Research

Predicting atmospheric optical properties for radiative transfer computations using neural networks

Menno A. Veerman[1], Robert Pincus[2,3], Robin Stoffer[1], Caspar M. van Leeuwen[4], Damian Podareanu[4] and Chiel C. van Heerwaarden[1]

[1]Meteorology and Air Quality Group, Wageningen University and Research, Wageningen, The Netherlands
[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, CO, USA
[3]NOAA Physical Sciences Laboratory, Boulder, CO, USA
[4]SURFsara, Amsterdam, The Netherlands

MAV, 0000-0002-4869-3948; RP, 0000-0002-0016-3470; CMvL, 0000-0003-4407-6675; CCvH, 0000-0001-7202-3525

The radiative transfer equations are well known, but radiation parametrizations in atmospheric models are computationally expensive. A promising tool for accelerating parametrizations is the use of machine learning techniques. In this study, we develop a machine learning-based parametrization for the gaseous optical properties by training neural networks to emulate a modern radiation parametrization (RRTMGP). To minimize computational costs, we reduce the range of atmospheric conditions for which the neural networks are applicable and use machine-specific optimized BLAS functions to accelerate matrix computations. To generate training data, we use a set of randomly perturbed atmospheric profiles and calculate optical properties using RRTMGP. Predicted optical properties are highly accurate and the resulting radiative fluxes have average errors within $0.5\,\mathrm{W\,m^{-2}}$ compared to RRTMGP. Our neural network-based gas optics parametrization is up to four times faster than RRTMGP, depending on the size of the neural

networks. We further test the trade-off between speed and accuracy by training neural networks for the narrow range of atmospheric conditions of a single large-eddy simulation, so smaller and therefore faster networks can achieve a desired accuracy. We conclude that our machine learning-based parametrization can speed-up radiative transfer computations while retaining high accuracy.

This article is part of the theme issue 'Machine learning for weather and climate modelling'.

1. Introduction

Accurate calculations of radiative fluxes are key to capturing the coupling between radiation, the atmosphere and the surface. Unlike many parametrizations of subgrid processes, the equations governing radiative transfer are well known. This integration is often parametrized with correlated-k distribution methods [1–3] to drastically reduce the number of quadrature points. Even with this approximation radiative transfer schemes in weather and climate models remain a large computational burden. An important part of radiative transfer parametrizations is therefore to find approaches or approximations that further reduce the computational costs, for example by coarsening the spatial and temporal resolution of the radiative transfer computations [4,5] or random sampling in spectral space [6].

A promising and increasingly explored approach to accelerate or improve parametrizations is the use of machine learning techniques [7]. The application of machine learning to accelerate expensive radiative transfer computations was one of the first uses of machine learning in forward modelling in the atmospheric sciences [8], and a range of studies have used machine learning to predict vertical profiles of longwave [8–13] and shortwave [10,11,13] radiative fluxes in weather and climate models. These end-to-end approaches, i.e. predicting radiative fluxes by fully emulating a radiative transfer scheme, may result in speed-ups of more than one order of magnitude [8,9,11], with root mean squared errors (MSEs) of the heating rates within $0.2\,\mathrm{K\,d^{-1}}$ for shortwave and $0.5\,\mathrm{K\,d^{-1}}$ for longwave radiation. However, such an approach is inflexible with respect to changes in model configuration including the vertical discretization as the neural networks are trained for a fixed number of vertical layers [8–11]. Moreover, this approach does not respect the well-understood underlying physics.

In this study, we describe a machine learning approach for accelerating radiative transfer computations that respects the well-understood governing equations, using machine learning only to accelerate the data-driven aspects of the calculations. We exploit the fact that radiation calculations are composed for two distinct steps: the mapping of the physical and chemical state of the atmosphere to a problem in radiative transfer and the subsequent solution of that radiative transfer problem. The first step converts temperature, pressure and composition into atmospheric optical properties that determine how much radiation is emitted (Planck source function), absorbed or scattered (optical depth, single scattering albedo), and the direction of scattering (asymmetry parameter). Unlike the solution of the radiative transfer equation, the calculation of gas optics relies on empiricism and large amounts of tabulated data.

Here, we use machine learning to emulate the calculation of gaseous optical properties of a modern RRTM for general circulation model applications–parallel (RRTMGP) [14]. We use neural networks as a computationally efficient tool to replace the lookup-tables used in RRTMGP and train the networks for a narrowed range of atmospheric conditions, e.g. by neglecting variations of many trace gases and limiting the range of temperatures and pressures. Constraining the range and number of inputs allows further optimization of the neural networks, which helps to reduce the computational costs of our parametrization compared to RRTMGP. This approximation can be particularly suitable for limited-area models, such as large-eddy simulations (LESs), in which the range of values of thermodynamic variables is smaller than is required by general-purpose tools.

2. Training data generation

We train three sets of artificial neural networks to predict all optical properties:

— one set (NWP) is trained for a wide range of atmospheric conditions, roughly representing the variability expected in global numerical weather prediction, but with all gases except water vapour and ozone kept constant;

— two sets (Cabauw, RCEMIP) are trained for only the narrow range of atmospheric conditions within one LES simulation each, in order to estimate the performance gains of this LES-specific tuning.

Because the LES-tuned networks are trained for a narrow range of atmospheric conditions, they may contain substantially fewer layers and nodes than the NWP-tuned networks and can therefore be faster. This may provide a great speed-up when an LES simulation is repeated for large number of numerical experiments.

To generate the data required to train and validate the neural networks we use the new radiation package RTE+RRTMGP [14]. RRTMGP is a parametrization of gas optics, computing all optical properties from temperature (T), pressure (p) and the concentrations of a wide range of gases. Given these optical properties, RTE (radiative transfer for energetics) calculates the radiative fluxes throughout each column.

RRTMGP covers the spectral range of radiation relevant to atmospheric problems using a correlated k-distribution [3] with 14 shortwave (0.2–12 μm) bands, 16 longwave (3–1000 μm) bands and 16 g-points per band. We therefore need to predict 224 (14×16) values for the short wave optical properties and 256 (16×16) for the longwave optical depth. Additional to calculating the Planck source functions of each layer from the layer temperature, RRTMGP also calculates the Planck source functions at the layer interfaces from the interface temperatures and the wavelength to g-point mapping of the layers below and above each interface, separately. Each layer interface thus has two Planck source functions, representing the upward and downward emission. Since we aim to predict the same output as RRTMGP, we also need to predict the upward and downward emission at each layer interface, resulting in 768 ($3 \times 16 \times 16$) values per grid cell.

The NWP-tuned neural networks are trained with only T, p, H_2O and O_3 as input, which are time-varying three-dimensional variables in global weather prediction models. To create training data, we begin with the set of 100 atmospheric profiles of temperature, pressure and several gas concentrations from the Radiative Forcing Model Intercomparison Project (RFMIP) [15]. These 100 profiles were chosen to reproduce the global, annual mean radiative forcing between present-day and pre-industrial conditions [15], but do not represent the full diversity of atmospheric conditions on earth. RFMIP assumes that all gases except H_2O and O_3 are well mixed and we train only on present-day concentrations to reduce the degrees of freedom. Since larger amounts of training data is required to train neural networks, we generate extra data, spanning a slightly wider range of atmospheric conditions, by generating permutations of this set of 100 profiles with random perturbations in T, p, H_2O and O_3:

$$H_2O^i = (1 + c_1 r_1)H_2O^i, \ O_3^i = (1 + c_2 r_2)O_3, \ T^i = T^i + c_3 r_3, \ p^i = (p^{i+1} - p^{i-1})r_4 + p^{i-1},$$

where r_1, r_2 and r_3 are random numbers between -1 and 1, r_4 is a random number between 0.05 and 0.95, and i is the index of each layer. The choice of constants c_1, c_2 and c_3, which define the maximum width of the permutations, determines the range of atmospherics conditions the networks are trained for, but also affect the risk of generating unrealistic data. Here, we set constants c_1, c_2 and c_3 to 3/4, 3/4 and 5, respectively. Since the four random numbers are generated independently, the perturbations are uncorrelated. This gives a larger variation of different atmospheric conditions in the training data and reduces the risk of overfitting, but may result in combinations of T, p, H_2O and O_3 that are unlikely to occur in reality. Water vapour mixing ratio is constrained to be less than saturation. By generating r_4 between 0.05 and 0.95 and

not between 0 and 1, we aim to prevent unrealistically small layer depths. To obtain a wider range of surface conditions, we randomly choose the surface temperature between $T_0 - 10$ K and $T_0 + 10$ K, where T_0 is the temperature at the lowest pressure level. Note that the surface temperature is not used for training, since we compute the surface emission from the emission of the lowest model layer (equation (3.2)), but only to compute radiative fluxes from the predicted optical properties.

For the two LES-tuned sets of neural networks, which are trained for a narrow range of atmospheric conditions, we compute O_3 as a monotonic function of pressure following [16] with a lower boundary of 5×10^{-3} ppmv. Consequently, we train these networks using only T, p and H_2O. For the `Cabauw` networks, we run a 10 h LES simulation (07 UTC to 17 UTC) of a developing convective boundary layer over grassland near the Cabauw Experimental Site for Atmospheric Research (CESAR) in the Netherlands with shallow cumulus clouds forming in the afternoon (see [17,18] for a detailed description of the case). The simulation is performed using the Dutch atmospheric large-eddy simulation (DALES) [19], with domain size of $19.2 \times 19.2 \times 5.47$ km^3 and a resolution of $100 \times 100 \times 24$ m^3. For RCEMIP networks, we run a 100-day simulation with MicroHH [20] following the specification for cloud-resolving models (see `RCE_small300` in [16]) of the Radiative Convective Equilibrium Model Intercomparison Project (RCEMIP) [16]. This is a case with deep convection over a tropical ocean with an atmosphere in radiative convective equilibrium, meaning that radiative cooling is balanced by convective heating [16]. The simulation is performed with a domain size of $100 \times 100 \times 32$ km^3, a horizontal resolution of 1 km and 72 vertical levels. For each LES-tuned set, we then determine the minimum and maximum H_2O and T values of each vertical layer and subsequently generate random profiles of p, H_2O and T to cover the full parameter space of the corresponding simulation. We generate 1000 profiles for the `Cabauw` networks and 3000 profiles for the `RCEMIP` networks, because the RCEMIP simulation has a higher domain top and therefore spans larger range of atmospheric conditions. To deal with negative or unrealistically low water vapour concentrations in the simulations, we set a lower H_2O limit in these profiles of 16 ppmv and 5 ppmv for the `Cabauw` and `RCEMIP` simulation, respectively. Since the LES-tuned networks are trained for a narrow range of atmospheric conditions, they need to be retrained to be used for a LES simulation with a different range of conditions, which takes on the order of 1 hour on a single compute node.

For every combination of T, p, O_3 and H_2O (`NWP`) or T, p and H_2O (`Cabauw`, `RCEMIP`), we then calculate the optical properties at each g-point using RRTMGP. During both training and inference, we log-scale the optical depths, Planck source function, H_2O, (O_3,) and p, which improves the convergence of the neural networks because the range of these variables may span multiple orders of magnitude. For the log-scaling, we use a fast approximation of the natural logarithm to keep the computational effort feasible

$$\ln x = \lim_{n \to \infty} \left(\sqrt[n]{x} - 1 \right) n, \tag{2.1}$$

where we take $n = 16$. This approximation is consistent with the fast approximation of the exponential we use during inference (equation (3.2)).

3. Artificial neural networks

(a) Data preparation

Before training our neural networks we normalize all variables to a zero mean and unit variance, with different means and standard deviations for the upper ($p < 9948$ Pa) and lower ($p > 9948$ Pa) atmosphere per variable. A random 95% of the dataset is used to train the neural networks and 5% of the data is used for validation. In total, the training datasets for the `NWP`, `Cabauw` and `RCEMIP` networks consist of about 5.7×10^5 (lower atmosphere: 3.4×10^5, upper atmosphere: 2.3×10^5), 2.2×10^4 (lower atmosphere only) and 6.8×10^4 (lower atmosphere: 3.8×10^4, upper atmosphere: 3.0×10^4) data points, respectively.

Table 1. Properties of hidden layers, number of nodes, and number of weights (including biases) for all tested neural networks.

name	layer 1	layer 2	layer 3	no. of weights
1L–32	32	—	—	8608
1L–64	64	—	—	35 232
2L–32_32	32	32	—	18 272
2L–64_64	64	64	—	73 312
3L–32_64_128	32	64	128	116 928

(b) Network architecture and training

The neural networks are designed in and trained with TensorFlow [21], v.1.11/1.12. We need to predict four different optical properties: the single scattering albedo w_0 (shortwave only), the shortwave τ_{sw} and longwave τ_{lw} optical depth, and the Planck source function B. Because optical depth depends on layer depth, the optical depths are normalized by the layer thickness (Δp). We therefore predict $\tau_{sw}/\Delta p$ and $\tau_{lw}/\Delta p$, which are the proportional to the absorption coefficients (Pa^{-1}), so we do not have to account for layer thickness. For each optical property, we train two neural networks, for the upper atmosphere ($p < 9948$ Pa) and lower ($p > 9948$ Pa) atmosphere, a distinction that also made in RRTMGP. By training separate neural networks for each optical property, we can reduce the complexity of each network. Furthermore, this allows us to (re-)train the networks for the different optical properties independently. The neural networks predict the optical properties of each grid cell or layer from the values of T, p, (O$_3$) and H$_2$O of only that grid cell or layer. As such, each network has either 4 (NWP-tuned set) or 3 (LES-tuned sets) inputs and either 224 ($\tau_{sw}/\Delta p$, w_0), 256 ($\tau_{lw}/\Delta p$) or 768 (B) outputs.

All networks are feed-forward multi-layer perceptrons with densely connected layers. We use a leaky ReLu activation function [22] with a slope of 0.2 in all hidden layers and a linear activation function in the output layer. We train for 500 epochs (one epoch is one iteration over the entire training dataset). During training, we use the Adam optimizer [23] to optimize the weights, with a batch size of 128 and an initial learning rate of 0.01 that decays every 10 epochs. As the loss function, we use the MSE

$$\text{MSE} = \frac{1}{N_{\text{batch}}N_{gpt}} \sum_m^{N_{\text{batch}}} \sum_n^{N_{gpt}} (NN_{m,n} - RR_{m,n})^2, \quad (3.1)$$

where N_{batch} and N_{gpt} are the batch size and number of g-points, respectively. NN and RR are the optical properties predicted by the neural networks and calculated by RRTMGP, respectively.

Although several studies have already shown successful attempts to optimize neural network architecture for accuracy with machine learning [24,25], choosing the numbers of hidden layers and the number of nodes per layer is often still a matter of manual tuning. Wider and deeper networks are able to learn more complex functions, with the risk of overfitting, but are slower during both training and inference. We test several network sizes (table 1) to investigate the trade off between accuracy and performance for different network sizes. We also tested a network without hidden layers that performs only linear regression as a reference. However, the prediction errors of the linear networks were over 2 order of magnitude higher than the errors of the other networks (not shown), suggesting that a linear parametrization cannot capture the complex relationship between atmospheric conditions and optical properties.

(c) Implementation

We use the trained neural networks as a parametrization for the gas optics in the RTE+RRTMGP framework. We replace the RRTMGP gas optics and source function routine and pair the new parametrization with the RTE radiative transfer solver. The new parametrization gets as input

one or more columns of T, p, H_2O and O_3 (latter only for NWP) and outputs, for each layer of each column, the log-scaled and normalized optical properties for all 224, 256 or 768 g-points. The workflow for a neural network with two hidden layers of 64 nodes, is as follows.

(i) Initialize the weights matrices $W_1[N_1, N_{in}]$, $W_2[N_2, N_1]$, $W_3[N_{out}, N_2]$ and bias vectors $\beta_1[N_1]$, $\beta_2[N_2]$, $\beta_3[N_{out}]$ of the trained neural networks, where N_{in} is the number of inputs (3 or 4), N_{out} the number of outputs (224, 256 or 768), and N_1 and N_2 the number of nodes of the first and second hidden layer (both 64).

(ii) Create input matrix $I[N_{in}, N_{batch}]$ from the 4 input columns, where N_{batch} is the batch size, i.e. the number of grid cells computed simultaneously, where the number of grid cells is the number of atmospheric profiles multiplied by the amount of vertical layers per profile.

(iii) Apply log-scaling on p, H_2O and O_3 (NWP only) and normalize all input variables.

(iv) Calculate the first hidden layer ($L_1[N_1, N_{batch}]$):
    a. Calculate matrix product $L_1 = W_1 I$
    b. Add β_1 to each column of L_1
    c. Apply Leaky ReLu activation function

(v) Calculate the second hidden layer ($L_2[N_2, N_{batch}]$):
    a. Calculate matrix product $L_2 = W_2 L_1$
    b. Add β_2 to each column of L_2
    c. Apply Leaky ReLu activation function

(vi) Calculate output matrix ($O[N_{out}, N_{batch}]$):
    a. Calculate matrix product $O = W_3 L_2$
    b. Add β_3 to each column of O

(vii) Denormalize output matrix O, take the exponential of all values and multiply $\frac{\tau_{sw}}{\Delta p}$ $\frac{\tau_{lw}}{\Delta p}$ with layer thickness Δp.

The matrix products are the computationally most expensive parts of the neural network-solver. We therefore make use of the level 3 functions of the basic linear algebra subprograms (BLAS) library for which machine-specific optimized versions exist. In our implementation, we use the C-interface to the sgemm function of Intel's math kernel library [26] (MKL). For the exponentiation, we use a fast approximation consistent with the natural logarithm approximation (equation (2.1))

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n, \tag{3.2}$$

where we take $n = 16$. The exponential in step vii is omitted for the single scattering albedo w_0. The Planck source function at the surface B_{sfc} is calculated from the Planck source function of the lowest layer B_{lay0} with a function of the form

$$B_{sfc} = B_{lay0} \left(\frac{\alpha T_{sfc} - 1}{\alpha T_{lay0} - 1}\right)^\beta, \tag{3.3}$$

where T_{sfc} and T_{lay0} are the temperature of the surface and of the lowest layer, respectively, and coefficients α and β are fitted (per spectral band) from the lookup tables of RRTMGP.

4. Results and discussion

(a) NWP-tuned networks

(i) Prediction skill

Since our loss function is the MSE of the predicted optical properties with respect to the optical properties calculated by RRTMGP, a useful indication of the accuracy of the neural networks is
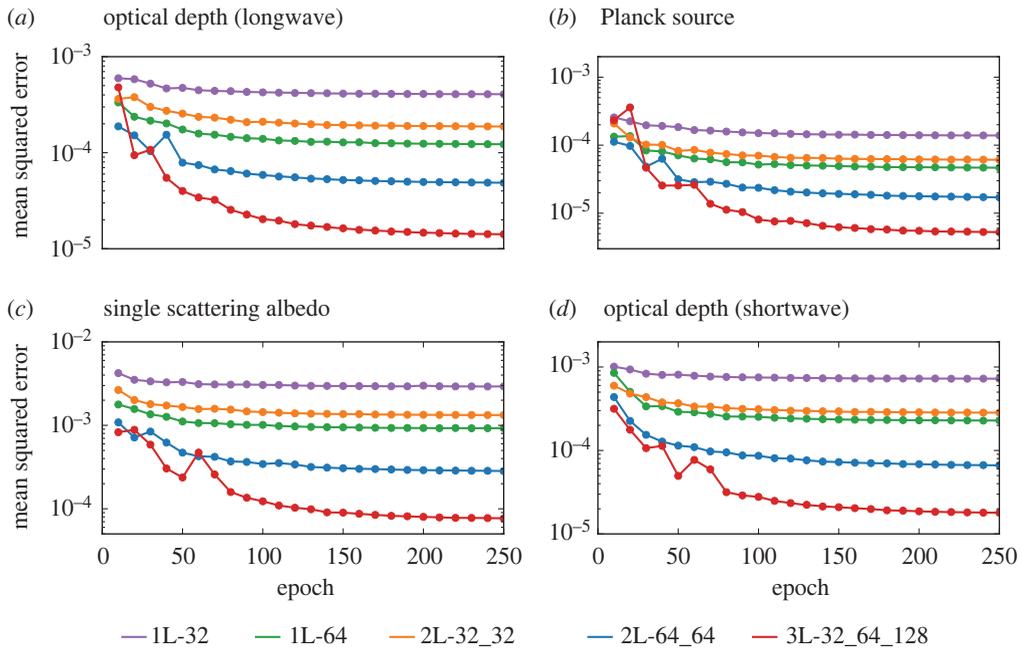
Figure 1. Mean squared errors, as function of the number of epochs, of the longwave optical depth (*a*), Planck source function (*b*), shortwave optical depth (*c*) and single scattering albedo (*d*) for the six different network sizes (table 1). The mean squared errors are based on the full validation dataset and the validation is performed every 10 epochs. (Online version in colour.)

the evolution of the MSE as training progresses (figure 1). We generally see that as the size of the networks increases, the number of epochs needed to reach convergence increases and the MSE at the end of the training decreases. The relatively large difference between the MSE's of the 2L-64_64 and 3L-32_64_128 networks suggests that we may still be able to strongly reduce the MSE by increasing the network size (at the cost of additional computational cost). The 1L_64 networks are slightly more accurate than the 2L_32_32 networks despite having the same number of hidden nodes: due to the larger number of output nodes (224/256), the 1L_64 networks have more connections between nodes (table 1). This means that more complex functions can be learned, but may also lead to higher computational costs.

To test whether the neural network predictions correlate well with RRTMGP, we generate a new set of 100 randomly perturbed profiles and calculate R-squared values (R^2) between the optical properties computed by the neural networks and by RRTMGP. The R^2-values are determined for each g-point separately and subsequently averaged over all 224 or 256 g-points ($\overline{R^2}$) to represent the overall performance. The networks typically have $\overline{R^2} > 0.9998$ for the optical depths and the Planck source function and $\overline{R^2} > 0.998$ for the single scattering albedo. These high correlations give us confidence that the neural networks are able to predict the optical properties with very high accuracy.

(ii) Radiative fluxes

The neural networks predict the optical properties very accurately, but for atmospheric modelling applications we require accurate radiative fluxes through the atmosphere and at the surface. To assess whether the accuracy of the neural networks is sufficient, we use our implementation of the neural networks in RTE+RRTMGP to calculate radiative fluxes based on the optical properties predicted by the neural networks.

The errors of the radiative fluxes based on the neural network-predicted optical properties generally decrease as the complexity of the neural networks increases (figure 2). Mean flux errors
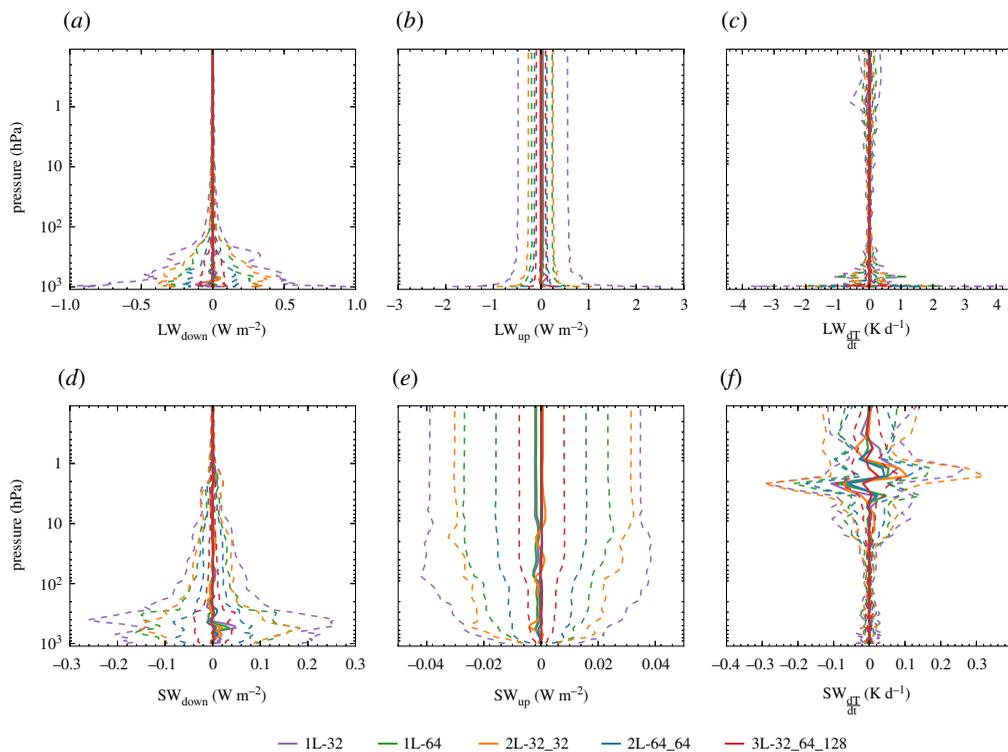
Figure 2. For all network sizes, vertical profiles of the errors of the radiative fluxes and heating rates based on the neural network-predicted optical properties with respect to the radiative fluxes based on the optical properties from RRTMGP. Shown are the mean (solid) and twice the standard deviation (dashed) of the error, for the downwelling (*a*) and upwelling (*b*) longwave radiation, longwave heating rates (*c*), the downwelling (*d*) and upwelling (*e*) shortwave radiation and shortwave heating rates (*f*). A zenith angle of 42° and an albedo of 0.07 is used for the shortwave fluxes. (Online version in colour.)

differ only slightly, whereas the spread of the flux errors is clearly smaller for the more complex networks. This indicates that reducing the size of the neural networks used to predict optical properties does not introduce a significant bias in the radiative fluxes, but results in larger flux errors for individual atmospheric profiles because smaller networks predict less accurate optical properties.

Our average errors in the downwelling surface fluxes and upwelling top of atmosphere fluxes with respect to RRTMGP (figure 2) are similar to or smaller than the average errors of RRTMGP with respect to the line-by-line radiative transfer model (LBLRTM) [27], as reported by [14] for the original set of RFMIP profiles. However, the accuracy of our neural network-approach may be lower for individual atmospheric profiles, especially with the smallest networks. Since the neural networks are trained against RRTMGP, the similar mean errors suggest that further increasing the predictive skill of the networks will not improve the radiative fluxes much on average, as the flux errors compared to the 'ground truth' will then be dominated by the errors of RRTMGP with respect to LBLRTM.

The errors of shortwave radiative heating rates, which are proportional to the divergence of the radiative fluxes, are within about $0.05\,\mathrm{K\,d^{-1}}$ in troposphere and most of the stratosphere, which is similar to the accuracy of RRTMGP with respect to LBLRTM [14]. However, the shortwave heating rates errors of the smallest networks are over $0.05\,\mathrm{K\,d^{-1}}$ near the top of the atmosphere, which is larger than the error of RRTMGP with respect to LBLRTM. The longwave radiative heating rates have errors within $1\,\mathrm{K\,d^{-1}}$ for most of the profile, which is higher than the error of RRTMGP with respect to LBLRTM, and the smaller networks give significantly larger errors at the surface. The
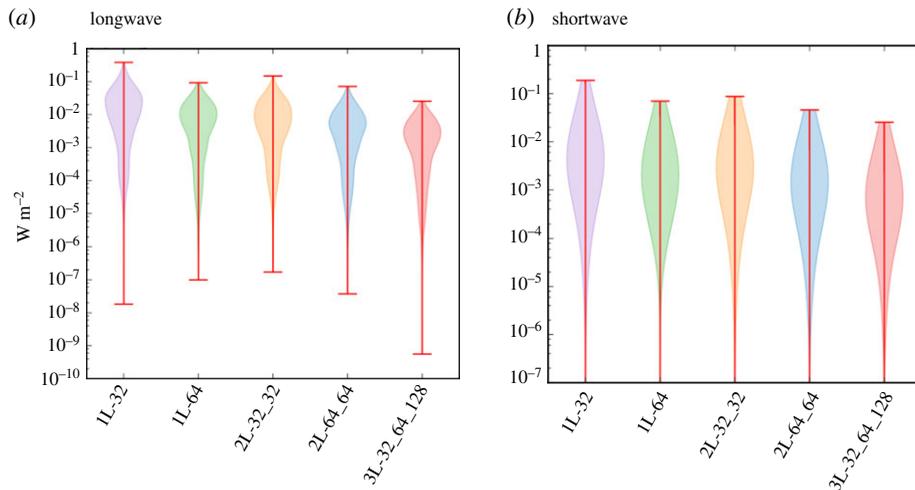
Figure 3. For both longwave (*a*) and shortwave radiation (*b*) and for all network sizes, violin plots of the absolute errors per spectral band of the downwelling surface fluxes with respect to RRTMGP. The radiative flux per spectral band is the integral of the radiative fluxes over the 16 *g*-points of each band. For each network size, the minimum errors in (*b*) are on the order of 10^{-13} and correspond to the spectral band with the shortest wavelength, which is almost completely absorbed in the stratosphere. (Online version in colour.)

largest longwave heating rates errors presumably occur mostly where the temperature gradients of the perturbed profiles are very large, in which case we find that the absolute heating rates of RRTMGP may be over $10^2 \, \mathrm{K} \, \mathrm{d}^{-1}$.

Additionally, we determine the absolute errors of the downwelling radiative surface fluxes per spectral band, which help to assess to what extent a single band contributes to the broadband flux errors. The accuracy of the radiative fluxes per band may also be of interest for predictions of UV-index [28] or photosynthetically activate radiation (PAR). The maximum errors of the flux per band range between $0.026 \, \mathrm{W} \, \mathrm{m}^{-2}$ and $0.39 \, \mathrm{W} \, \mathrm{m}^{-2}$ in the longwave spectrum and between $0.025 \, \mathrm{W} \, \mathrm{m}^{-2}$ and $0.18 \, \mathrm{W} \, \mathrm{m}^{-2}$ in the shortwave spectrum (figure 3), depending on network size. Using the neural network-predicted optical properties, we can thus calculate the radiative fluxes of each spectral band with errors below $0.5 \, \mathrm{W} \, \mathrm{m}^{-2}$.

(iii) Computational performance

We have demonstrated that neural networks can accurately reproduce optical properties determined from RRMTGP's lookup tables but the approach is only useful if it can accelerate radiation computations. To evaluate the difference in runtime between RRTMGP and the neural network-solver, we generate 11 additional sets of 100 profiles from the RFMIP profiles. The optical properties of these sets are evaluated sequentially and the runtimes of the last 10 sets are averaged. The runtime of the first set is neglected to allow some spin-up time, mainly for the initialization of BLAS. Since the neural networks run in single precision, we also run RRTMGP in single precision for this benchmark to have a fairer comparison. However, it must be noted that the shortwave solver of RTE currently only works in double precision. The benchmarks are performed on a single core (compute node: 2×12-core Intel Xeon E5-2690 v3, 2.6 GHz, 64 GB memory)

The neural network-based gas optics parametrization is up to about four times faster than RRTMGP, depending on size of the networks (figure 4). The speed-up of the neural network-solver generally decreases for increasing network complexity. This is as expected, because the number of matrix multiplications scales with the number of layers and because the size of the matrices, and thus the computational effort required for the matrix multiplications, scales with the number of nodes per layer. The choice of a network size thus depends on the acceptable
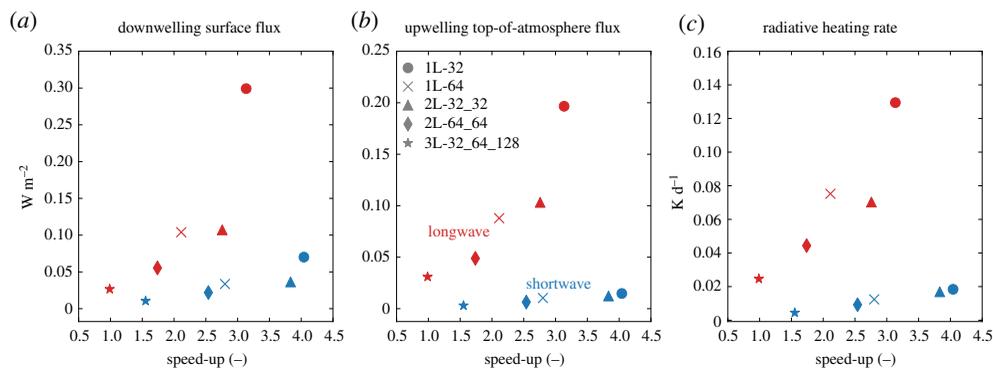
Figure 4. For all network sizes, the speed-up of the neural networks-solver against the mean absolute errors of the radiative heating rates (*a*), the upwelling top-of-atmosphere flux (*b*) and the downwelling surface flux (*c*), for shortwave (blue) and longwave (red) radiation. (Online version in colour.)

accuracy for a particular range of atmospheric conditions. If the neural networks would be trained for the full range of all 19 gases specified in RFMIP, or for the even larger set of gases treated by RRTMGP, they would have much more input nodes and likely also need more nodes in the hidden layer, which would reduce the speed-up achieved by our neural network approach.

(b) Large-eddy simulations-tuned networks

The NWP-tuned networks are trained for a wide range of atmospheric conditions, but in typical LES simulations we may expect only a narrow range of conditions. For this reason, we investigate whether smaller and therefore faster network suffice when the networks are trained for the narrow range of atmospheric conditions in a single LES simulation. The predicted optical properties of LES-tuned neural networks (Cabauw, RCEMIP) are also very accurate. To further test the LES-tuned neural networks, we sample 100 random atmospheric profiles each from the Cabauw and RCEMIP simulations and calculate profiles of optical properties with the Cabauw and RCEMIP sets of neural networks, respectively. For the Cabauw networks, we generally observe an improvement in the accuracy of the radiative heating rates, the downwelling surface fluxes and the upwelling top-of-atmosphere fluxes (figure 5) due to the LES tuning. This improvement is especially large in the shortwave spectrum, where for some networks sizes the MSE of the surface and top-of-atmosphere fluxes are over an order of magnitude lower with the Cabauw networks than with the NWP networks. The lower MSE of the Cabauw neural networks show that with LES tuning, we can use relatively small networks (e.g. 1L-32) to achieve similar or even higher accuracy than the more complex networks (2L-64_64, 3L-32_64_128) of the NWP set, which results in a larger speed-up compared to RRTMGP (figure 3). With the RCEMIP neural networks, we also observe a general increase in the accuracy of the shortwave and longwave fluxes and heating rates, (figure 5), but the improvements are often less than with the Cabauw networks. It should be noted that the improvement in accuracy achieved by LES-tuning is not fully consistent. Especially with the smaller networks, the accuracy of the LES-networks may be slightly lower than the accuracy of the NWP-tuned networks.

The mean absolute errors of the NWP networks on the profiles of the Cabauw (figure 5) and RCEMIP (figure 5) simulations are frequently larger than the errors of the NWP networks on the RFMIP-based profiles (figure 2). This might be an indication that not all atmospheric conditions occurring in the Cabauw and RCEMIP simulations are well-enough represented in the training data based on the RFMIP profiles, but the lower errors of NWP networks on the RFMIP-based profiles may also be a sign of overfitting due to insufficiently independent training and testing data. Nevertheless, given that the mean absolute errors are well within $0.5\,\mathrm{W\,m^{-2}}$ we are still
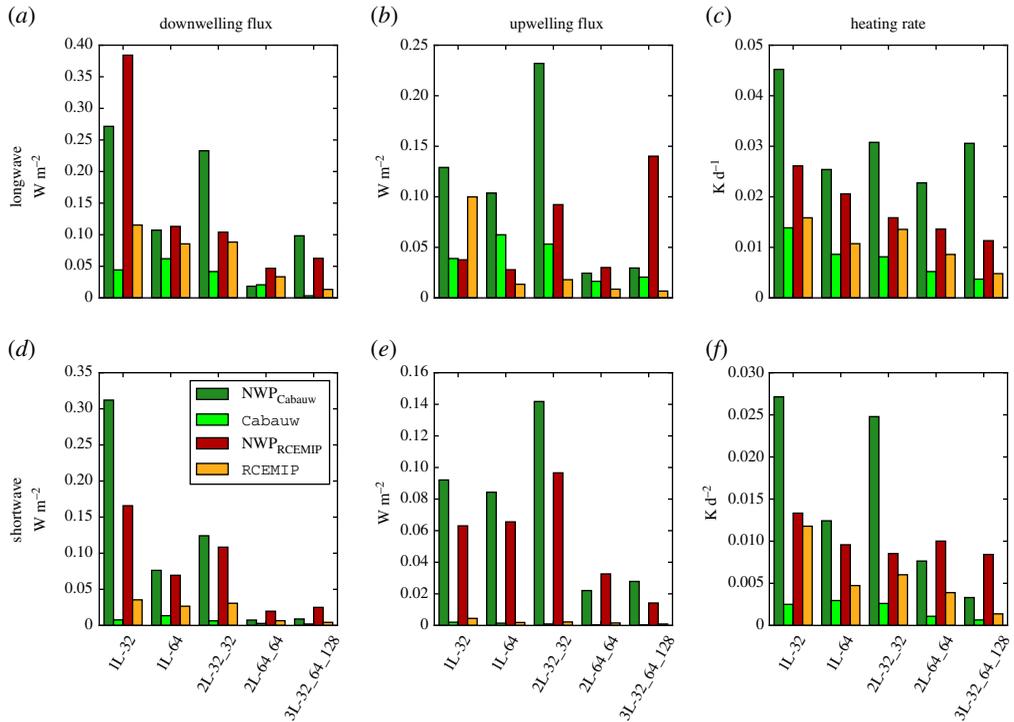
Figure 5. For all network sizes of the `NWP`, `Cabauw` and `RCEMIP` networks, the mean absolute errors with respect to RRTMGP of the radiative heating rates (a,d), upwelling radiative fluxes at the top of atmosphere (b,e) and downwelling radiative fluxes at the surface (c,f) for the longwave (a–c) and shortwave (d–f) spectrum. Radiative fluxes and heating rates are based on 100 random profiles of the `Cabauw` simulation (NWP$_{Cabauw}$, `Cabauw`) or the `RCEMIP` simulation (NWP$_{RCEMIP}$, `RCEMIP`). (Online version in colour.)

confident that the `NWP` neural networks can be accurately used on a relatively wide range of atmospheric conditions.

5. Conclusion

We developed a new parametrization for the gas optics by training multiple neural networks to emulate the gaseous optical properties calculations of RRTMGP [14]. The neural networks are able to predict the optical properties with high accuracy and errors of the radiative fluxes based on the predicted optical properties are mostly within $2\,W\,m^{-2}$. The resulting radiative heating rates are also accurate, especially in the shortwave spectrum. Radiative heating rate errors may be over $4\,K\,d^{-1}$ in the longwave spectrum, mainly near the surface, but we expect these errors to decrease rapidly after adjustment of the surface and air temperatures.

The neural network-based gas optics parametrization tested in this study is up to about four times faster than RRTMGP, depending on network size. The larger networks achieve lower speed-ups than the small networks, but result in more accurate radiative fluxes and heating rates, clearly showing a trade-off between accuracy and computational speed. To further investigate this trade-off, we trained two additional sets of neural networks; each is tuned for the narrow range of conditions of a single LES simulation (`Cabauw`, `RCEMIP`). In general, these LES-tuned networks are more accurate on profiles of their respective simulations than the `NWP` networks, especially for shortwave radiation. This indicates that with LES tuning, smaller and therefore faster neural networks suffice to achieve a desired accuracy.

Given that RRTMGP uses linear interpolation from look-up tables to compute optical properties [14], the computational efficiency of our neural network-based parametrization may be surprising. We attribute the speed-ups achieved by our parametrization to a large extent to the case-specific tuning, i.e. considering only a few gases or greatly limiting the range of atmospheric conditions (`Cabauw` and `RCEMIP` only), which reduces the problem size for which the neural networks have to be trained. Furthermore, the matrix computations required to solve the neural networks allow the use of machine-specific optimized BLAS libraries and reduces the memory use and access at the expense of floating-point operations.

The speed-ups we achieve are less than those achieved by end-to-end approaches that emulate full radiative transfer parametrizations [8–11], which may be up to 80 times faster than the original radiative transfer schemes. An advantage of our machine-learning approach is that it still respects the governing radiative transfer equations, at the cost of having to perform the spectral integration by predicting optical properties and calculating fluxes for all g-points. A promising future approach would be the application of machine learning to optimize the spectral integration. With such a machine learning approach the radiative transfer equations will still be solved, while the number of quadrature points may be reduced, e.g. by training neural networks to predict broadband fluxes from a small set of g-point. This would speed-up both the computations of both optical properties and the resulting radiative fluxes. The benefit of case-specific neural network-training also raises the question to what extent RRMTGP can be accelerated by reducing the number of input gases, which may result in smaller lookup tables and fewer computations. This was not investigated in this study, but the use of case-specific lookup tables in RRTMGP would be interesting for further studies.

References

1. Goody R, West R, Chen L, Crisp D. 1989 The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *J. Quant. Spectrosc. Radiat. Transfer* **42**, 539–550. (doi:10.1016/0022-4073(89)90044-7)
2. Lacis AA, Oinas V. 1991 A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. *J. Geophys. Res. Atmos.* **96**, 9027–9063. (doi:10.1029/90JD01945)
3. Fu Q, Liou KN. 1992 On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *J. Atmos. Sci.* **49**, 2139–2156. (doi:10.1175/1520-0469(1992)049<2139:OTCDMF>2.0.CO;2)
4. 2000 On the effects of temporal and spatial sampling of radiation fields on the ECMWF forecasts and analyses. *Mon. Weather Rev.* **128**, 876–887. (doi:10.1175/1520-0493(2000)128<0876:OTEOTT>2.0.CO;2)
5. Hogan RJ, Bozzo A. 2018 A flexible and efficient radiation scheme for the ECMWF model. *J. Adv. Model. Earth Syst.* **10**, 1990–2008. (doi:10.1029/2018MS001364)

6. Pincus R, Stevens B. 2009 Monte Carlo spectral integration: a consistent approximation for radiative transfer in large eddy simulations. *J. Adv. Model. Earth Syst.* **1**, 1–9. (doi:10.3894/JAMES.2009.1.1)

7. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat . 2019 Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204. (doi:10.1038/s41586-019-0912-1)

8. Chevallier F, Chéruy F, Scott NA, Chédin A. 1998 A neural network approach for a fast and accurate computation of a longwave radiative budget. *J. Appl. Meteorol.* **37**, 1385–1397. (doi:10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2)

9. Krasnopolsky VM, Fox-Rabinovitz MS, Chalikov DV. 2005 New approach to calculation of atmospheric model physics: accurate and fast neural network emulation of longwave radiation in a climate model. *Mon. Weather Rev.* **133**, 1370–1383. (doi:10.1175/MWR2923.1)

10. Krasnopolsky VM, Fox-Rabinovitz MS. 2006 Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Netw.* **19**, 122–134. (doi:10.1016/j.neunet.2006.01.002)

11. Krasnopolsky VM, Fox-Rabinovitz MS, Hou YT, Lord SJ, Belochitski AA. 2010 Accurate and fast neural network emulations of model radiation for the NCEP coupled climate forecast system: climate simulations and seasonal predictions. *Mon. Weather Rev.* **138**, 1822–1842. (doi:10.1175/2009MWR3149.1)

12. Belochitski A, Binev P, DeVore R, Fox-Rabinovitz M, Krasnopolsky V, Lamby P. 2011 Tree approximation of the long wave radiation parameterization in the NCAR cam global climate model. *J. Comput. Appl. Math.* **236**, 447–460. (doi:10.1016/j.cam.2011.07.013)

13. Pal A, Mahajan S, Norman MR. 2019 Using deep neural networks as cost-effective surrogate models for super-parameterized E3SM radiative transfer. *Geophys. Res. Lett.* **46**, 6069–6079. (doi:doi.org/10.1029/2018GL081646)

14. Pincus R, Mlawer EJ, Delamere JS. 2019 Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *J. Adv. Model. Earth Syst.* **11**, 3074–3089. (doi:10.1029/2019MS001621)

15. Pincus R, Forster PM, Stevens B. 2016 The radiative forcing model intercomparison project (RFMIP): experimental protocol for CMIP6. *Geoscientific Model Dev.* **9**, 3447–3460. (doi:10.5194/gmd-9-3447-2016)

16. Wing AA, Reed KA, Satoh M, Stevens B, Bony S, Ohno T. 2018 Radiative–convective equilibrium model intercomparison project. *Geoscientific Model Dev.* **11**, 793–813. (doi:10.5194/gmd-11-793-2018)

17. Vilà-Guerau de Arellano J, Ouwersloot HG, Baldocchi D, Jacobs CMJ. 2014 Shallow cumulus rooted in photosynthesis. *Geophys. Res. Lett.* **41**, 1796–1802. (doi:10.1002/2014GL059279)

18. Pedruzo-Bagazgoitia X, Ouwersloot HG, Sikma M, van Heerwaarden CC, Jacobs CMJ, Vilà-Guerau de Arellano J. 2017 Direct and diffuse radiation in the shallow-cumulus vegetation system: enhanced and decreased evapotranspiration regimes. *J. Hydrometeorol.* **18**, 1731–1748. (doi:10.1175/JHM-D-16-0279.1)

19. Heus T, van Heerwaarden CC, Jonker HJJ, Pier Siebesma A, Axelsen S, Geoffroy O, Moene AF, Pino D, de Roode SR, Vilà-Guerau de Arellano J. 2010 Formulation of the Dutch atmospheric large-eddy simulation (dales) and overview of its applications. *Geoscientific Model Dev.* **3**, 415–444. (doi:10.5194/gmd-3-415-2010)

20. van Heerwaarden C, Van Stratum BJ, Heus T, Gibbs JA, Fedorovich E, Mellado JP. 2017 Microhh 1.0: a computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows. *Geoscientific Model Dev.* **10**, 3145–3165. (doi:10.5194/gmd-10-3145-2017)

21. Abadi M *et al.* 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems. *CoRR* abs/1603.04467.

22. Maas AL, Hannun AY, Ng AY. 2013 Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML*, **30**, 3.

23. Kingma DP, Ba J. 2014 Adam: a method for stochastic optimization. (http://arxiv.org/abs/1412.6980).

24. Baker B, Gupta O, Naik N, Raskar R. 2016 Designing neural network architectures using reinforcement learning. *CoRR* abs/1611.02167.

25. Zoph B, Le QV. 2016 Neural architecture search with reinforcement learning. *CoRR* abs/1611.01578.
26. Intel. Intel Math Kernel Library. https://software.intel.com/en-us/intel-mkl (accessed 24 June 2020).
27. Clough S, Shephard M, Mlawer E, Delamere J, Iacono M, Cady-Pereira K, Boukabara S, Brown P. 2005 Atmospheric radiative transfer modeling: a summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer* **91**, 233–244. (doi:10.1016/j.jqsrt.2004.05.058)
28. WHO. 2002 *Global solar UV index: a practical guide*. Geneva, Switzerland: World Health Organization.

14

royalsocietypublishing.org/journal/rsta *Phil. Trans. R. Soc. A* **379**: 20200095