Deep Learning for Scientific Inference from Geophysical Data: The Madden-Julian Oscillation as a Test Case

Benjamin A. Toms\textsuperscript{1,}*, Karthik Kashinath\textsuperscript{2}, Prabhat\textsuperscript{2}, Da Yang\textsuperscript{2,3}

\textsuperscript{1}Department of Atmospheric Science, Colorado State University, Fort Collins, CO
\textsuperscript{2}Lawrence Berkeley National Laboratory, Berkeley, California
\textsuperscript{3}University of California, Davis, Davis, California

Key Points:

- Deep learning models can accurately identify and classify the state of complex geophysical phenomena
- ...and they can understand the correct spatial structure of geophysical phenomena
- ...and they can independently learn variables important for the characterization of geophysical phenomena

\textsuperscript{*}1371 Campus Delivery, Fort Collins, CO 80523

Corresponding author: Benjamin A. Toms, ben.toms@colostate.edu
Abstract
Deep learning can recognize complex geophysical phenomena by inferring which variables are important for their identification and understanding their spatial characteristics.

We use a particular mode of multiscale tropical atmospheric variability, the Madden-Julian oscillation (MJO), to study the capabilities of deep learning, a form of artificial intelligence and machine learning, in identifying spatial geophysical phenomena. The MJO is characterized by its spatial and temporal evolution of cloud patterns, and an extensive body of literature has examined its defining characteristics. By applying a convolutional neural network (CNN), a type of deep learning model, to the task of identifying the state of the MJO, we show that deep learning can correctly identify geophysical phenomena by “learning” the variables and spatial patterns important to their evolution. In a broader sense, these findings suggest that deep learning models are interpretable and viable for scientific inference in geoscientific applications.

1 Motivation
Machine learning techniques have been applied in a broad range of subfields of atmospheric science, including in the development of climate model parameterizations (Brenowitz & Bretherton, 2018; Gentine et al., 2018; O’Gorman & Dwyer, 2018; Rasp et al., 2018), the improvement of convective-scale forecasting (Herman & Schumacher, 2018a, 2018b; Jin et al., 2008), and the supplementation of forecaster decision making tools (McGovern et al., 2017). A new subset of AI techniques has emerged in recent years in response to a surge in data availability and computational capabilities. These techniques, colloquially dubbed deep learning (LeCun et al., 2015), offer a fresh landscape of opportunity for AI-driven discovery within atmospheric science and the broader geosciences and therefore deserve thorough exploration. To this end, we present a novel application of a type of deep learning, convolutional neural networks, to spatio-temporal geophysical patterns within the tropical atmosphere.

2 Introduction
Recent research has shown deep learning to be an effective method for identifying weather and climate patterns such as atmospheric rivers and tropical cyclones within geospatial fields (Kurth et al., 2017; Liu et al., 2016). Unlike the aforementioned phenomena, some atmospheric phenomena are characterized by superimposed waves and are therefore less visually apparent in geospatial fields, which makes the extraction of scientifically meaningful information more difficult. The Madden-Julian oscillation (MJO) is such a phenomenon.

The MJO is a recurring equatorial wave with a periodicity of approximately 30-to-80 days, and is observed as a region of increased cloudiness that forms over the Indian Ocean and propagates toward the central Pacific (Hendon & Salby, 1994; Madden & Julian, 1971, 1972; Zhang, 2004). A broad range of atmospheric processes constitute its existence, ranging from deep convective clouds on the scale of tens to hundreds of kilometers within the Indo-Pacific region to an equatorial circumnavigating Kelvin wave response on the scale of thousands of kilometers (Hendon & Liebmann, 1994; Powell, 2017). The MJO has been the focus of many publications that have expanded our understanding of both its tropical (Adames & Wallace, 2015, 2014; Kiladis et al., 2014; Wheeler & Hendon, 2004) and extratropical (Henderson et al., 2017; Garfinkel et al., 2014; Tseng et al., 2017) character. Through this work, a basis for the physical mechanisms driving its existence has been developed, including which variables co-vary throughout its evolution.
Given the well-established knowledge about its characteristics, the MJO offers an opportunity to test the capabilities of deep learning in the geosciences, particularly for multi-scale spatio-temporal phenomena. Deep learning models are notorious for being physically indecipherable, and are colloquially dubbed “black box models”. And so an important first step is to ensure the method can first corroborate established science before extending into the unknown. In the case of the MJO, this test may be in the form of recognizing the covariance and spatial evolution of various atmospheric state variables, such as temperature and humidity.

We approach the application of deep learning to the MJO by answering two questions: 1) Can a deep learning method be used to identify the state of the MJO with high accuracy?; 2) If so, is the deep learning model accurate because it “understands” the covariance of atmospheric state variables and their spatial evolution throughout the life-cycle of the MJO? We first guide the reader through a brief introduction to convolutional neural networks (CNNs), then delve into the application of a CNN to the identification of the MJO.

3 Deep Learning Architecture: Convolutional Neural Networks

We use a CNN to classify the MJO according to spatial fields of atmospheric state variables. CNNs are designed to extract information from an image in its raw form directly without vectorizing the information. The MJO, which is detailed in the next section, is intrinsically defined by its spatial structure, so the usage of a CNN preserves its spatial identity. The specific details of the CNN architecture are available within the supporting information or on GitHub (https://github.com/benatoms/MJO_DeepLearning). Its general design is detailed in Figure 1.

We will briefly discuss the general structure and logic of CNNs, although the reader is referred to other resources such as Goodfellow et al. (2016) for additional details. CNNs consist of a series of layers that are constructed from convolution matrices. Each convolution matrix of size $N \times N$, the size of which is set by the user, has $N^2$ weights and a single bias parameter that are iteratively adjusted throughout the training of the CNN. For classification, the CNN’s task is to learn the weights within these convolution matrices that, when applied to the image, extract information that helps differentiate between the image classes. The outputs from each convolution layer are known as filters, the number of which is specified by the user during the creation of the CNN. The values input into each subsequent layer are the outputs, or filters, from the previous convolution layer that have been passed through a nonlinear activation function. We use an activation function called the leaky rectified linear unit (leaky ReLU), which is simply a function that applies the following operator to each point within an image, after the convolution has been applied:

$$f(x) = \begin{cases} 
    x & \text{if } x > 0 \\
    ax & \text{if } x < 0
\end{cases}$$

where $a$ is typically a rather small value, in this case 0.003.

The CNN’s success is scored using a loss function, which measures the difference between estimated and truth values. In our case, we use the negative log likelihood loss function, defined as:

$$L = \sum_{n=1}^{N} l_n$$

where $L$ is the total loss and $l_n$ is the logarithm of the probability that an image can be classified into any of the $N$ classes, which is output by the CNN. The logarithm is applied to punish cases where the probability is split evenly across multiple classes and to
reward the model when it is more confident in its decision. The samples are passed through the network in batches, which are subsets of the training dataset, and the loss is calculated after each batch. We used a batch size of 64 samples.

After a batch has passed through the network, the weights and biases of each convolution matrix are updated by first calculating the gradient of the loss function (Equation 2) via backpropagation (a series of chain-rule operations). An increment is then taken along this gradient, which leads toward the minimum of the loss function which is further associated with a minimum in the classification error of the CNN. We use the Adam method to determine how large of an increment should be taken (Kingma & Ba, 2014) with an initial learning rate of 1E-3.

4 The Test Case: The Madden-Julian Oscillation

4.1 Defining the Madden-Julian Oscillation

We define the MJO according to the Bimodal Intraseasonal Oscillation Index (BISO) of Kikuchi et al. (2012) and focus on the wintertime mode, which generally occurs from December through April (Kikuchi et al., 2012). BISO is constructed using principal component analysis of outgoing longwave radiation (OLR) anomalies, which are a proxy for cloud cover based on the amount of longwave radiation that is emitted from earth into space, within the 30°S to 30°N band. BISO defines the state of the MJO according to its progression through eight phases of cloud pattern evolution (Figure 2). Phase 1 characterizes the initiation of the cloud anomalies of the MJO, from which it typically progresses eastward, sequentially through the eight phases.

4.2 Data

The CNN is trained on daily OLR data from the National Oceanic and Atmospheric Administration OLR Daily Climate Data Record (Lee, 2014) and daily-mean atmospheric state variables from MERRA-2 reanalysis (Gelaro et al., 2017) for the years 1980 through 2017. Sixteen atmospheric fields are input into the CNN for each sample: OLR and 850 hPa, 500 hPa, and 200 hPa zonal and meridional wind components, temperature, specific humidity, and geopotential (example shown in Figure 3). The point of including more variables than is likely necessary to identify the MJO is to test whether the CNN can independently infer which variables are important for describing the MJO. It is important to note that because we are using the BISO index, the classification of the MJO is based solely on OLR anomalies. All fifteen other fields are therefore additional information that the CNN can use according to its own “judgment”. We temporally filter the input variables using the method proposed by Kikuchi et al. (2012), and each variable is normalized across the 1980 through 2017 period.

Active MJO days are defined when the BISO principal component magnitude is greater than 1 standard deviation (σ) in magnitude. The principal component magnitude threshold for inactive days ensures an approximately equal number of days in each of the nine categories of MJO activity and is therefore set to 0.4-σ. In total, approximately 3,500 days within the 1980 to 2017 period are used to train the model, approximately evenly split across the nine states of the MJO (MJO phases 1 through 8 and inactive MJO).

Monthly periods of the data are randomly distributed into training and validation datasets, with 90% of the samples for training and the remaining 10% for validation. The separation of the data into monthly chunks before randomly selecting the training and validation data limits the issue of autocorrelation within the MJO spatial fields when testing the accuracy of the CNN using the validation dataset. Figure 3 shows the distribution of the training and validation datasets within the MJO phase space. The val-
idation dataset is similar to the training dataset, since the mean phase and amplitude of the validation samples are similar to those of the training samples for each phase.

5 CNN Accuracy

The accuracy of the CNN is measured based on whether it can correctly identify, or classify, the phase of the MJO into one of nine MJO phases: MJO phases one through eight or inactive MJO. During training, the CNN is given the class of the input sample based solely on the BISO index, which only uses OLR information. The CNN correctly classifies the MJO phase for approximately 90% of the validation samples, and has a one-phase error for 9% of the validation samples. The CNN struggles to correctly classify the MJO phase along the phase transitions, which is likely related to the fact that the MJO is a continuous wave and the end of one discretized phase is appreciably similar to the start of the next (Figure 4). Since the CNN can accurately classify the phase of the MJO with only slight confidence errors during the transitions between phases, we subsequently use the model to test the capabilities of CNNs in independently extracting meaningful information about the processes they are tasked with identifying.

6 Interpretation of the CNN

The objective of this study is to test not only whether CNNs can be used to accurately identify complex geophysical patterns, but also whether they do so because they capture the correct physical characteristics of those patterns. Regardless of the application, CNNs have proven difficult to physically interpret given the large number of parameters, such as convolutional matrix weights, that are updated throughout training (in this case 180,000). The investigation of each individual parameter is therefore impractical and alternative techniques must be considered.

The spatial patterns the CNN considers important for identifying each class can be tested using a method called feature optimization, wherein an initially meaningless input image is iteratively adjusted to maximize the CNN's confidence that it belongs in a certain class by minimizing the loss function as defined in Equation 2 (in this case the initial image is all zeros). Figure 5 shows the optimized patterns for the identification of each phase of the MJO. The eastward propagating OLR anomaly within the Indo-Pacific is immediately apparent, and there are numerous other regions of anomalous cloudiness separated from the primary cloud anomaly, particularly over the Americas. The spatial patterns of cloud identified by the CNN are similar to those of the BISO index used for the truth classification, although there are some minor differences, particularly over the eastern Pacific and the Americas.

Permutation importance (PI) is a common method for estimating how important a trained machine learning model “thinks” a variable is for the correct classification of a sample (Strobl et al., 2008). In PI, the samples are passed through the classifier but with one of the input variables permuted (i.e. shuffled), and variables that are more important for the model's classification will cause a greater decrease in classification accuracy when permuted. While the truth classification for MJO phase is based solely on OLR (as discussed within Section 4), the CNN considers 850 hPa and 200 hPa zonal winds (U850 and U200) to be of substantial importance, as well (Figure 6a). Lower and upper tropospheric zonal wind anomalies have been identified by an extensive number of studies to be important descriptors of the overall evolution of the MJO. In fact, numerous indices of the MJO include information about zonal wind anomalies (Maloney & Hartmann, 2001; Wheeler & Hendon, 2004; Zhang & Hendon, 1997), whereas the index used in this study only uses information about OLR. The CNN also deems 200 hPa temperature anomalies (T200) to be of importance, although the reasoning for this is less clear. Overall, the results of the feature importance and PI tests support the idea that the de-
tails the CNN is learning about the MJO corroborate previous research, and the CNN is therefore correctly characterizing the MJO.

7 Discussion and Conclusions

We have shown that convolutional neural networks (CNNs), a form of deep learning, are capable of accurately identifying the state of multi-scale spatio-temporal geophysical phenomena by independently inferring which variables are most important for their characterization. Furthermore, CNNs can accurately reproduce the most fundamental spatial patterns associated with the geophysical phenomena that they are trained to identify. These findings suggest that when deep learning is used for scientific applications in geoscience, if the model is accurate it is possible that it is accurate because the model is capturing patterns that are physically reasonable, although one must check to ensure this is the case using methods such as those presented within this paper.

It is likely still important that deep learning models be trained using information that can guide the model to a reliable answer, as is the case with this study. After being offered information about the phase of the MJO using one variable (outgoing longwave radiation), the model is able to independently infer which other variables are important for describing the phase evolution of the MJO, as corroborated by other process-oriented studies (lower- and upper-tropospheric zonal wind anomalies; e.g. Adames and Wallace (2014) and Wheeler and Hendon (2004)). This capability offers the opportunity to more expediently explore hypotheses related to variable importance, along with confidence that accurate deep learning models can independently learn which variables are most important for the description of the phenomena. Deep learning has the potential to transform how geoscientists approach data analytics, hypothesis testing, and parameterization development, particularly if the “black box” is unraveled and the deep learning outputs are shown to be consistent with physical theory.

Acknowledgments

This work was funded by the Department of Energy Computational Science Graduate Fellowship via grant DE-FG02-97ER25308. Thanks to Aryeh Drager and Jessica Tomaszewski for help improving the manuscript. All of the data used in this study is publicly available as detailed in the text.

References

Adames, A. F., & Wallace, J. M. (2014). Three-dimensional structure and evolution of the MJO and its relation to the mean flow. *Journal of the Atmospheric Sciences, 71*(6), 2007-2026. doi: 10.1175/JAS-D-13-0254.1

Adames, A. F., & Wallace, J. M. (2015). Three-dimensional structure and evolution of the moisture field in the MJO. *Journal of the Atmospheric Sciences, 72*(10), 3733-3754. doi: 10.1175/JAS-D-15-0003.1

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters, 45*(12), 6289-6298. doi: 10.1029/2018GL078510

Garfinkel, C. I., Benedict, J. J., & Maloney, E. D. (2014). Impact of the MJO on the boreal winter extratropical circulation. *Geophysical Research Letters, 41*(16), 6055-6062. doi: 10.1002/2014GL061094

Gelaro, R., McCarty, W., Surez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate, 30*(14), 5419-5454. doi: 10.1175/JCLI-D-16-0758.1

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could
machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11), 5742-5751. doi: 10.1029/2018GL078202

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (http://www.deeplearningbook.org)

Henderson, S. A., Maloney, E. D., & Son, S.-W. (2017). Madden–Julian Oscillation pacific teleconnections: The impact of the basic state and MJO representation in general circulation models. *Journal of Climate*, 30(12), 4567-4587. doi: 10.1175/JCLI-D-16-0789.1

Hendon, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Hendon, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Hendon, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, S. A., Maloney, E. D., & Son, S.-W. (2017). Madden–Julian Oscillation pacific teleconnections: The impact of the basic state and MJO representation in general circulation models. *Journal of Climate*, 30(12), 4567-4587. doi: 10.1175/JCLI-D-16-0789.1

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Hendon, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.

Henderson, H. H., & Salby, M. L. (1994). The life cycle of the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, 51(15), 2225-2237. doi: 10.1175/1520-0469(1994)051⟨2225:TLCOTM⟩2.0.CO;2

Henderson, H. H., & Liebmann, B. (1994). Organization of convection within the Madden–Julian oscillation. *Journal of Geophysical Research: Atmospheres*, 99(D4), 8073–8083.
to improve real-time decision-making for high-impact weather. 
Bul-
elin of the American Meteorological Society, 98(10), 2073-2090. 
doi: 10.1175/BAMS-D-16-0123.1

O’Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameter-
ize moist convection: Potential for modeling of climate, climate change and 
extreme events. ArXiv e-prints.

Powell, S. W. (2017). Successive MJO propagation in MERRA-2 reanalysis. Geo-
physical Research Letters, 44(10), 5178-5186. doi: 10.1002/2017GL073399

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid 
processes in climate models. Proceedings of the National Academy of Sciences,
115(39), 9684–9689. doi: 10.1073/pnas.1810286115

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008, Jul 11). 
Conditional variable importance for random forests. BMC Bioinformatics,
9(1), 307. doi: 10.1186/1471-2105-9-307

Tseng, K.-C., Barnes, E. A., & Maloney, E. D. (2017). Prediction of the midlatitude 
response to strong Madden–Julian Oscillation events on S2S time scales. Geo-
physical Research Letters, 45(1), 463-470. doi: 10.1002/2017GL075734

Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate 
MJO index: Development of an index for monitoring and prediction. Monthly 
Weather Review, 132(8), 1917-1932. doi: 10.1175/1520-0493(2004)132<1917:
AARMJI2.0.CO;2

Zhang, C. (2004). Madden–Julian Oscillation. Reviews of Geophysics, 43(2). doi: 10
.1029/2004RG000158

Zhang, C., & Hendon, H. H. (1997). Propagating and standing components of the 
intrasessional oscillation in tropical convection. Journal of the atmospheric sci-
ces, 54(6), 741–752.
Figure 1. The general structure of the CNN architecture used in this study. Note that the dimensions of the inputs and outputs of the convolution layers and pooling layers are meant as examples, and do not reflect the true progression of image sizes throughout the CNN. Refer to the supporting information for a specific list of operators within the CNN. Each sample contains sixteen variables and the CNN ultimately classifies each sample into one of nine classes: MJO phases one through eight or inactive MJO. The CNN contains approximately 180,000 parameters (weights and biases).
Figure 2.  a) The spatial phase evolution of outgoing longwave radiation anomalies of the MJO derived from the BISO index (Kikuchi et al., 2012). The anomalies are for an event with a 1-σ principal component magnitude of the BISO index. The hollow white circles denote the approximate centroid of the region of increased cloudiness. b) A phase space perspective of the MJO, representing the linear combination of the two principal components within the BISO index. The red dots denote the locations within the phase space that correspond to the spatial perspective in (a) for each respective phase. The red arrows show the direction the MJO evolves throughout the phase space.
Figure 3. a) Example training sample from January 26, 2014, when the MJO was in phase 7 with a principal component magnitude of 1.29. The sixteen variables shown are included within each sample input into the CNN. b) Training (gray) and validation (blue) samples within the MJO phase space. The blue (gray) circles denote the mean validation (training) dataset principal component magnitude and phase for each respective MJO phase. The red dot denotes the phase of the sample shown in a.
Figure 4. a) Classification for all samples (including both testing and validation) within the MJO phase space. Blue dots denote a one-phase classification error, while red dots denote a classification error of two or more phases. b) Two-dimensional histogram of probabilistic classification accuracy showing the correct phase (y-axis) and the predicted phase from the CNN (x-axis), with the average probabilities assigned by the CNN for each class listed in the center of each box.
Figure 5. a) The definition of the MJO based on its eight-phase evolution of outgoing long-wave radiation anomalies according to Kikuchi et al. (2012); b) the CNN interpretation of the outgoing longwave radiation patterns important for correctly classifying the phase of the MJO. Note that the scale of importance in (b) is rescaled between -1 and 1, since feature optimization does not necessarily retain the scaling of the original field.
Figure 6. Permutation importance for the sixteen variables input into the CNN. The y-axis denotes the reduction in classification accuracy when one of the variables is permuted, while the abbreviation for each variable shown in Figure 3 is listed along the x-axis. The red lines denote the median importance, while the upper and lower quartiles are denoted by the bounding boxes and the 5th and 95th percentile confidence bounds are denoted by the whiskers. The uncertainty bounds are estimated by permuting each variable 100 times.