the magnitude and the shape of 2-D gaussian constant and changing only the gaussian's location (Fig. 4.3; see experiment #19 in Table A1). An example gaussian is shown in the grey contours where the "x" denotes its center and the color shading beneath represents the predicted jet shift (Fig. 4.3a) and the predicted uncertainty (Fig. 4.3b) from the temperature tendency. The gaussian tendency is moved across the latitude-pressure plane and the shading in Figure 4.3 represents the jet sensitivity to the location of the heating as learned by the CNN.

Figure 4.3 exhibits multiple known features of the jet response to tropospheric thermal forcings. For example, thermal forcings located higher in the troposphere are known to be more effective at perturbing the jet than thermal forcings located lower in the troposphere (Kim et al., 2021). This feature is learned by the CNN and is shown in Figure 4.3a as darker shading at higher pressure levels. In addition, warming in the tropical upper troposphere has been previously shown to cause the jet to shift poleward (Chen et al., 2008; Lim and Simmonds, 2009; Butler et al., 2010), and this poleward jet shift is also seen in Figure 4.3a (denoted by the red shading in the tropical upper troposphere). Finally, heating at the polar surface has been shown to cause the jet to shift equatorward (Butler et al., 2010; Deser et al., 2010; Screen et al., 2013) and this is also seen in Figure 4.3a by the blue shading, albeit weak, near the polar surface. The equatorward jet shift driven by polar surface warming and the poleward jet shift driven by tropical upper tropospheric warming indicate that the CNN has learned the correct sign of the jet shift as supported by prior research.

Figure 4.3 also highlights how moving the center of the heating by a few degrees or pressure levels can change the direction of the jet shift. Take for instance heating at the polar surface, where moving the heating from 80° latitude to 75° latitude changes the jet response from an equatorward shift to a poleward shift. Baker et al. 2017 investigates the jet senstivity by running 306 dry

core experiments with a gaussian shaped temperature tendency that is moved around the latitude-pressure plane, just as we have done here with a trained CNN. In Baker et al. (2017), they show that changes in the latitude of the heating most strongly impacts the sign of the jet shift while changing the pressure level has very little impact. Similar behavior is found here with the CNN, although a few exceptions are found near the surface of the midlatitudes. A more in-depth discussion about the failure of the CNN to capture the correct direction of the jet shift at the surface of the midlatitudes is found in Appendix C.

Recall that the CNN predicts both the jet shift (μ) as well as its uncertainty (σ). Figure 4.3b displays the predicted uncertainty values and highlights three regions where the CNN is less certain than others. The CNN is less certain when heating occurs around 10° latitude and 600 hPa ($\sigma \approx 4°$) and additionally has large uncertainty when the heating is around 1.5°and is centered near 85° latitude and 900 hPa and the 60° latitude and 550 hPa ($\sigma \approx 1.5°$). The reasons for the greater uncertainty in these regions require further investigation.

4.2 Nonlinearities learned by the CNN

Ideally, a benefit of using a CNN is that it learns a nonlinear relationship between the temperature tendency input and the jet shift output. The hyperbolic tangent activation functions in the convolutional and dense layers of the CNN allow it to learn nonlinear relationships between the inputs and outputs if nonlinearity is present in the data. However, this does not necessarily mean the CNN has learned nonlinear relationships. To evaluate the nonlinearity learned by the CNN we complete a few analyses. The first analysis examines how the CNN predicted jet shift varies as a function of the input forcing magnitude. The second analysis looks at scenarios where two gaussian forcings are simultaneously present in the input forcing and explores whether the CNN

has learned a nonlinear interaction between the two. Keep in mind that neither of these analyses has a ground truth, and so we are exclusively exploring what the CNN has learned. In the next section we will then further test the accuracy of the CNN with additional dry core simulations.

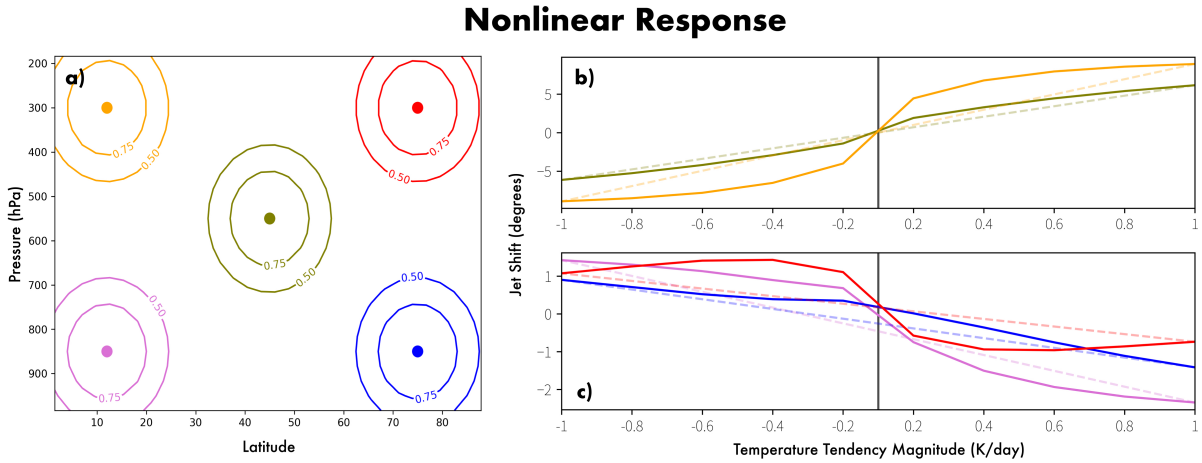

Figure 4.4: The CNN learned nonlinear responses where panel (a) shows example gaussians with magnitudes of 1 K day^{-1} at the five locations and panels (b) and (c) show the linear response (dashed line) and the predicted response from the CNN (solid lines). Note the different y axis limits in panels (b) and (c). Gaussian parameters are found in Table A1 #20-24.

The first nonlinear analysis explores how the jet shift varies as a function of the temperature forcing magnitude by inputing gaussians of different magnitudes in five different locations (Fig. 4.4a; see experiments #20-24 in Table A1). We use ten different magnitudes that vary from -1.0 K day^{-1} to 1.0 K day^{-1} in increments of 0.2 K day^{-1} for each location. For all cases, the input initial jet location is fixed at the average jet location during the dry core control run (42.4°). Figure 4.4b and 4.4c compare a linear relationship (dashed lines) and the CNN's learned relationship (solid lines) between the temperature forcing magnitude and the jet shift. Note the different y axes limits on panels (b) and (c). Temperature tendencies near the polar surface and the mid-tropospheric midlatitudes have the most linear response to varying temperature forcing magnitudes as shown

by the green line in Figure 4.4b and blue line in Figure 4.4c which have a predicted jet shift that is most similar to the linear dashed line. In contrast, the jet response to temperature tendencies in the tropics and the upper troposphere of the polar region have the largest nonlinear response (yellow line in Figure 4.4b, pink, and red lines in Figure 4.4c).

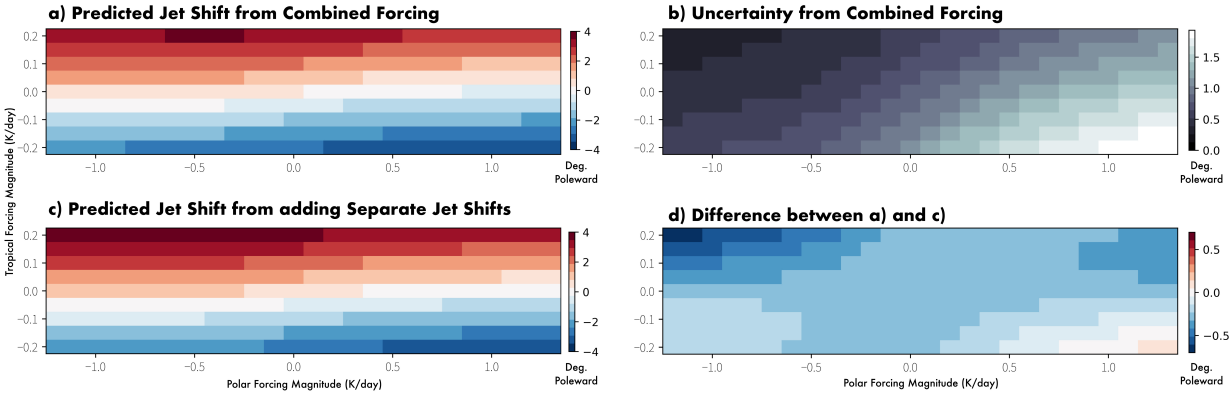

Figure 4.5: Panel (a) shows the CNN predicted jet shift from two temperature forcing with varying magnitudes. One at the surface of the pole (x axis) and the other in the upper troposphere of the tropics (y axis) and panel (b) shows the uncertainty for the predictions of panel (a), where darker shading is a more certain prediction. Panel (c) is similar to panel (a) but the CNN predicts the jet shift from the two forcings independently which are then added together. Panel (d) is the difference between panel (a) minus panel (c) which shows the CNN learned nonlinearity.

We next explore the nonlinearities learned by the CNN when two temperature forcings are present. The temperature forcings are centered on two key regions, the tropical upper troposphere and the polar surface. As discussed previously, warming in the tropical upper troposphere forces the jet to shift poleward (e.g. Chen et al., 2008; Lim and Simmonds, 2009; Butler et al., 2010) and warming at the polar surface forces the jet to shift equatorward (e.g. Butler et al., 2010; Deser et al., 2010; Screen et al., 2013). In reality these forcings do not act seperately, instead, they can occur simultaneously and force competing effects that can result in a tug-of-war scenario on the jet-stream (e.g. Harvey et al., 2015; Chen et al., 2020). To explore the jet sensitivities to this

climate change induced tug-of-war, the temperature tendency inputs take the form a 2-D gaussian at the polar surface and another in the upper atmosphere of the tropics. Both vary independently in magnitude during the analyses (see experiment #25 in Table A1) and Figure 4.5a shows the predicted jet shift (μ) and 4.5b shows the predicted uncertainty (σ) for each forcing pattern. With regards to the tug-of-war, studies use a variety of atmospheric models to show that despite warming in both regions that can force competing jet responses, the likely jet response is for the jet to shift poleward (Yin, 2005; Harvey et al., 2015). The upper right quadrant of Figure 4.5a depicts the situation where both temperature forcings are positive. In this scenario, the CNN predicts a poleward shift of the jet in agreement with past work, however, the CNN is not equally certain for all predictions. As shown in Figure 4.5b, the CNN is more confident with cooling at the pole and warming in tropics and less confident with warming in the pole combined with cooling in the tropics. Understanding why these scenarios are more uncertain requires further investigation.

Figure 4.5a shows that the CNN predicted jet response when two temperature forcings are present in the temperature tendency input. To test whether the CNN has learned a nonlinear impact on the jet from two forcings, we task the CNN to predict the jet shift from the two temperature tendency forcings independently (upper tropical troposphere and polar surface) and we add the two predicted together subsequently. If the CNN exclusively learned a linear response between two forcings, Figure 4.5a and Figure 4.5c would be identical, as predicting a jet shift from combined forcings would be equal to predicting the jet shifts separately and adding them together. Instead, Figure 4.5d shows their difference and provides evidence of the nonlinearity learned by the CNN where larger magnitudes indicate greater learned nonlinearity. Temperature forcing inputs that contain stronger thermal forcings (scenarios in the corners of 4.5d) have the greater learned nonlinearity.

4.3 Out-of-sample tests

Thus far we have compared the CNN predicted jet shifts to our established baselines, true jet shifts from the control run, and to past work. We next evaluate the ability of the CNN to inform upon the jet response to a sustained temperature forcing. The fluctuation-dissipation theorem (FDT) states that the response of a nonlinear system to external forcing can be related to the internal variability of the system. Under the assumption that FDT holds, our CNN trained on internal variability may also be able to predict a forced response. We next explore this by comparing the true forced jet shift calculated from additional dry core experiments to the predicted jet shift by the CNN.

We perform 14 additional forced heating experiments with the dry dynamical core (see Methods and Table A1 experiments #1-14). These 14 heating experiements are motivated by the tug-of-war on the jet as a result of anthropogenic climate change (Harvey et al., 2015; Chen et al., 2020; Stendel et al., 2021). To mimic the tug-of-war, each experiment contains a temperature forcing in the tropical upper troposphere and at the polar surface. The left side of Figure 4.6 includes the temperature tendency magnitude of each forcing. The forced jet response from the dry dynamical core experiments and the CNN predicted jet response from 14 heating experiments are shown on the right side of Figure 4.6. Comparing the forced jet shift simulated by the dry core (green dots) and the predicted jet shift by the CNN trained on internal variability (black lines), we see that across all experiments, the CNN accurately captures the sign of the jet shift. Experiments #1 through #7 exhibit a negative jet shift and #8 through #14 exhibit a positive jet shift. Furthermore, nearly all of the experiments (excluding #3, #4, and #14) have forced jet shifts that fall well within

the uncertainty bounds predicted by the CNN ($\pm 2\sigma$; grey boxes). Additional discussion of other heating experiments that do not focus on the tug-of-war on the jet can be found in Appendix C.

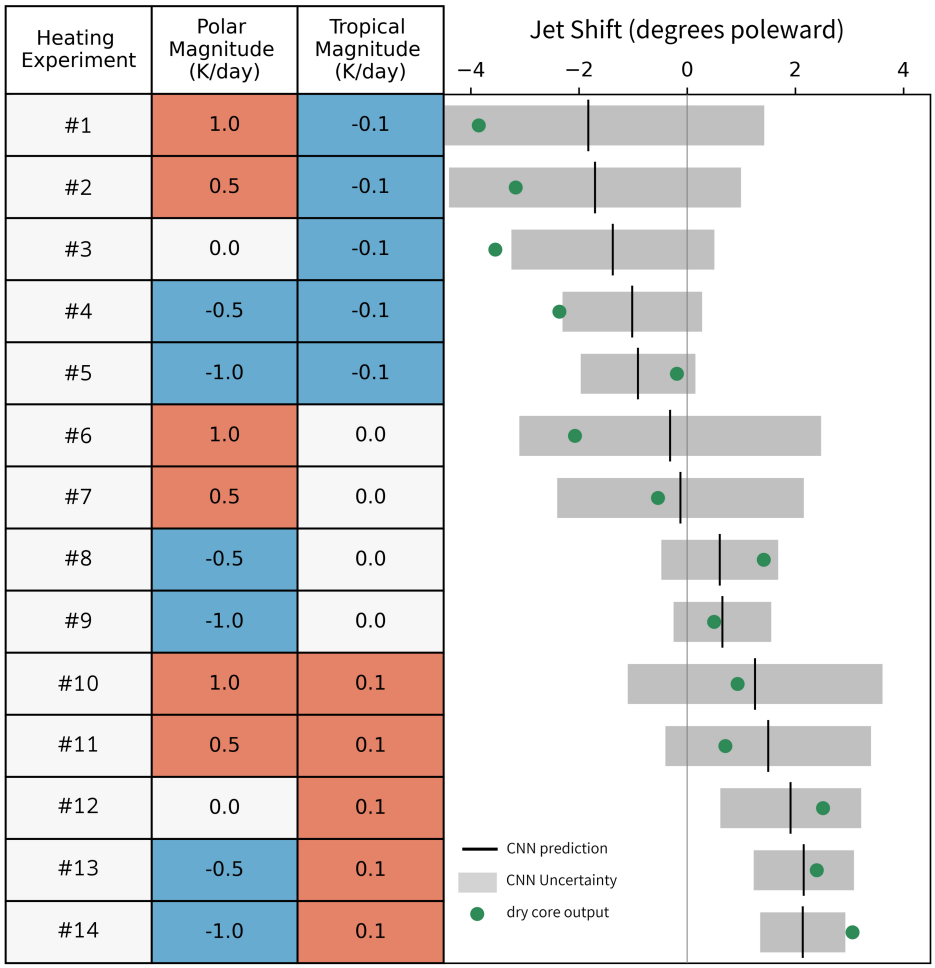

Figure 4.6: The 14 heating experiments, their resulting jet shift in the dry core, and the jet shift as predicted by the CNN. The left side shows the magnitude of the polar and tropical temperature tendencies for each heating experiment. The right side, shows the true forced dry core jet shift (green dots), the predicted CNN jet shift (black line), and the CNN predicted uncertainty (grey boxes; $\pm 2\sigma$).

Heating experiments #6, #7, #8, and #9 contain only a temperature forcing at the polar surface (no temperature forcing in the tropical upper troposphere) and are therefore useful for investigating the difference in CNN predicted uncertainty between polar warming and polar cooling. In heating experiments #6 and #7, which contain polar warming, the CNN is less certain (larger σ), in contrast

to heating experiments #8 and #9, which contain polar cooling, where the CNN is more certain (smaller σ).

The CNN's uncertainty when there is a temperature tendency in the upper troposphere of the tropics is more difficult to discern from Figure 4.6 because the CNN's uncertainty is impacted considerably by the temperature tendency at the polar surface. However, heating experiments #3 and #12 include only a temperature tendency in the tropical upper troposphere, one cooling and one warming respectively. These heating experiments suggest that the CNN is more certain when the upper tropical troposphere is cooling rather than warming.

Chapter 5

Discussion

Given the CNN's ability to replicate the sign of the jet-stream's forced response as validated with the additional dry dynamical core experiments, we propose that our approach could be used to support experiment design of model simulations by providing the opportunity to explore responses to a variety of forcings. As a proof of concept, we next demonstrate the CNN's ability to explore jet sensitivities by recreating a sensitivity test completed by a previous study. Butler et al. (2010; afterwards B10) acknowledged that the jet response is sensitive to location, shape, and magnitude of the forcing and worked to identify and document how the atmospheric circulation in a dry dynamical core was sensitive to these small variations in the imposed thermal forcings. To test the sensitivity, B10 ran multiple heating experiments with the Colorado State University general circulation model (Ringler et al., 2000). Although B10 did not quantify the jet shift in the same way as we do here, they did show the zonal-mean zonal wind response and discuss which heating experiments resulted in a stronger wind response. From this we infer the relative magnitudes of the midlatitude jet shifts. Four of the heating experiments performed by B10 are repeated here to analyze how sensitive the circulation is to the height and shape of tropical upper tropospheric heating (see experiments #26-29 in Table A1). Examples of the temperature tendencies are found in Figure 5.1a-d and will be referred to as heating experiments #26, #27, #28, and #29 for this discussion. B10 found that the jet shifts poleward in response to all heating experiments, but the magnitudes of the shifts varied. The results of B10 show heating experiment #26 has the strongest jet response and compressing the forcing vertically (heating experiment #27) or compressing the

forcing in the meridional direction (heating experiment #28) weakens the wind response. In the last scenario (heating experiment #29), B10 shows that when the forcing is compressed vertically and moved lower in the troposphere, resulting in an even weaker zonal wind response.

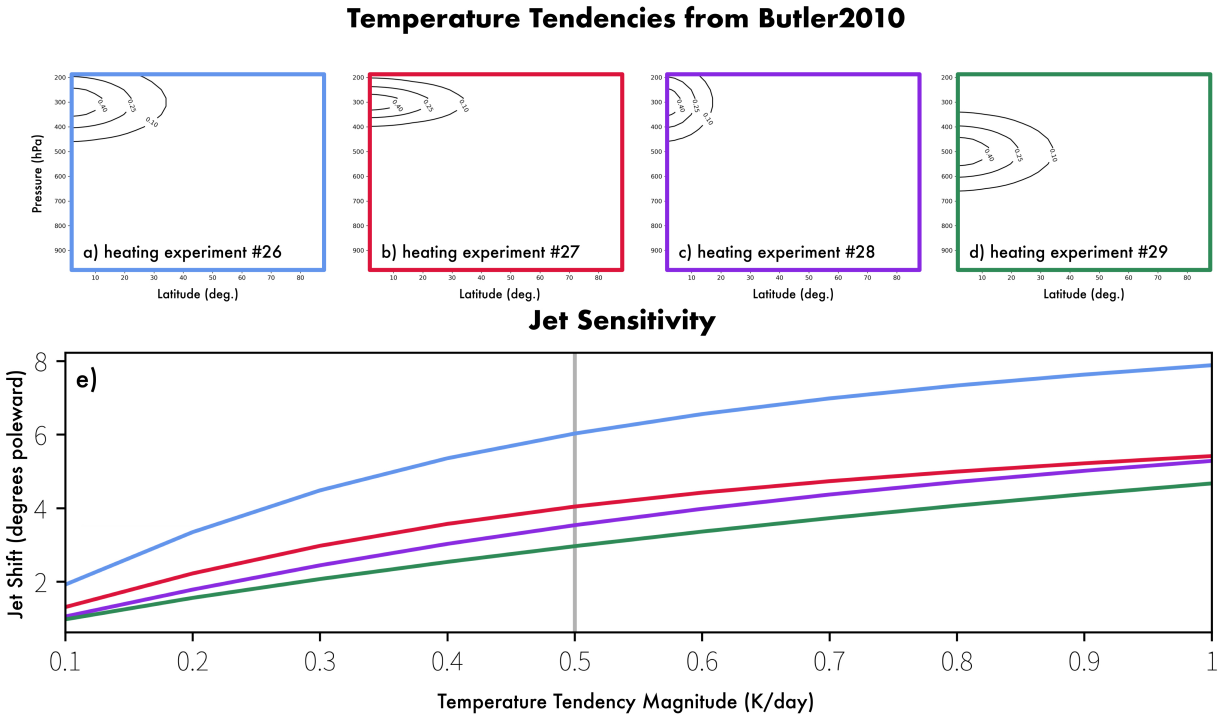

Figure 5.1: Panels (a) - (d) shows example gaussians with magnitudes of 0.5 K day^{-1} representing the four heating experiments used in Butler et al. 2010, gaussian parameters are found in Table A1 #26-29. The panel outline corresponds the predicted jet shifts (y axis) in panel (e) which are shown as a function of the magnitude of the temperate forcing (x axis). The vertical grey line at 0.5 K day^{-1} corresponds to the magnitude of the heating experiments used in Butler et al. 2010.

Following the B10, we input the same forcing patterns (Fig. 5.1a-d) into the CNN with the initial jet location set as the average jet location from the long control run (42.4°). In addition to the original forcing magnitude used in B10 (q_o = 0.5 K day $^{-1}$), here the jet sensitivity to the magnitude of the temperature tendency is also included since it is trivial to explore once the CNN is trained. Figure 5.1e shows the predicted jet shift from the four heating experiments with

varying magnitudes, and the vertical grey line indicates the 0.5 K day $^{-1}$ magnitude used in B10. The CNN predicts the same relative relationship between the heating experiments as found in B10, where heating experiment #26 exhibits the strongest jet response and heating experiment #29 exhibits the weakest. Interestingly, by also exploring the heating magnitudes, Figure 5.1e shows new information about how the jet is sensitive to the strength of the temperature forcing. For example, as the forcing magnitude increases, heating experiment #27 (compressed vertically) and heating experiment #28 (compressed meridionally) converge to the same jet shift. At lower forcing magnitudes, heating experiment #28 and heating experiment #29 (compressed vertically and lower in the troposphere) have a similar jet response.

In this chapter we show a successful example of using the CNN to explore the jet sensitivities inside the dry dynamical core. The CNN is not perfect in its predictions which may be due to lack of predicability, non-optimal training of the CNN, a breakdown of FDT, or a combination of the three. However, as demonstrated throughout this paper, the CNN predicted jet shifts have the same sign and similar magnitudes to the forced jet shifts from the dry core in response to a range of temperature tendencies and once trained, can make predictions quickly. We wish to emphasize that this method should not replace the need to run designed climate model experiments. Rather, training a CNN on an existing long control run could provide the opportunity to explore a large number of forcing experiments before any forced model runs are simulated and be especially helpful for planning forced experiments in dynamical model simulations.

Chapter 6

Conclusion

We explore the jet-stream's response to external forcing by training a CNN on temperature tendencies from a dry dynamical core long control run to predict a shift in the jet-stream's location 30 days later. The main motivation of this work is to explore the potential for training a CNN on internal variability alone and then using it to examine possible nonlinear responses of the jet to tropospheric thermal forcing that more closely resembles forcing due to anthropogenic climate change. Since the CNN is trained entirely on data from a control simulation, it exclusively learns from internal variability. Nevertheless, by comparing the CNN predicted jet shifts to established baselines, peer reviewed literature, and additional dry core heating experiments, we show that the CNN can predict the forced jet shift to sustained forcing. The trained CNN is then used to investigate jet sensitivities to scenarios that mimic the tug-of-war between the tropics and poles under anthropogenic climate change (e.g. Harvey et al., 2015; Stendel et al., 2021). Given the CNN's ability to predict the jet response to thermal forcings, we propose training a CNN to explore model sensitivities to various forcings as a tool to aid in early stage climate model experiment design. Future work could include extending this method to evaluate whether it generalizes to different experimental frameworks, including, but not limited to, evaluating it with three dimensional predictors, training on coupled climate model simulations, and learning complex non-linear climate responses from forced simulations where fluctuation-dissipation theorem does not hold.

Chapter 7

Future Work

Throughout this thesis, we provide strong evidence that we successfully trained a CNN to learn a forced response. Nevertheless, future research should further develop this method to improve the CNN skill and to explore the possible frameworks (i.e. different number of inputs/outputs or different atmospheric variables and models) in which this method can be applied. It is important to note that for the initial application, we attempt to keep the method as simple as possible (i.e. using a dry core, zonally averaging etc.). Increasing the CNN skill and showing that this method works for a range of frameworks will make the method increasingly more useful.

The CNN skill is impacted by choices in network architecture, training parameters, and data processing, and while we investigated these choices when possible (as mentioned throughout thesis), additional combinations and choices should be explored in future work. Modifications to our methods could produce a network that more skillfully predicts a forced jet response from a thermal forcing.

In terms of exploring additional frameworks in future work, one approach to transfer our method to a different framework involves increasing the number of outputs. Instead of predicting a single value that denotes the jet response, include enough outputs so that the CNN can predict the response of the entire wind field. Past work shows that the thermal forcings that comprise the tug-of-war impact the vertical wind shear within the jet (Lee et al., 2019; Stendel et al., 2021). Setting up the CNN to predict the entire wind field provides the opportunity to investigate the response of atmospheric features such as vertical wind shear to a thermal forcing. While this example focuses

on increasing the number outputs, another extension of this work includes increasing the number of inputs. In this work, we use a zonally averaged temperature tendency and an initial jet location, but the dimensions of the input and the variable can vary. For example, the entire zonally average zonal wind field could be included as a CNN input rather than just the single value that denotes the jet location. The inputs and outputs of the CNN are dependent on the goal of the project. Adding more variables may increase the CNN skill but may also make it less effective at investigating a response to a forcing of interest, requiring careful choice by the user.

Applying our method to atmospheric models that simulate more complex interactions, such as coupled climate models, is another way in which evaluation of how well this method transfers to more complex frameworks could be explored. Using a long control run from a coupled climate model to train a CNN would be very beneficial in designing experiments that will later be run in the same coupled climate model. Consider the Polar Amplification Model Intercomparison Project (PAMIP; (Smith et al., 2019)) which assesses the roles of sea ice and sea surface temperatures in driving polar amplification. PAMIP expliores how the climate system responds to changes is sea ice extent by running coupled climate models that are forced with variations in sea ice and sea surface temperatures. Applying the method developed in this thesis, a CNN can be trained to investigate circulation responses to sea ice forcing before running any of the experiments in the coupled climate models. Then, the PAMIP experiments can be designed with the additional information provided by the CNN predictions.

Extending this method to aid in designing experiments for a project such as PAMIP would require evaluating whether the method successfully transfers to a coupled climate model. Further, considering our current method relies on the fluctuation-dissipation theorem (FDT), for this method to extend to systems that do not satisfy FDT, such as atmospheric responses to changes sea ice

extent, our method may require modification such as, but not limited to, different data processing steps or larger neural network. While PAMIP experiments have already been completed, one could apply this method to a future intercomparison project if the method successfully extends to coupled climate models and systems that do not satisfy the FDT.

In this thesis we successfully train a CNN to predict a jet shift from a temperature forcing using only internal variability. Future studies should focus on investigating the transferability and reliability of this approach to different frameworks. Thus exploring ways in which this method is useful beyond exploring the tug-of-war on the jet-stream in a dry dynamical core.

Bibliography

Abbot, D. S., and E. Tziperman, 2008: A high-latitude convective cloud feedback and equable climates. *Quart. J. Roy. Meteor. Soc.*, **134 (630)**, 165–185.

Athanasiadis, P. J., J. M. Wallace, and J. J. Wettstein, 2010: Patterns of wintertime jet stream variability and their relation to the storm tracks. *J. Atmos. Sci.*, **67 (5)**, 1361–1381.

Baker, H. S., T. Woollings, and C. Mbengue, 2017: Eddy-Driven jet sensitivity to diabatic heating in an idealized GCM. *J. Clim.*, **30 (16)**, 6413–6431.

Barnes, E., R. Barnes, and M. DeMaria, 2022: Sinh-arcsinh-normal distributions to add uncertainty to neural network regression tasks: applications to tropical cyclone intensity forecasts.

Barnes, E. A., D. L. Hartmann, D. M. W. Frierson, and J. Kidston, 2010: Effect of latitude on the persistence of eddy-driven jets. *Geophys. Res. Lett.*, **37 (11)**.

Barnes, E. A., and L. Polvani, 2013: Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models. *J. Clim.*, **26 (18)**, 7117–7135.

Bibi, A., K. Ullah, Z. Yushu, Z. Wang, and S. Gao, 2020: Role of westerly jet in torrential rainfall during monsoon over northern pakistan. *Life Support Biosph. Sci.*, **7 (2)**, e2019EA001 022.

Blackport, R., and J. A. Screen, 2020: Insignificant effect of arctic amplification on the amplitude of midlatitude atmospheric waves. *Sci Adv*, **6 (8)**, eaay2880.

Blunden, J., and D. S. Arndt, 2012: State of the climate in 2011. *Bull. Am. Meteorol. Soc.*, **93 (7)**, S1–S282.

Boggs, P. T., and J. E. Rogers, 1990: Orthogonal distance regression. *Contemporary Mathematics*, **112**, 183–194.

Bourdin, D. R., T. N. Nipen, and R. B. Stull, 2014: Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system. *Water Resour. Res.*, **50 (4)**, 3108–3130.

Butler, A. H., D. W. J. Thompson, and R. Heikes, 2010: The Steady-State atmospheric circulation response to climate change–like thermal forcings in a simple general circulation model. *J. Clim.*, **23 (13)**, 3474–3496.

Chen, G., J. Lu, and D. M. W. Frierson, 2008: Phase speed spectra and the latitude of surface westerlies: Interannual variability and global warming trend. *J. Clim.*, **21 (22)**, 5942–5959.

Chen, G., P. Zhang, and J. Lu, 2020: Sensitivity of the latitude of the westerly jet stream to climate forcing. *Geophys. Res. Lett.*, **47 (4)**, 345.

Cohen, J., and Coauthors, 2014: Recent arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7 (9)**, 627–637.

Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Clim. Chang.*, **2 (7)**, 491–496.

Dai, A., D. Luo, M. Song, and J. Liu, 2019: Arctic amplification is caused by sea-ice loss under increasing CO2. *Nat. Commun.*, **10 (1)**, 121.

Deser, C., R. Tomas, M. Alexander, and D. Lawrence, 2010: The seasonal atmospheric response to projected arctic sea ice loss in the late Twenty-First century. *J. Clim.*, **23 (2)**, 333–351.

Druckenmiller, M. L., and Coauthors, 2021: The arctic. *Bull. Am. Meteorol. Soc.*, **102 (8)**, S263–S316.

Dunn-Sigouin, E., C. Li, and P. J. Kushner, 2021: Limited influence of localized tropical sea-surface temperatures on moisture transport into the arctic. *Geophys. Res. Lett.*, **48 (8)**.

Elsberry, R. L., and S. D. Raney, 1978: Sea surface temperature response to variations in atmospheric wind forcing. *J. Phys. Oceanogr.*, **8 (5)**, 881–887.

Erdenebat, E., and T. Sato, 2016: Recent increase in heat wave frequency around mongolia: role of atmospheric forcing and possible influence of soil moisture deficit. **17 (2)**, 135–140.

Evans, K. J., and Coauthors, 2014: A spectral transform dynamical core option within the community atmosphere model (CAM4). *J. Adv. Model. Earth Syst.*, **6 (3)**, 902–922.

Fuchs, D., S. Sherwood, and D. Hernandez, 2015: An exploration of multivariate fluctuation dissipation operators and their response to sea surface temperature perturbations. *J. Atmos. Sci.*, **72 (1)**, 472–486.

Fukushima, K., 1980: Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.*, **36 (4)**, 193–202.

Garfinkel, C. I., D. W. Waugh, and E. P. Gerber, 2013: The effect of tropospheric jet latitude on coupling between the stratospheric polar vortex and the troposphere. *J. Clim.*, **26 (6)**, 2077–2095.

Gerber, E. P., S. Voronin, and L. M. Polvani, 2008: Testing the annular mode autocorrelation time scale in simple atmospheric general circulation models. *Mon. Weather Rev.*, **136 (4)**, 1523–1536.

Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Series B Stat. Methodol.*, **69 (2)**, 243–268.

Gordon, E. M., and E. A. Barnes, 2022: Incorporating uncertainty into a regression neural network enables identification of decadal state-dependent predictability in CESM2. *Geophys. Res. Lett.*, **49 (15)**.

Graversen, R. G., and P. L. Langen, 2019: On the role of the atmospheric energy transport in 2 $\times$ CO2–Induced polar amplification in CESM1. *J. Clim.*, **32 (13)**, 3941–3956.

Grise, K. M., and L. M. Polvani, 2016: Is climate sensitivity related to dynamical sensitivity? *J. Geophys. Res.*, **121 (10)**, 5159–5176.

Gritsun, A., and G. Branstator, 2007: Climate response using a Three-Dimensional operator based on the Fluctuation–Dissipation theorem. *J. Atmos. Sci.*, **64 (7)**, 2558–2575.

Harvey, B. J., L. C. Shaffrey, and T. J. Woollings, 2015: Deconstructing the climate change response of the northern hemisphere wintertime storm tracks. *Clim. Dyn.*, **45 (9)**, 2847–2860.

Held, I. M., 1993: Large-Scale dynamics and global warming. *Bull. Am. Meteorol. Soc.*, **74 (2)**, 228–242.

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Am. Meteorol. Soc.*, **75 (10)**, 1825–1830.

Hunter, J. D., 2007: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, **9 (3)**, 90–95, https://doi.org/10.1109/MCSE.2007.55.

Hwang, Y.-T., and D. M. W. Frierson, 2010: Increasing atmospheric poleward energy transport with global warming. *Geophys. Res. Lett.*, **37 (24)**.

Ingrosso, A., and S. Goldt, 2022: Data-driven emergence of convolutional structure in neural networks. *Proc. Natl. Acad. Sci. U. S. A.*, **119 (40)**, e2201854 119.

Kattsov, V. M., V. E. Ryabinin, J. E. Overland, M. C. Serreze, M. Visbeck, J. E. Walsh, W. Meier, and X. Zhang, 2010: Arctic sea-ice change: a grand challenge of climate science. *J. Glaciol.*, **56 (200)**, 1115–1121.

Kidston, J., and E. P. Gerber, 2010: Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *GEOPHYSICAL RESEARCH LETTERS*, **37 (9)**.

Kim, D., S. M. Kang, T. M. Merlis, and Y. Shin, 2021: Atmospheric circulation sensitivity to changes in the vertical structure of polar warming. *Geophys. Res. Lett.*, **48 (19)**.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. 1412.6980.

Kornhuber, K., D. Coumou, E. Vogel, C. Lesk, J. F. Donges, J. Lehmann, and R. M. Horton, 2019: Amplified rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nat. Clim. Chang.*, **10 (1)**, 48–53.

Kraichnan, R. H., 1959: Classical Flucationation-Relaxation theorem. *Institute of Mathematical Sciences*, **113 (5)**.

Krawczyk, B., 2016: Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, **5 (4)**, 221–232.

LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521 (7553)**, 436–444.

Lee, S., 2014: A theory for polar amplification from a general circulation perspective. *Asia-Pacific Journal of Atmospheric Sciences*, **50 (1)**, 31–43.

Lee, S. H., P. D. Williams, and T. H. A. Frame, 2019: Increased shear in the north atlantic upper-level jet stream over the past four decades. *Nature*, **572 (7771)**, 639–642.

Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.*, **32 (10)**, 2022–2026.

Lim, E.-P., and I. Simmonds, 2009: Effect of tropospheric temperature change on the zonal mean circulation and SH winter extratropical cyclones. *Clim. Dyn.*, **33 (1)**, 19–32.

Ma, Y., H. He, H. He, and Y. Ma, 2013: *Imbalanced Learning : Foundations, Algorithms, and Applications*. John Wiley & Sons, Incorporated, Somerset, UNITED STATES.

MacDonald, G., 1992: Persistence in climate. *The MITRE Coporation*, **JSR–91–340**, 22 102–3481.

Madonna, E., C. Li, C. M. Grams, and T. Woollings, 2017: The link between eddy-driven jet variability and weather regimes in the north Atlantic-European sector. *Quart. J. Roy. Meteor. Soc.*, **143 (708)**, 2960–2972.

Mahlstein, I., O. Martius, C. Chevalier, and D. Ginsbourger, 2012: Changes in the odds of extreme events in the atlantic basin depending on the position of the extratropical jet. *Geophys. Res. Lett.*, **39 (22)**.

Majda, A., R. V. Abramov, and M. J. Grote, 2005: *Information Theory and Stochastics for Multiscale Nonlinear Systems*, Vol. 25. American Mathematical Soc.

Manabe, S., D. G. Hahn, and J. Leith Holloway, 1974: The seasonal variation of the tropical circulation as simulated by a global model of the atmosphere. *J. Atmos. Sci.*, **31 (1)**, 43–83.

Manabe, S., and R. J. Stouffer, 1980: Sensitivity of a global climate model to an increase of CO2concentration in the atmosphere. *J. Geophys. Res.*, **85 (C10)**, 5529.

Marconi, U. M. B., A. Puglisi, L. Rondoni, and A. Vulpiani, 2008: Fluctuation–dissipation: Response theory in statistical physics. *Phys. Rep.*, **461 (4)**, 111–195.

Mattingly, K. S., J. T. McLeod, J. A. Knox, J. M. Shepherd, and T. L. Mote, 2015: A climatological assessment of greenland blocking conditions associated with the track of hurricane sandy and historical north atlantic hurricanes. *Int. J. Climatol.*, **35 (5)**, 746–760.

Mbengue, C., and T. Schneider, 2013: Storm track shifts under climate change: What can be learned from Large-Scale dry dynamics. *J. Clim.*, **26 (24)**, 9923–9930.

McGraw, M. C., and E. A. Barnes, 2016: Seasonal sensitivity of the Eddy-Driven jet to tropospheric heating in an idealized AGCM. *J. Clim.*, **29 (14)**, 5223–5240.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: THE WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Am. Meteorol. Soc.*, **88 (9)**, 1383–1394.

Nakamura, H., T. Sampe, Y. Tanimoto, and A. Shimpo, 2004: Observed associations among storm tracks, jet streams and midlatitude oceanic fronts. *Earth's Climate: The Ocean–Atmosphere Interaction, Geophys. Monogr*, **147**, 329–345.

Nipen, T., and R. Stull, 2011: Calibrating probabilistic forecasts from an NWP ensemble. *Tellus Ser. A Dyn. Meteorol. Oceanogr.*, **63 (5)**, 858–875.

Nix, D. A., and A. S. Weigend, 1994: Estimating the mean and variance of the target probability distribution. *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, Vol. 1, 55–60 vol.1.

Nix, D. A., and A. S. Weigend, 1995: Learning local error bars for nonlinear regression. *Advances in Neural Information Processing Systems*, **7 ((NIPS'94))**, 489–496.

Phipps, S. J., 2010: *The CSIRO Mk3L climate system model v1.2*. 4 121 pp., CRC Tech. rep.

Pithan, F., and T. Mauritsen, 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nat. Geosci.*, **7 (3)**, 181–184.

Pithan, F., B. Medeiros, and T. Mauritsen, 2014: Mixed-phase clouds cause climate model biases in arctic wintertime temperature inversions. *Clim. Dyn.*, **43 (1)**, 289–303.

Polvani, L. M., 2002: Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophys. Res. Lett.*, **29 (7)**.

Prechelt, L., 2012: Early stopping — but when? *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 53–67.

Ringler, T. D., R. P. Heikes, and D. A. Randall, 2000: Modeling the atmospheric general circulation using a spherical geodesic grid: A new class of dynamical cores. *Mon. Weather Rev.*, **128 (7)**, 2471–2490.

Röthlisberger, M., S. Pfahl, and O. Martius, 2016: Regional-scale jet waviness modulates the occurrence of midlatitude weather extremes. *Geophys. Res. Lett.*, **43 (20)**, 10,989–10,997.

Rousi, E., K. Kornhuber, G. Beobide-Arsuaga, F. Luo, and D. Coumou, 2022: Accelerated western european heatwave trends linked to more-persistent double jets over eurasia. *Nat. Commun.*, **13 (1)**, 3851.

Rousi, E., F. Selten, S. Rahmstorf, and D. Coumou, 2021: Changes in north atlantic atmospheric circulation in a warmer climate favor winter flooding and summer drought over europe. *J. Clim.*, **34 (6)**, 2277–2295.

Schubert, S., H. Wang, and M. Suarez, 2011: Warm season subseasonal variability and climate extremes in the northern hemisphere: The role of stationary rossby waves. *J. Clim.*, **24 (18)**, 4773–4792.

Screen, J. A., I. Simmonds, C. Deser, and R. Tomas, 2013: The atmospheric response to three decades of observed arctic sea ice loss. *J. Clim.*, **26 (4)**, 1230–1248.

Shaw, T. A., and Coauthors, 2016: Storm track processes and the opposing influences of climate change. *Nat. Geosci.*, **9**, 656–664.

Sherwood, and S. Nidhi Nishant, 2015: Atmospheric changes through 2012 as shown by iteratively homogenized radiosonde temperature and wind data (IUKv2). *Environ. Res. Lett.*, **10 (5)**, 054 007.

Smith, D. M., and Coauthors, 2019: The polar amplification model intercomparison project (PAMIP) contribution to CMIP6: investigating the causes and consequences of polar amplification. *Geosci. Model Dev.*, **12 (3)**, 1139–1164.

Stendel, M., J. Francis, R. White, P. D. Williams, and T. Woollings, 2021: Chapter 15 - the jet stream and climate change. *Climate Change (Third Edition)*, T. M. Letcher, Ed., Elsevier, 327–357.

Swart, N. C., and J. C. Fyfe, 2012: Observed and simulated changes in the southern hemisphere surface westerly wind-stress. *Geophys. Res. Lett.*, **39 (16)**.

Tjernström, M., J. Sedlar, and M. D. Shupe, 2008: How well do regional climate models reproduce radiation and clouds in the arctic? an evaluation of ARCMIP simulations. *J. Appl. Meteorol. Climatol.*, **47 (9)**, 2405–2422.

Vihma, T., 2014: Effects of arctic sea ice decline on weather and climate: A review. *Surv. Geophys.*, **35 (5)**, 1175–1214.

Virtanen, P., and Coauthors, 2020: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272, https://doi.org/10.1038/s41592-019-0686-2.

Woollings, T., A. Hannachi, and B. Hoskins, 2010: Variability of the north atlantic eddy-driven jet stream. *Quart. J. Roy. Meteor. Soc.*, **136 (649)**, 856–868.

Wyrtki, K., 1975: El niño-the dynamic response of the equatorial pacific ocean to atmospheric forcing. *Journal of Physical Oceanography*, **5 (4)**, 572–584.

Yann, L., B. Leon, B. Yoshia, and P. Haffner, 1998: Gradient-Based learning applied to document recognition. *Proceedings of the IEEE*.

Yeh, S.-W., and Coauthors, 2018: ENSO atmospheric teleconnections and their response to greenhouse gas forcing. *Reviews of Geophysics*, **56 (1)**, 185–206.

Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.*, **32 (18)**.

Zeiler, M. D., and R. Fergus, 2014: Visualizing and understanding convolutional networks. *European conference on computer vision*, Springer, 818–833.

Appendix A

Table of 2-D Gaussians parameters

Experiment	Θ_0 (degree poleward)	Θ_w (degree poleward)	p_0 (hPa)	p_w (hPa)	q0 (K)
#1	90	16	1000	250	1.0
	0	27	300	125	-0.1
#2	90	16	1000	250	0.5
	0	27	300	125	-0.1
#3	0	27	300	125	-0.1
#4	90	16	1000	250	-0.5
	0	27	300	125	-0.1
#5	90	16	1000	250	-1.0
	0	27	300	125	-0.1
#6	90	16	1000	250	1.0
#7	90	16	1000	250	0.5
#8	90	16	1000	250	-0.5
#9	90	16	1000	250	-1.0
#10	90	16	1000	250	1.0
	0	27	300	125	0.1
#11	90	16	1000	250	0.5
	0	27	300	125	0.1
#12	0	27	300	125	0.1
#13	90	16	1000	250	-0.5
	0	27	300	125	0.1
#14	90	16	1000	250	-1.0
	0	27	300	125	0.1
#15	60	15	1000	200	0.25
#16	60	15	700	200	0.25
#17	30	15	1000	200	0.25
#18	30	15	700	200	0.25
#19	V	10	V	150	1.0
#20	12	15	850	200	V
#21	75	15	850	200	V
#22	45	15	550	200	V
#23	12	15	300	200	V
#24	75	15	300	200	V
#25	90	16	1000	250	V
	0	27	300	125	V
#26	0	27	300	125	V
#27	0	27	300	75	V
#28	0	13.5	300	125	V
#29	0	27	500	125	V

Table A.1: Parameters for 2-D gaussians from forced dry core heating experiments and CNN temperature forcings, "V" indicates varying parameter.

Appendix B

PIT Hisogram

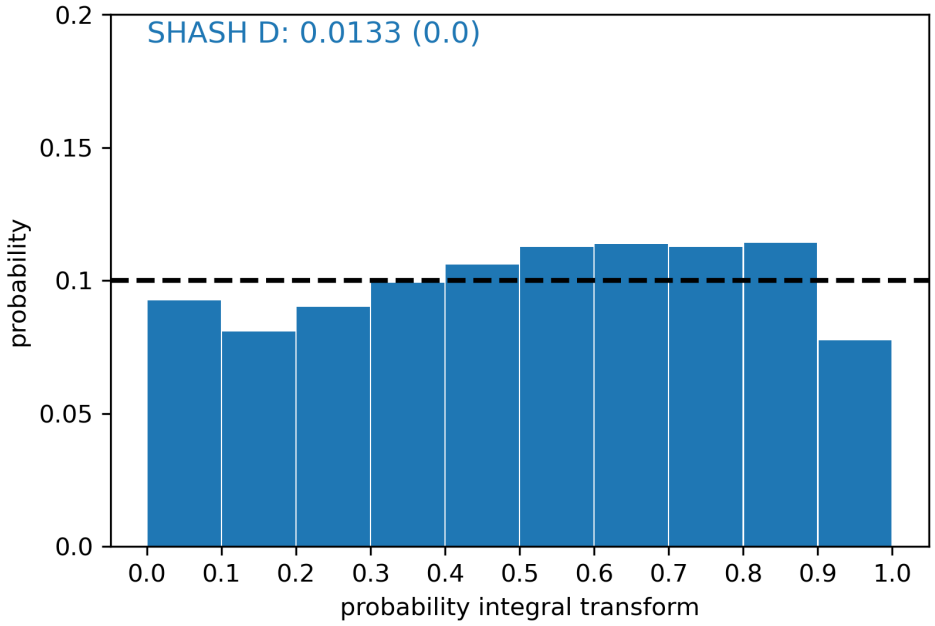

Figure B.1: PIT histogram from testing data. The deviation (D) from flatness is found in upper right corner along with the expected calibration deviation in parenthesis.

The convolutional neural network used in this study makes a probabilistic prediction (i.e. predicting a jet shift and an uncertainty value). To ensure the network is calibrated, we employ the probability integral transform (PIT) calibration metric (Nipen and Stull, 2011). Determining the calibration of the network increases reliability in the predicted uncertainty values. A calibrated network would result in a flat PIT histogram (Gneiting et al., 2007), and the PIT histogram for the CNN used in this study is shown in Figure B.1. We can calculate the deviation (D) from a flat histogram using Equation B.1 (Nipen and Stull, 2011; Bourdin et al., 2014):

$$D = \sqrt{\frac{1}{B}\sum_{i=1}^{B}(\frac{b_i}{T} - \frac{1}{B})^2} \tag{B.1}$$

where i iterates through the number of bins (B) and b_i is the number of samples in that bin. The flatter the histogram, the smaller the value. The deviation from flatness for this network, found in the upper left corner of Figure B.1, is 0.0133. The expected value of D for a calibrated network is calculated using Equation B.2 (Nipen and Stull, 2011; Bourdin et al., 2014):

$$E[D_{perfect}] = \sqrt{\frac{1 - B^{-1}}{TB}} \tag{B.2}$$

For this network, the expected deviation, also found in the upper left corner of Figure B.1, is 0.0. The expected deviation is a function of the number of bins (B) and the amount of data (T) used to calculate the PIT histogram. In this case, the PIT histogram is calculated with all 1,399,558 testing samples. This many samples causes Equation B.1 to approach zero.

We can investigate how the deviation value (D) changes when a CNN is trained on more data. Using the same methods and CNN outlined in the Methods, we train six additional CNNs with different amounts of training data. All six networks have the same validation and testing data. Figure B.2 shows the deviation value (D) of each PIT histogram, and the mean absolute error for six different CNN's trained with increasing amounts of data. As seen by the red line, which represents the mean absolute error of the network, training a CNN with more than 400,000 samples does

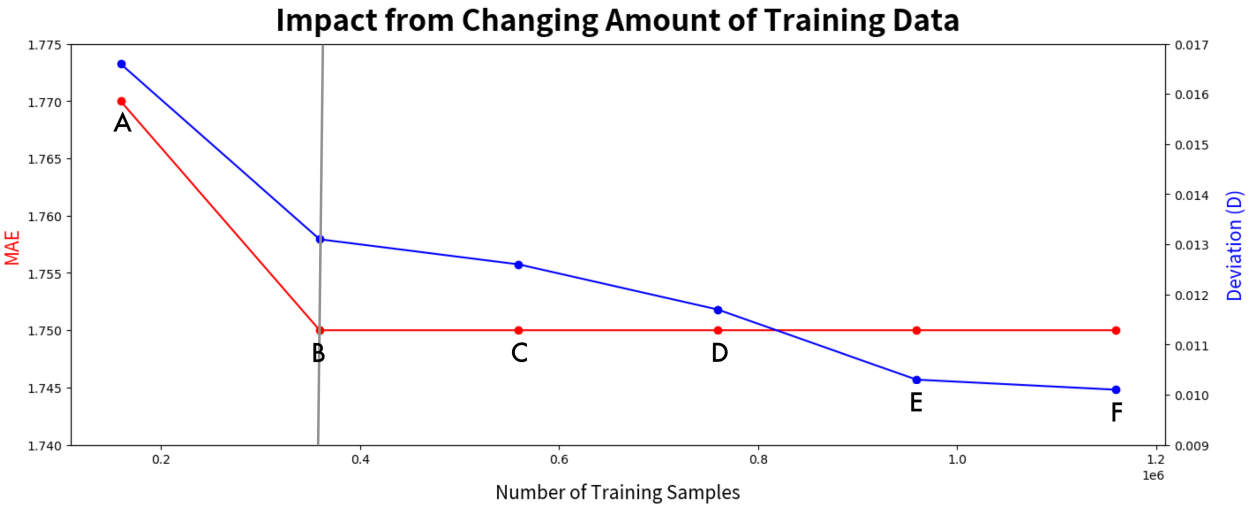

Figure B.2: Six CNNs, labeled with letters in figure, trained on a different number of samples (x axis). Red line indicates the mean absolute error for each network. Blue line indicates the Deviation form flatness in the PIT histogram for each network. Grey line indicates the amount of data used to train the CNN used throughout this study.

not improve the network's deterministic mean absolute error but it does decrease the PIT D value. Based on Figure B.2, adding more data would continue to improve the network calibration. While we have the ability to continuously add data to our training dataset, we have chosen not to here since not every model or study has the ability to make an unlimited amount of data. Importantly, while the values of the predicted uncertainties changed between CNN B (used in this study) and CNN F (more calibrated CNN) the results discussed throughout this paper do not change (not shown).

Appendix C

CNN skill at the midlatitude (30° and 60° latitude) surface

Predicted jet shift's from the CNN do not always have the same sign as forced jet shifts from the CESM dry dynamical core. Noted in Section 4.1 in the discussion of Figure 4.3, the CNN does not agree in all locations with the forced jet shifts seen in Baker et al. (2017). At the surface of 30° latitude the CNN predicts an equatorward shift and at the surface of 60° latitude the CNN predicts a poleward shift, inverse in sign of the forced jet shift found in Baker et al. (2017). We ran four heating experiments in the CESM dry dynamical core to investigate this region of conflicting jet shifts to establish whether these inconsistencies were a failure of the CNN or an outcome of different methods between the two studies.

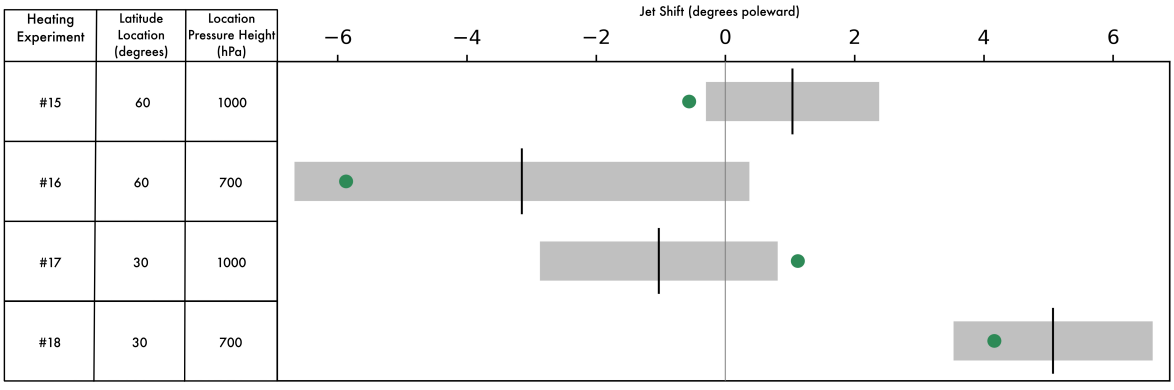

Figure C.1: The 4 heating experiments, their resulting jet shift in the dry core, and the jet shift as predicted by the CNN, that are similar to Baker et al. 2017. The left hand side shows the gaussian experiment and its location. On the right hand side, green dots represent the true jet shift from the dry core, the black line represents the predicted jet shift from the CNN and the grey boxes denote the CNN uncertainty ($\pm 2\sigma$) defined as the sigmas output.

The four additional heating experiment parameters can be found in Appendix A experiment #15-18. The left side Figure C.1 shows the location of the four heating experiments (q_o = 0.25 K day $^{-1}$) and the right side shows the forced dry core jet shift (green dot), the CNN predicted jet shift (black line), and the CNN predicted uncertainty (2σ; grey box). Heating experiment #18 has a CNN predicted jet shift of the same sign and similar magnitude while heating experiment #16 has the same predicted sign and a different magnitude, though the true jet shift is captured in the large uncertainty. The heating experiments at the surface, #15, and #17, have CNN predicted jet shifts of different signs and that are not within the predicted uncertainty bounds. These heating experiments are examples of where the CNN fails to capture the correct jet response. Interestingly, Gritsun (2007) documented that fluctuation-dissipation theorem (FDT) failed to replicate the jet response at the surface due to true weak responses in the system. Since our approach of training a CNN on internal variability to predict the forced jet response fundamentally relies, in part, on FDT, it stands to reason that regions and situations where FDT fails, the CNN may also fail.