TOWARDS A DATA-DRIVEN UNDERSTANDING OF CLOUD STRUCTURE FORMATION

Ann-Christin Wörl & Michael Wand
Institute of Computer Science
Johannes Gutenberg University, Mainz, Germany
{awoerl, wandm}@uni-mainz.de

Peter Spichtinger
Institute for Atmospheric Physics
Johannes Gutenberg University, Mainz, Germany
spichtin@uni-mainz.de

ABSTRACT

The physics of cloud formation and evolution is still not fully understood and constitutes one of the highest uncertainties in climate modeling. We are working on an approach that aims at improving our understanding of how clouds of different structures form from a data-driven perspective: By predicting the visual appearance of cloud photographs from physical quantities obtained from reanalysis data and subsequently attributing the decisions to physical quantities using "explainable AI" methods, we try to identify relevant physical processes. At the current stage, this is just a proof of concept, being at least able to identify basic meteorologically plausible facts from data.

1 INTRODUCTION

Current numerical weather and climate prediction models face challenges in effectively modeling cloud physics whose dynamics cover a large breadth of spatial and temporal scales. Achieving a comprehensive understanding of cloud physics is a significant objective within the field of atmospheric modeling. In this paper, we describe work-in-progress towards a data-driven approach with the goal of supporting atmospheric physicists in understanding the relevant processes better.

Our approach is rather simple and straightforward: We use ERA5-reanalysis data – i.e., a "gold standard" reconstruction of the past physical state of the atmosphere – and use a regressor based on a deep convolutional network (U-Net) trained to predict the visible channels of high-resolution satellite data (gray-scale images taken by NASA satellites in the visible spectrum over water, thus primarily depicting clouds) from physical state. This turns out to work surprisingly well, even exceeding the matching quality in terms of the normalized-cross-correlation (NCC) score by almost a factor of two over just taking the built-in "cloud cover" channel of the same ERA5 dataset.

Next, we use "masking" of input channels (optimizing for the minimum amount of input channels through an added sparsity prior) to identify the input most relevant for the predictions. This reveals which physical quantities are most predictive of the cloud structure. Some findings (e.g., temperatures being disregarded in the tropics, but being important in mid latitudes) are not surprising, but show that in principle such an approach can uncover physical insights from observational data.

2 RELATED WORK

Machine learning has been applied to a wide range of fields in the physical sciences. Gentine et al. (2018) utilize machine learning techniques to improve the parameterization of convection in climate simulations at a coarse scale. Rolnick et al. (2019) provide a comprehensive overview of the current and potential applications of machine learning in addressing climate change. Toms et al. (2020) use layerwise relevance propagation (Bach et al., 2015) to prove that neural networks can identify the coherent spatial patterns of known modes of Earth system variability.

As neural networks gain popularity, explanation methods are becoming increasingly relevant. Not only to increase confidence in the prediction, but also to establish a clear connection between the input and output and to provide a physical interpretation. Most of them can be categorized as either gradient-based or non-local perturbation methods.

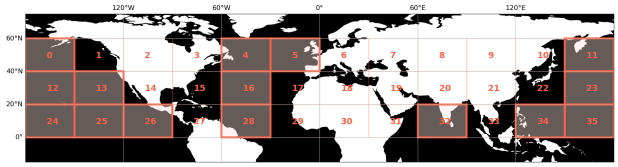


Figure 1: Generated subpatches between $0°$ and $60°$ Northern latitude with used regions highlighted

Gradient-based methods like SmoothGrad (Smilkov et al., 2017), Integrated Gradients (Sundararajan et al., 2017) or GradCAM (Selvaraju et al., 2017) use some kind of back propagation of the output score with respect to the input to assign importance scores, while non-local perturbation methods like DeepLIFT (Shrikumar et al., 2019), SHAP (Lundberg & Lee, 2017) or LIME (Ribeiro et al., 2016) modify the input in a specific way. A subcategory of perturbation methods are masking approaches. A distinction can be made between removing (Fong & Vedaldi, 2017; Qi et al., 2020) or preserving evidence (Khorram et al., 2021).

3 METHOD

3.1 DATA

ECMWF Reanalysis v5 (ERA5): In our work, we try to predict cloud structures based on physical conditions. Therefore, we use the ERA5 dataset (Hersbach et al., 2023) of global atmospheric reanalyses (i.e., a best fit of an atmospheric simulation against most available observational data). It offers high-resolution data with a spatial resolution about 30 kilometers (0.25 degrees in longitude and latitude) and hourly temporal resolution. For vertical resolution, we choose model levels, which resolve the atmosphere from the surface up to a height of approximately 24 kilometers. The ERA5 dataset covers a variety of physical quantities. We restrict ourselves to the most basic ones: temperature, specific humidity, cloud liquid and ice water content, horizontal wind and vertical velocity.

Satellite Images: The second dataset used in this work consists of satellite images based on polar orbiting satellites Terra and Aqua, hosted by the NASA. These satellites capture multi-spectral imagery in various spectral bands, providing comprehensive coverage of Earth's surface and atmosphere. In this work, we restrict ourselves to two-dimensional cloud images derived from the visible spectrum. The satellites cross each location on Earth two times a day. Since we are interested in the visible spectrum, we only use the daytime passing.

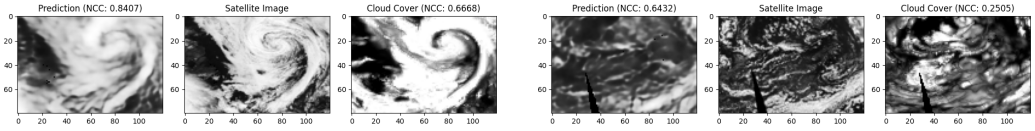
Data Preparation: We run our algorithm in a sliding-window fashion, dividing the world into subpatches of $30°$ longitude and $20°$ latitude. To avoid the distortion at the polar caps, we limit ourselves to the range of $0°$ and $60°$ North. To clearly separate the clouds from the background we only consider regions that are mostly over water. Fig. 1 shows the resulting tiling with used regions highlighted. So far, we only use 2 years of data, split into training and validation sets, and map satellite and ERA5 data accordingly.

3.2 ARCHITECTURE & METRICS

Cloud Structure Prediction: Our first goal is to predict two-dimensional cloud structures based on physical data. Therefore, we use a U-Net architecture (Ronneberger et al., 2015), which is a convolutional neural network commonly used for segmentation tasks. In our experiment, we perform image reconstruction (with a least-squares-loss) instead. We adapt the architecture to work with high-dimensional data and use ERA5 subpatches of shape $343 \times 120 \times 80$ as input. Grayscale satellite images of shape 120×80 serve as the target.

We train our network for 100 epochs, employing an ADAM optimizer (Kingma & Ba, 2014) and integrating a one-cycle learning rate scheduler (Smith & Topin, 2017). To account for the variability of cloud types across regions and their corresponding physical conditions, we train individual neural networks for each region. This strategy ensures that our models capture and adapt to the unique characteristics of different geographical areas.

Predictive Input Identification: The primary goal of our work is to establish a correlation between physical quantities and cloud appearance in a data-driven manner. Traditional saliency methods



(a) (b)

Figure 3: Comparison of cloud structures. Left: Our model prediction, Middle: Satellite image, Right: ERA5 cloud cover

(such as input-domain gradient methods) highlight regions that the classifier responds to in an unspecific way; the utility for uncovering causal connections between input and output has thus been drawn into question heavily in the literature (Adebayo et al., 2018; Dombrowski et al., 2019; Woerl et al., 2023). We use a channel masking approach instead, which is known to work much better for such tasks (Khorram et al., 2021).

Mamalakis et al. (2022) suggest that the choice of the right baseline is crucial in saliency methods and needs to be considered. We decide to use the channels of real data samples as baseline, because we think that this guarantees that we stay on the right distribution. Let $\boldsymbol{M} \in [0,1]^c$ be a weight matrix, whose entries correspond to the importance scores of the input channels c_i. Furthermore, let $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^{c \times w \times h}$ be two ERA5 data samples. The masked version of $\boldsymbol{x}$ is then calculated as

$$\tilde{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{x} + (1 - \boldsymbol{M})\boldsymbol{z}. \tag{1}$$

Let $\mathcal{P}$ be our cloud prediction network and $\mathcal{P}(\boldsymbol{x})$ be the predicted cloud field. Our goal is to determine the optimal solution for $\boldsymbol{M}$ such that $\|\mathcal{P}(\boldsymbol{x}) - \mathcal{P}(\tilde{\boldsymbol{x}})\| \rightarrow \min.$ with an additional sparsity prior on $\boldsymbol{M}$. This problem is solved by a second neural network.

Normalized-Cross-Correlation (NCC): To evaluate the quality of our cloud structure prediction we use the normalized-cross-correlation (Goshtasby, 2012) to measure the similarity between our prediction and the real satellite image. The NCC is a commonly used measure capturing linear correlation between two images with input scale invariance, yielding outputs between -1 and 1.

4 Results

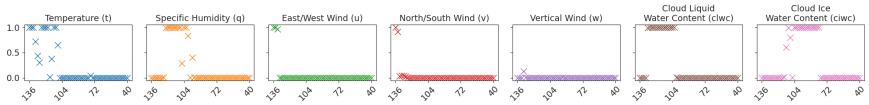
4.1 Nowcasting Cloud Prediction

Two example predictions of our model are shown in fig. 3, with additional examples available in the appendix. The cloud field generated by our model is shown in the left column, while the original satellite image is displayed in the middle column. The cloud cover used in ERA5 is shown in the right column for comparison purposes only, as it was not utilized in the training process.

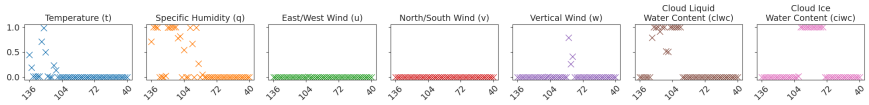
From a visual inspection, it can be seen that our model provides a better representation of the overall cloud structure. To not only rely on visual inspection, we computed the NCC for both our model and the ERA5 cloud cover. The distributions for the whole validation set are shown in fig. 5. The ERA5 cloud cover distribution is plotted in orange with a mean of 0.57, while our model distribution is plotted in blue with a mean of 0.76.

4.2 Predictive Input Identification

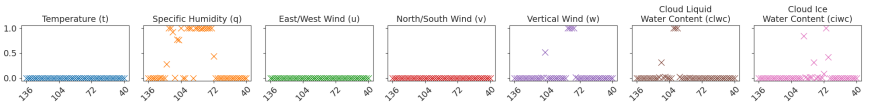
We obtain importance scores for physical input quantities at different height levels from their masking weights. We want to study whether location has an effect on the explanation and thus train separate models for each region. Fig. 4 shows one region each from the mid latitudes (4a), the subtropics (4b) and the tropics (4c). The colored subplots correspond to the seven variables of ERA5. On the x-axis are the model levels, ranging from 136 at ground level to 40 at an altitude of about 24 km. The y-axis shows the importance score between zero and one with values in between being artefacts due to the sparsity prior. To further examine the relevance of physical quantities, we compute the principal components of variation over all 15 regional importance scores. We obtain (mostly) one dominant axis of variation from our 15 sample regions. Fig. 6 shows the projection on this axis color-coded from blue to red. It shows a pronounced gradual change in relevance structure from North to South, with the same but smaller changes when moving away from coastlines.



(a) Region 05: mid latitudes



(b) Region 12: subtropics



(c) Region 35: tropics

Figure 4: Importance scores for different variables and multiple regions

4.3 Forecasting Cloud Structures

So far, we have focused on predicting and explaining cloud structures in the present. In order to put more focus on cloud formation processes, we also try to predict cloud formations one day in advance. To achieve this, we combine satellite images with ERA5 data from 24 hours prior. Similar to the nowcasting model, we computed the NCC for the entire validation dataset. The distribution is depicted in green in fig. 5. The cloud structures overall become less similar to the nowcasting model (with a mean of 0.65 compared to 0.76), but they are still larger to the ERA5 cloud cover, which is an indicator of feasibility.

When comparing the importance scores for different regions with and without time shift, it appears that the wind, particularly the u- and v-components, becomes more significant. In contrast, the other physical quantities become less relevant. It seems that knowledge of individual layers is sufficient to make good predictions. This is also a plausible and not unexpected observation since wind speed and direction directly affect the physical conditions one day ahead.

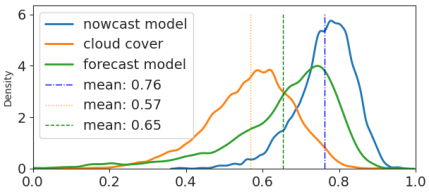


Figure 5: Comparison of NCC distributions for satellite image validation data: ERA5 Cloud Cover vs. Nowcast Model vs. Forecast Model

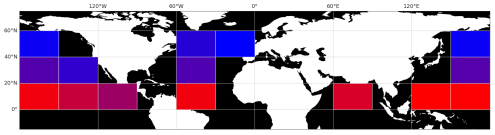


Figure 6: Projection of each regional importance scores on first principal component of variation in a color-coded fashion

5 Discussion

In this work, we explore the correlation between physical data and actual cloud appearance using a data-driven approach. We show that neural networks can produce two-dimensional cloud fields from physical data that surpass the performance of the algorithms typically utilized in contemporary numerical weather prediction models, even when using data one day in advance. Additionally, we proposed a data-driven method for generating a physical explanation of cloud structures.

Limitations and future work: Limiting our experiments to only 15 different regions and two years of training data restricts the expressiveness of our analysis. To address this, we plan to expand our dataset to cover a wider time period and spatial area. We also intend to adjust the tiling size to provide a more detailed explanation.

REFERENCES

Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. *Advances in neural information processing systems*, 31, 2018.

Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7):e0130140, 2015.

Ann-Kathrin Dombrowski, Maximilian Alber, Christopher J. Anders, Marcel Ackermann, Klaus-Robert Müller, and Pan Kessel. Explanations can be manipulated and geometry is to blame, 2019. URL https://arxiv.org/abs/1906.07983.

Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE international conference on computer vision*, pp. 3429–3437, 2017.

Pierre Gentine, Mike Pritchard, Stephan Rasp, Gael Reinaudi, and Galen Yacalis. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018.

A Ardeshir Goshtasby. Similarity and dissimilarity measures. *Image registration: Principles, tools and methods*, pp. 7–66, 2012.

H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J-N. Thépaut. Era5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023.

Saeed Khorram, Tyler Lawson, and Li Fuxin. Igos++: Integrated gradient optimized saliency by bilateral perturbations. In *Proceedings of the Conference on Health, Inference, and Learning*, CHIL '21, pp. 174–182, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383592. doi: 10.1145/3450439.3451865. URL https://doi.org/10.1145/3450439.3451865.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL https://arxiv.org/abs/1412.6980.

Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017. URL https://arxiv.org/abs/1705.07874.

Antonios Mamalakis, Elizabeth A. Barnes, and Imme Ebert-Uphoff. Carefully choose the baseline: Lessons learned from applying xai attribution methods for regression tasks in geoscience, 2022.

Zhongang Qi, Saeed Khorram, and Li Fuxin. Visualizing deep networks by optimizing with integrated gradients. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34 (07):11890–11898, April 2020. ISSN 2159-5399. doi: 10.1609/aaai.v34i07.6863. URL http://dx.doi.org/10.1609/aaai.v34i07.6863.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier, 2016. URL https://arxiv.org/abs/1602.04938.

David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning, 2019.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences, 2019.

Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda B. Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *CoRR*, abs/1706.03825, 2017. URL http://arxiv.org/abs/1706.03825.

Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates, 2017. URL https://arxiv.org/abs/1708.07120.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks, 2017. URL https://arxiv.org/abs/1703.01365.

Benjamin A Toms, Elizabeth A Barnes, and Imme Ebert-Uphoff. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS002002, 2020.

Ann-Christin Woerl, Jan Disselhoff, and Michael Wand. Initialization noise in image gradients and saliency maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1766–1775, 2023.