Supporting Information for "Deep learning based cloud cover parameterization for ICON"

Arthur Grundner[1,2], Tom Beucler[3], Pierre Gentine[2], Fernando Iglesias-Suarez[1], Marco A. Giorgetta[4], and Veronika Eyring[1,5]

[1]Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[2]Columbia University, Center for Learning the Earth with Artificial intelligence And Physics (LEAP), New York, NY 10027, USA

[3]University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland

[4]Max Planck Institute for Meteorology, Hamburg, Germany

[5]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

Contents

Introduction This supplementary information provides more detailed information concerning the data and the neural networks (NNs). It describes the variables that were used as input features for the NNs, illustrates the architecture of the three NN types, the space of hyperparameter we explored and the preprocessing and amount of (training) data for each network. Table S1 specifies the parameterization schemes used in the NARVAL and QUBICC simulations. The cross-validation split for the QUBICC (R2B5) models is

depicted in Figure S3. Figure S1 illustrates the coefficients of a multiple linear model trained on the NARVAL (R2B4) data. Figures S4 and S5 cover aspects of the generalization capability of the NARVAL networks across regions and resolutions. Lastly, Figure S6 shows that SHAP values do not strongly depend on the base value.

---

1. Definition and Choice of Input Parameters for the NNs

1. **land**: The land fraction (in $[0, 1]$) is used in the ICON-A cloud cover scheme to discern whether one might have to artificially increase relative humidity in order to take thin maritime stratocumuli into account.

2. **lake**: The lake fraction (in $[0, 1]$) is a parameter closely related to the land fraction. A supply of moisture from the ground very likely influences the distribution of moisture in the atmospheric column above, especially in the presence of convection.

3. **Cor.**: The Coriolis parameter (in $1/s$) allows the cloud cover parameterization to vary between different latitudes, which can be especially useful with global training data.

4. $\mathbf{q_v}$, $\mathbf{T}$, $\mathbf{p}$, $\mathbf{z_g}$: Specific humidity (in kg/kg), air temperature (in K), pressure (in Pa) and geometric height at full levels (in m). These are the most important input variables for the original ICON-A cloud cover scheme (to compute relative humidity).

5. $\mathbf{q_c}$, $\mathbf{q_i}$: The specific cloud water content and the specific cloud ice content (in kg/kg). They have a direct influence on cloudiness as the presence of cloud water or ice is a necessary requirement for the presence of clouds. In this spirit, they are for instance used in an alternative 0-1 cloud cover scheme in ICON-A, which sets cloud cover to 1 when a certain threshold of cloud condensate is crossed.

6. $\boldsymbol{\rho}$: Air density (in kg/m^3). We left it out for the R2B5 NNs, since air density can mostly be derived from p, T and q_v by using the ideal gas law and is therefore redundant.

7. $\mathbf{u}$, $\mathbf{v}$: Zonal/eastward wind and meridional/northward wind (in m/s). Vertical wind shear can induce a large difference between cloud area fraction and cloud cover.

8. $\mathbf{clc_{t-1}}$: The cloud cover estimate (in $[0, 100]\%$) from the previous timestep (1 hour before). Undeniably, clouds have a memory effect on this time scale. However, a model that relies on previous cloudiness cannot be used in the first time step.

The features ρ, u, v are also used in the Tompkins scheme of cloud cover (Tompkins, 2002).

2. Preprocessing

The preprocessing, which we define as distinct from coarse-graining, consists of up to four steps:

1. **For all cell-based and QUBICC neighborhood-based models (N1, Q1 and Q3)**: Ensure that the amount of data samples with $clc \neq 0$ is as large (for the Q1 model twice as large to reduce the data size) as the one with $clc = 0$, by downsampling the latter class of cloud-free data samples.

2. **For the neighborhood-based NARVAL models (N3)**: Remove the cloud cover from the first time step of each day of the NARVAL data from the output. We cannot predict it, because there is no previous cloud cover value which the neighborhood-based NARVAL model would require as input.

3. **QUBICC data**: Remove the first time steps of the simulations because that output incorrectly consists of an entirely cloud-free atmosphere. Scale the cloud cover to be in $[0, 100]\%$. Convert the data from float64 to float32 to reduce the data size.

4. **For the QUBICC cell- and neighborhood-based models (Q1 and Q3)**: Subsample only every third hour from the QUBICC data set to reduce the data size. Assuming a high temporal correlation, we should not lose a lot of information. Remove condensate-free clouds ($\sim 7\%$ of all clouds).

5. **For all models (N1-N3, Q1-Q3)**: Normalize the actual training data so that each input feature to the NN is distributed according to a Gaussian with zero mean and unit variance. In the column-based models this means that the normalization is done on a level-by-level basis and for the cell-based and neighborhood-based models we have one level-independent mean and standard deviation per input feature. According to Brenowitz and Bretherton (2019), we expect the impact on our results due to these different choices of normalization to be very small. This step of normalization can only be done after splitting the set of all training data samples into subsets of training, validation and test data.

3. Space of Hyperparameters

We explored the following space of hyperparameters used in the neural network training:

1. Number of units per hidden layer: 16, 32, ..., 512
2. Number of hidden layers: From 1 to 4
3. Activation functions: ReLU, ELU, tanh, leaky ReLU with $\alpha \in \{0.01, 0.2\}$
4. Initial learning rate: From 10^{-4} to 1

5. Epsilon parameter in the optimizer: 10^{-8}, 10^{-7}, 0.1, 1

6. Dropout: With or without after each hidden layer with parameters from 0 to 0.5

7. L1/L2-regularization: With parameters from 0 to 0.01

8. Batch normalization: With or without after each layer

References

Barker, H. W., Stephens, G., Partain, P., Bergman, J., Bonnel, B., Campana, K., ... Yang, F. (2003). Assessing 1d atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds. *Journal of Climate*, *16*(16), 2676–2699.

Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2728-2744. doi: 10.1029/2019ms001711

Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M., ... others (2011). A description of the nonhydrostatic regional cosmo model, part ii: Physical parameterization. *Deutscher Wetterdienst, Offenbach, Germany*.

Mauritsen, T., Svensson, G., Zilitinkevich, S. S., Esau, I., Enger, L., & Grisogono, B. (2007). A total turbulent energy closure model for neutrally and stably stratified atmospheric boundary layers. *Journal of Atmospheric Sciences*, *64*(11), 4113–4126.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663–16682.

Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency, and

flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3074-3089. doi: 10.1029/2019MS001621

Raddatz, T., Reick, C., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., ... Jungclaus, J. (2007). Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate dynamics*, *29*(6), 565–574.

Raschendorfer, M. (2001). The new turbulence parameterization of lm. *COSMO newsletter*, *1*, 89–97.

Schrodin, R., & Heise, E. (2001). *The multi-layer version of the dwd soil model terra_lm.* DWD.

Schulz, J.-P., Vogel, G., Becker, C., Kothe, S., & Ahrens, B. (2015). Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil. In *Egu general assembly conference abstracts* (p. 6549).

Seifert, A. (2008). A revised cloud microphysical parameterization for cosmo-lme. *COSMO Newsletter*, *7*, 25–28.

Tompkins, A. M. (2002). A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *Journal of the Atmospheric Sciences*. doi: https://doi.org/10.1175/1520-0469(2002)059⟨1917:APPFTS⟩2.0.CO;2

Table S1. Parameterizations used in the NARVAL and QUBICC simulations

	NARVAL	**QUBICC**
Cloud Cover	Diagnostic PDF	All-or-nothing scheme based on cloud condensate
Microphysics	Single-moment scheme (Doms et al., 2011; Seifert, 2008)	Single-moment scheme (Doms et al., 2011; Seifert, 2008)
Radiation	RRTM scheme (Barker et al., 2003; Mlawer et al., 1997)	RTE+RRTMGP scheme (Pincus et al., 2019)
Turbulence	Prognostic TKE (Raschendorfer, 2001)	Total turbulent energy scheme (Mauritsen et al., 2007)
Land	Tiled TERRA (Schrodin & Heise, 2001; Schulz et al., 2015)	JSBach4-lite (Raddatz et al., 2007)

Table S2. Amount of training data samples for the NNs. The tuples denote either (time steps, vertical layers, horizontal fields) or (time steps, horizontal fields). Note that for the R2B4 neighborhood-based model we trained one NN per vertical layer, so the number of training samples is equal to the number of training samples for the R2B4 column-based model. Grid columns containing grid cells that were omitted during coarse-graining are excluded in the 'After coarse-graining'-column and are also not used for training.

	Original data ($\leq 21\,\mathrm{km}$)	After coarse-graining	After preprocessing
Cell-based			
R2B4 NARVAL	$5.6 \cdot 10^{11}$ $(1721, 66, 4887488)$	$4.5 \cdot 10^{7}$ $(1635, 27, 1024)$	$3.7 \cdot 10^{7}$
R2B5 QUBICC	$3.9 \cdot 10^{12}$ $(2162, 87, 20971520)$	$4.6 \cdot 10^{9}$ $(2162, 27, 78069)$	$8.8 \cdot 10^{8}$
Neighborhood-based			
R2B4 NARVAL	$8.4 \cdot 10^{9}$ $(1721, 4887488)$	$1.7 \cdot 10^{6}$ $(1632, 1024)$	$1.7 \cdot 10^{6}$
R2B5 QUBICC	$3.9 \cdot 10^{12}$ $(2162, 87, 20971520)$	$4.6 \cdot 10^{9}$ $(2162, 27, 78069)$	$1.2 \cdot 10^{9}$
Column-based			
R2B4 NARVAL	$8.4 \cdot 10^{9}$ $(1721, 4887488)$	$1.7 \cdot 10^{6}$ $(1635, 1024)$	$1.7 \cdot 10^{6}$
R2B5 QUBICC	$4.5 \cdot 10^{10}$ $(2162, 20971520)$	$1.7 \cdot 10^{8}$ $(2162, 78069)$	$1.7 \cdot 10^{8}$

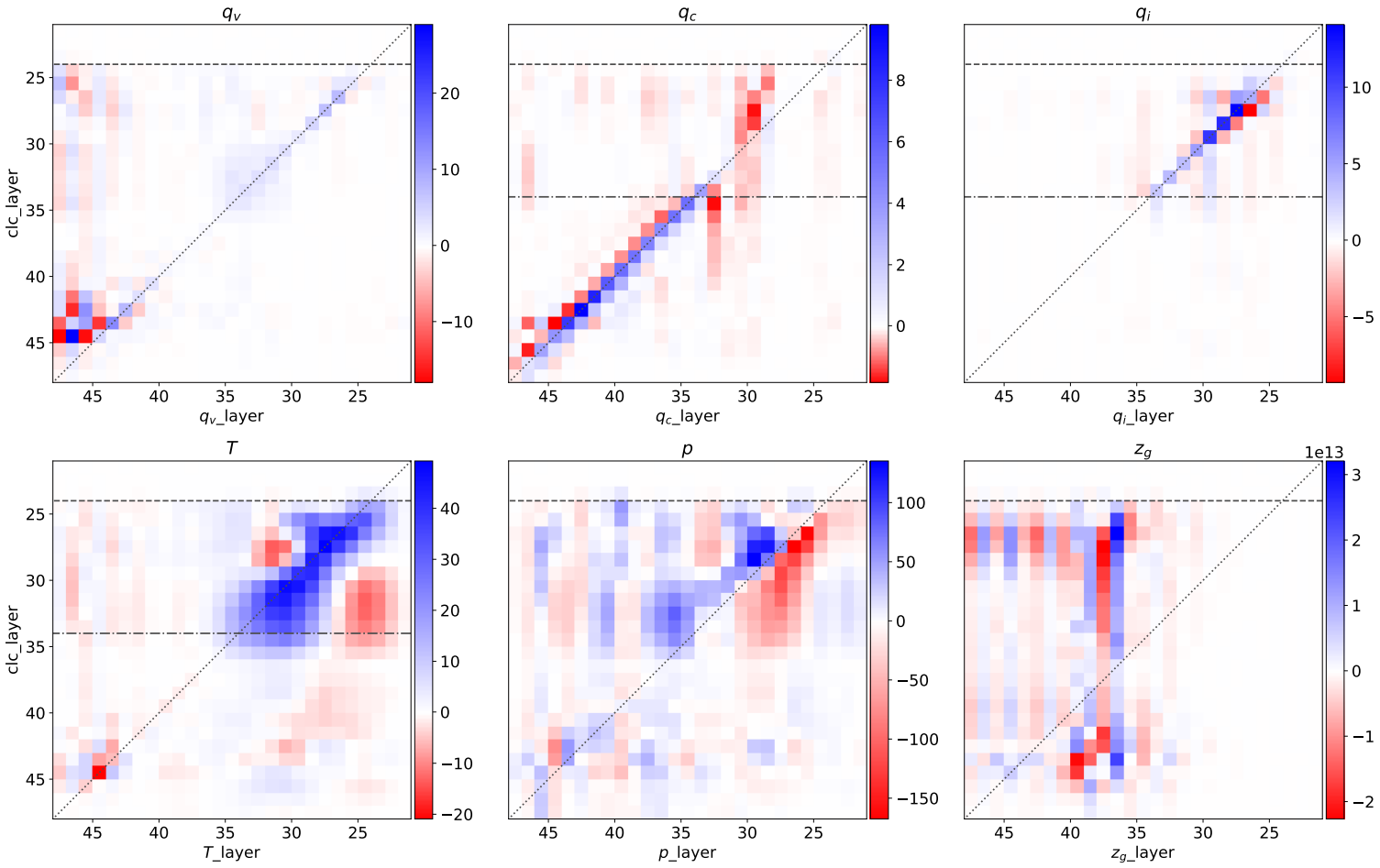


Figure S1. Coefficients of the best multiple linear model on standardized NARVAL R2B4 data. The dashed line shows the tropopause ($\approx 15\,\mathrm{km}$), the dash dotted line shows the freezing level (i.e. where temperatures are on average below 0 degrees) ($\approx 5\,\mathrm{km}$) and the dotted line visualizes the diagonal. The coefficients suggest that the problem of diagnosing cloud cover is non-local. The zg coefficients seem to dominate. An elevated grid cell on level 15 increases cloud cover significantly. However, due to the nature of the vertical grid, the layers below will also be elevated, driving a decrease of cloud cover. An increase in specific humidity, cloud water (at altitudes below the freezing level) and cloud ice (at altitudes above the freezing level) increase cloudiness in the same grid cell. In the upper troposphere, when we increase the pressure, we force the condensation of water vapor at the given level and above.

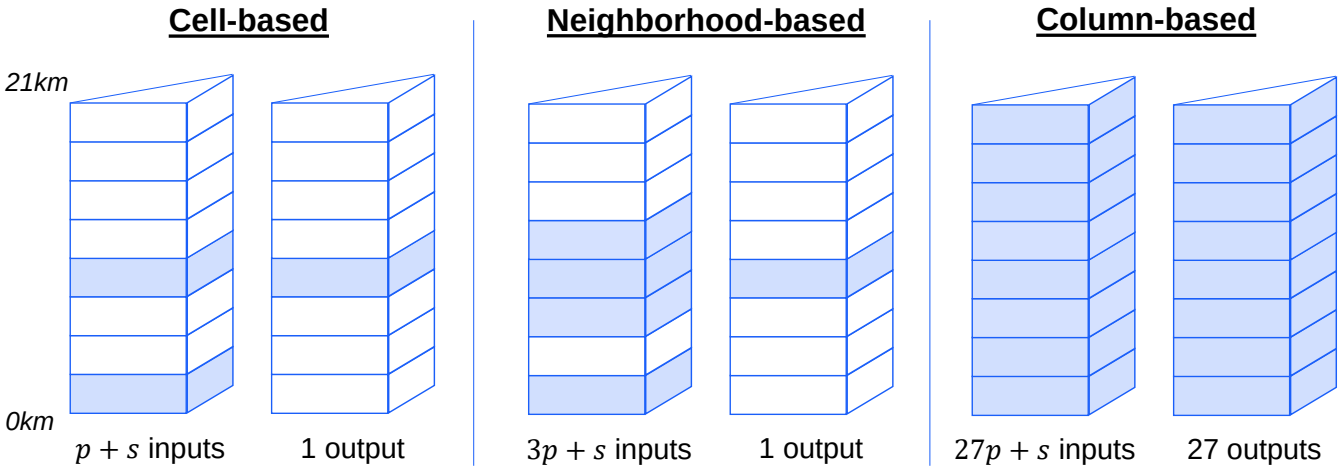


Figure S2. A sketch of the three NN types based on one grid column. The variable p denotes the number of input features from the grid cells and s is the number of extra variables from the surface. In this sketch, the neighborhood-based model uses two neighboring cells, which is only true for our QUBICC-trained NN.

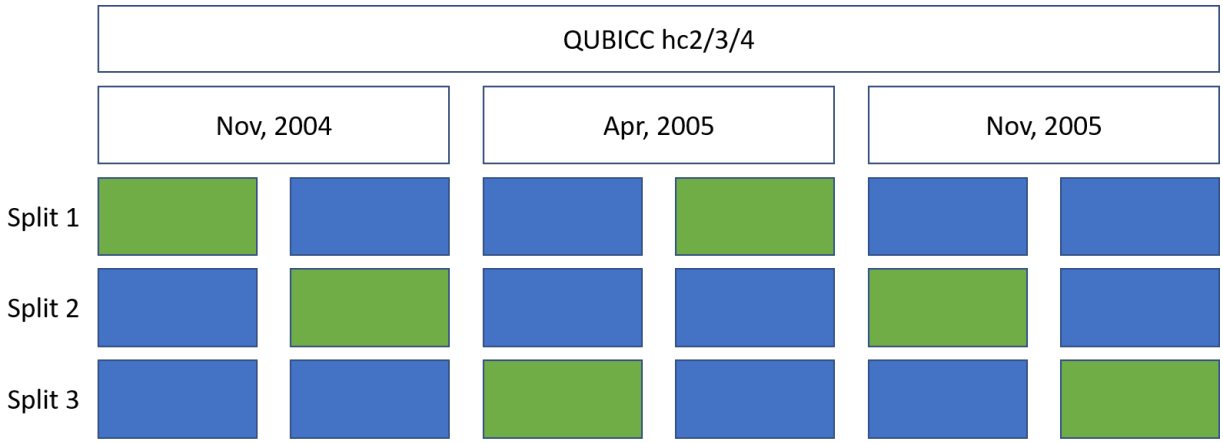


Figure S3. We split the R2B5 data using a three-fold temporally coherent cross-validation split. In each split, we train a network on the blue folds and validate it on the green folds. One fold covers approximately 15 days.

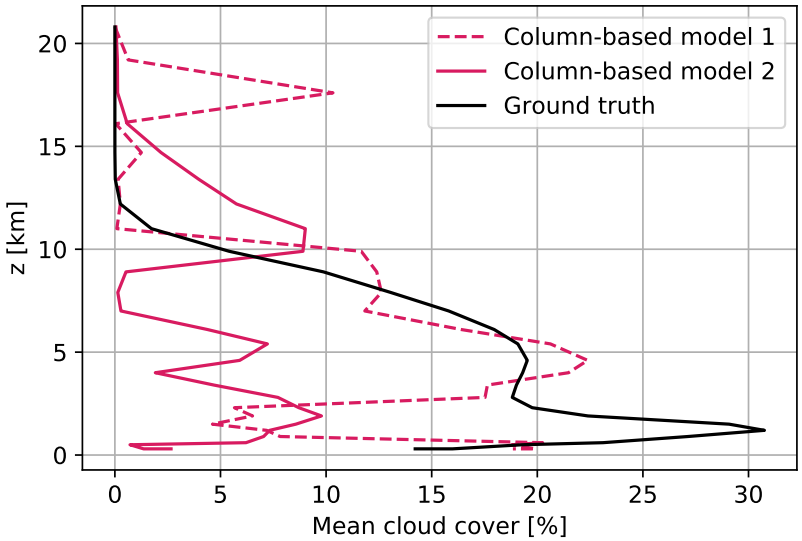
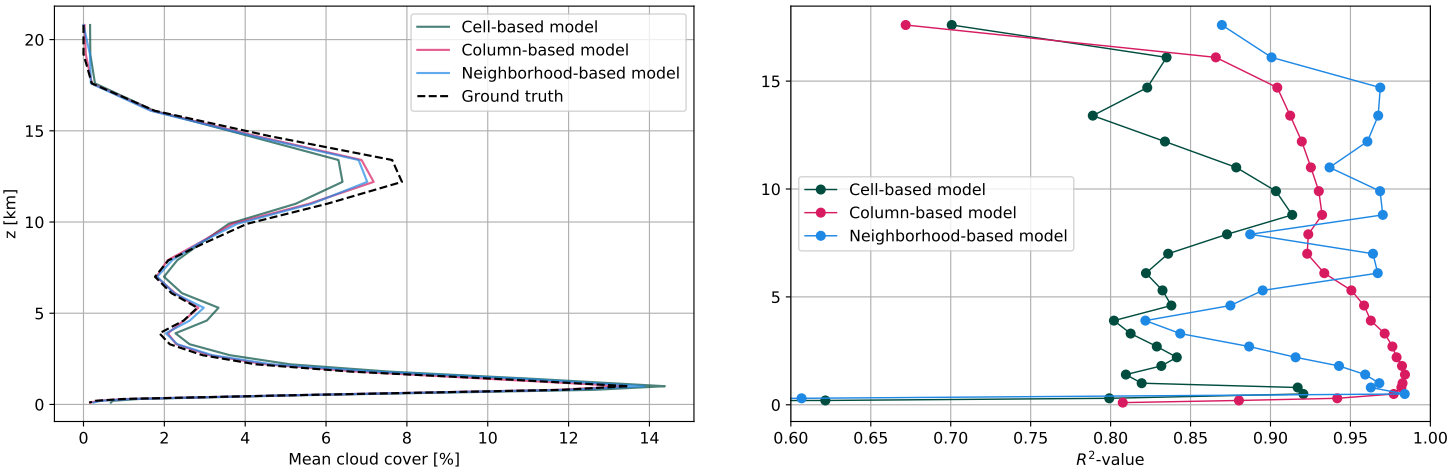


Figure S4. Two different column-based models trained on NARVAL R2B4 data evaluated on QUBICC R2B4 data over the Southern Ocean and Antarctica (< 60S). Models from the same type stop being consistent and deviate significantly from the ground truth.



(a) Cloud cover profiles

(b) Coefficients of determination (best value: 1)

Figure S5. The NNs trained on NARVAL R2B4 data evaluated on the coarse-grained and preprocessed NARVAL R2B5 data.

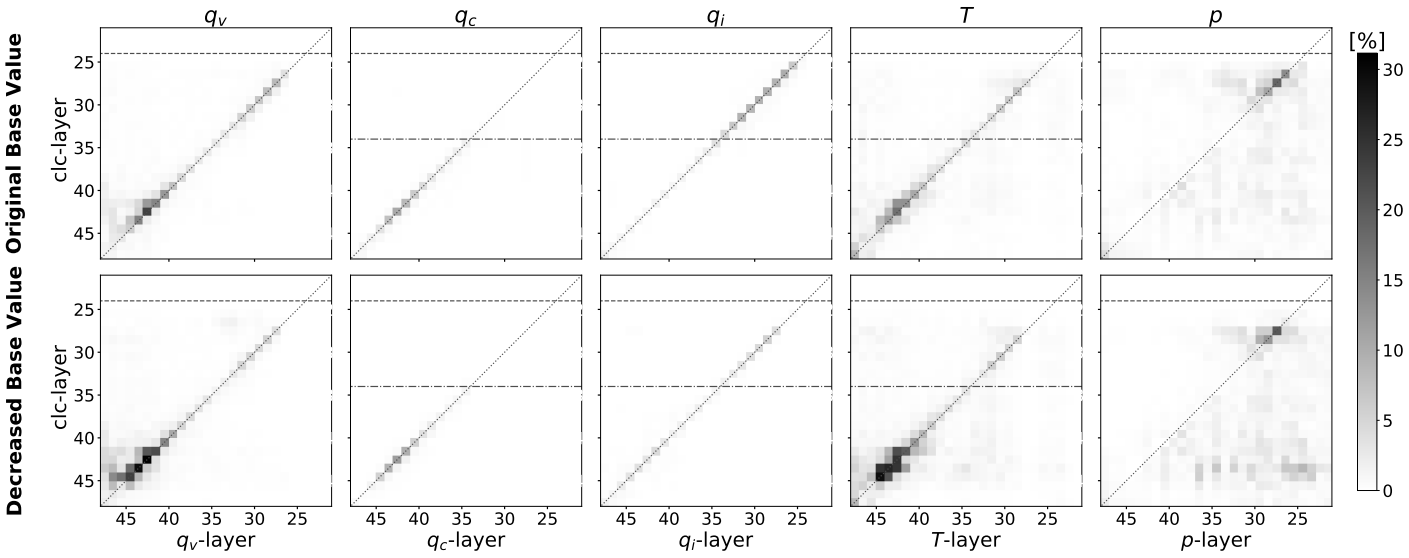


Figure S6. Average absolute SHAP values of the QUBICC R2B5 column-based model when applied to a sufficiently large subset of the NARVAL R2B5 data. By repeatedly drawing an appropriate training sample from the QUBICC training data we decrease its base values, aligning them closely with the cloud cover profile of the NARVAL R2B5 data. Tests with ten different seeds have shown the values from the lower row to be robust, with pixel values not differing absolutely by more than 1 or relatively by more than 20%. The input features that are not shown exhibit smaller absolute SHAP values ($z_g < 0.8\%$, $land/lake < 0.22\%$) everywhere and are thus omitted.