Posted on 25 May 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/essoar.168500302.25137268/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Sparse, Empirically Optimized Quadrature for Radiative Calculations

Paulina Czarnecki[1], Robert Pincus[1], and Lorenzo M Polvani[1]

[1]Columbia University

May 25, 2023

Abstract

Broadband (spectrally-integrated) radiation calculations are dominated by the expense of spectral integration, and many applications require fast parameterizations for computing radiative flux. Here we describe a novel approach using a linear weighted sum of monochromatic calculations at a small set of optimally-chosen frequencies. The empirically-optimized quadrature method is used to compute atmospheric boundary fluxes, net flux profiles throughout the atmosphere, heating rate profiles, and top-of-the-atmosphere forcing by CO2. We evaluate the method against two modern correlated k-distributions and find that we can achieve comparable errors with 32 spectral points. We also examine the effect of minimizing different cost functions, and find that in order to accurately represent heating rates and CO2 forcing, these quantities must be included in the cost function.

Sparse, Empirically Optimized Quadrature for Radiative Calculations

Paulina Czarnecki[1], Robert Pincus[2], Lorenzo Polvani[1,2,3]

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA
[2]Lamont-Doherty Earth Observatory, Palisades, NY, USA
[3]Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA

Key Points:

- Spectrally integrated radiative quantities can be estimated from a small weighted set of monochromatic calculations.
- Tradeoffs between computational cost and accuracy are commensurate with correlated k-distributions.
- Optimization priorities can be tailored to a specific applications.

Abstract

Broadband (spectrally-integrated) radiation calculations are dominated by the expense of spectral integration, and many applications require fast parameterizations for computing radiative flux. Here we describe a novel approach using a linear weighted sum of monochromatic calculations at a small set of optimally-chosen frequencies. The empirically-optimized quadrature method is used to compute atmospheric boundary fluxes, net flux profiles throughout the atmosphere, heating rate profiles, and top-of-the-atmosphere forcing by CO_2. We evaluate the method against two modern correlated k-distributions and find that we can achieve comparable errors with 32 spectral points. We also examine the effect of minimizing different cost functions, and find that in order to accurately represent heating rates and CO_2 forcing, these quantities must be included in the cost function.

Plain Language Summary

Quantifying the way radiation flows through the atmosphere is computationally expensive, and most applications require fast approximation. In this paper, we develop an empirically-optimized quadrature method that can compute quantities of interest in climate science and weather prediction such as radiative flux throughout the atmosphere, heating rates, and forcing by carbon dioxide. The method uses a simple optimization algorithm, combined with a linear model, to achieve accuracy and computational cost of similar order of magnitude to the modern correlated k-distributions. Depending on the focus of the application, the algorithm can easily be adapted to prioritize desired quantities.

1 Approximations for Spectral Integrals in Radiative Transfer Calculations

Radiation is the mechanism by which the Earth system achieves energy balance; radiative heating and cooling within the atmosphere also play a crucial role in setting the temperature profile of the atmosphere (Manabe & Wetherald, 1967). The strength of the interactions between radiation and the atmosphere depends strongly on frequency because the optical properties of gases, fundamentally determined by their molecular structure along with local temperature and pressure, vary rapidly in frequency by many or-

ders of magnitude. (The optical properties of condensed materials such as clouds and aerosols vary much more slowly with frequency and are not considered here.)

The optical properties of the gases in the atmosphere can be determined at given frequencies based on detailed spectroscopic information describing each absorption line for each species (e.g., HITRAN, Gordon et al., 2022). The spectrally-integrated (broadband) exchange of radiation can be determined from this information to near-arbitrary accuracy by direct integration at very high spectral resolution. Results from line-by-line models taking this approach (along with high angular resolution) may be considered reference calculations (Pincus et al., 2020).

Highly detailed calculations are, however, impractical in many contexts, such as when radiation must be computed as a diabatic forcing for a model of the atmosphere. In these contexts the radiative transfer integral is approximated using parameterizations, which reduce both spectral and angular detail. The state-of-the-art method for reducing spectral resolution is the correlated k-distribution (CKD) algorithm (Fu & Liou, 1992; Lacis & Oinas, 1991; Goody et al., 1989), which sorts frequencies according to their absorption coefficient k and determines the average at some set of quadrature points. The sorting and averaging creates a much smoother integral that can be evaluated with just tens or hundreds of quadrature points, as compared to the millions of points required by line-by-line models. These parameterizations seek to minimize the number of these quadrature points, called g-points, to limit computational cost (Hogan & Matricardi, 2022).

There is, however, no theory underlying k-distributions, so developers must make a range of somewhat arbitrary choices (Hogan & Matricardi, 2020). Moreover, the common experience is that the absorption coefficients determined with any algorithm need to be tuned to match reference calculations (Hogan & Matricardi, 2022; Pincus et al., 2019), making the link between the reference and approximate calculations somewhat obscure.

Here we describe an approximation to spectral integration in the longwave consisting of a weighted sum of monochromatic fluxes computed at a small set of discrete wavenumbers. We optimize both the members of this set and the linear weights to find the best estimate of the broadband integrated flux. The idea builds on approaches for computing narrow-band radiances originally developed for remote sensing applications (Buehler et al., 2010; Moncet et al., 2008) to sample the spectrum sparsely rather than averag-

ing absorption coefficients. Our work is in the same spirit as de Mourgues et al. (2023), though we take a more general approach and treat a wider range of applications including heating rates, boundary fluxes, and CO_2 forcing. We explore the relationship between the quantities targeted in the optimization and individual error profiles, test the effect of the vertical discretization of the training dataset, and compare the performance of the empirically-optimized quadrature method to state-of-the-art correlated-k methods. We find that weighted, sampled monochomatic calculations can predict fluxes, heating rates, and CO_2 forcing with error-cost relationships commensurate with modern correlated-k models, while being flexible to the needs of a given application.

2 Broadband Radiative Integral Calculation

The net broadband radiative flux at height z in the atmosphere, $F(z)$, is given by integrating monochromatic fluxes across all wavenumbers:

$$F(z) = \int F_\nu(z)\, d\nu \approx \sum_i F_{\nu_i}(z)\Delta\nu_i,$$

where ν denotes wavenumber and $F_\nu(z)$ is monochromatic net flux. We develop the method for the longwave in clear skies, so we disregard scattering.

Rather than calculating the line-by-line integral directly, we seek to compute the broadband net flux from a small subset of representative wavenumbers. This breaks the problem into two parts:

- Optimally select the representative wavenumbers, and
- Compute the net integrated flux from this subset.

Inspired by approaches originally developed for remote sensing applications (Buehler et al., 2010; Moncet et al., 2008), we select the representative subset via a combinatorial optimization algorithm called simulated annealing (Kirkpatrick et al., 1983), and calculate the net flux by performing a weighted sum of monochromatic fluxes found at the representative wavenumbers. In other words, we rewrite our problem as:

$$F(z) \approx \sum_{\nu \in S} w_\nu F_\nu(z), \tag{1}$$

where w_ν is the weight corresponding to wavenumber $\nu \in S$, and $F_\nu(z)$ is the monochromatic flux found at wavenumber ν at height z. Though this is similar to the formula-

tion of k-distributions (for example, Eqn. 3 in Lacis and Oinas (1991)), here we are sampling monochromatic fluxes rather than averaging absorption coefficients.

2.1 Line-by-Line Spectra and Correlated-K Data

In order to spectrally sample $F_{\nu \in S}$ and choose w_ν, we need accurate, finely-resolved line-by-line spectra at arbitrary frequencies, as well as corresponding integrated fluxes and heating rates throughout the atmosphere. Atmospheric profiles used for training must be independent and provide a representative sample of conditions for a given applications.

We use the set of atmospheric conditions developed for the Correlated K-Distribution Model Intercomparison Project (CKDMIP; see Hogan & Matricardi, 2020). The CKDMIP profiles are chosen from reanalysis data to represent the mean state as well as minimum and maximum extremes in water vapor concentration, ozone concentration, and temperature for the present day climate. We use software provided by CKDMIP to generate 100 independent present-day, clear-sky reference flux profiles, half for training and half for testing, integrated in 4 hemispheric angles. Heating rates are computed from fluxes using a simple finite difference. Finally, we perturb the CO_2 concentration of the 100 profiles to generate reference CO_2 forcing values for the Last Glacial Maximum (180 ppm), preindustrial (280 ppm), present day ($\sim$415 ppm), 2x PI (560 ppm), 4x PI (1120 ppm), 8x PI (2240 ppm) (Hogan & Matricardi, 2020), and 16x PI (4480 ppm) for extreme paleoclimate scenarios (Bice & Norris, 2002).

2.2 Predicting Net Flux from a Subset of Spectral Points

Given a subset of spectral points S of predetermined size $|S|$, we must obtain a set of weights such that the weighted sum of fluxes at those points predicts the broadband net flux across the spectrum. The weights w_ν are found by minimizing a linear cost function subject to several constraints.

First, all $w_\nu \geq 0$, since a negative weight would indicate redundant information (Buehler et al., 2010; Moncet et al., 2008). Also, the weights are normalized so that they sum to the length of the spectral domain, a general property of integration:

$$\sum w_\nu = \sum_i \Delta\nu_i \approx 3260 \text{ cm}^{-1},$$

as we focus on the longwave. This is achieved by formulating the cost function as a linear program and solving using the convex optimization package `cvxpy` in Python (Agrawal et al., 2018; Diamond & Boyd, 2016).

2.3 Wavenumber Subset Optimization Via Simulated Annealing

Next, we must identify an optimal subset at which to compute the predicted flux. We use simulated annealing, a stochastic optimization algorithm that can combinatorially evaluate different subsets in order to find a global minimum in a given cost function (Kirkpatrick et al., 1983). The algorithm is guaranteed to converge asymptotically to a global minimum in cost function with probability 1 (Dekkers & Aarts, 1991), and can do so at much lower computational cost than a brute force approach for a large state space.

Buehler et al. (2010) provides a useful tutorial on algorithm implementation. Annealing temperature is initialized such that 99% of moves will be accepted (Buehler et al., 2010). The optimization is organized in blocks of moves that the cost function value is averaged over (Buehler et al., 2010). Each block consists of 100 moves that must all be successful; otherwise, the algorithm is considered to have converged to a minimum and is terminated. Finally, when a block has a lower average cost function value than the previous block, the annealing temperature is decreased by a factor of 0.9 (Buehler et al., 2010).

Because the optimization process is stochastic, we perform 16 independent optimizations for each subset size and plot the standard deviation away from the mean error in our results. We refer to a given set of wavenumbers and corresponding weights as a model.

2.4 Cost Function Formulation

As we are interested in optimizing multiple radiative quantities – fluxes, heating rates, and top-of-the-atmosphere CO_2 forcing – we take inspiration from multi-objective optimization in the formulation of our cost function (Yang, 2014). Given multiple optimization targets F, H, and $\mathcal{F}$, the simplest way to minimize their respective cost functions c simultaneously is to minimize their weighted sum:

$$C = f_1\mathbf{c}(F) + f_2\mathbf{c}(H) + f_3\mathbf{c}(\mathcal{F}).$$

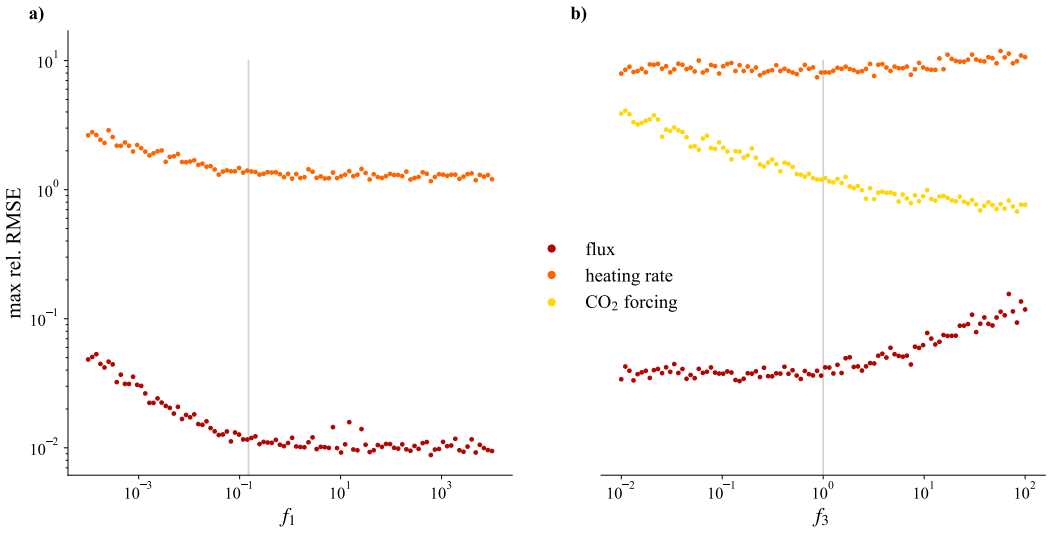

Figure 1. The relative root mean squared error (RMSE) of each radiative term of interest is plotted as weighting parameters vary. In **a)**, the weight on the heating rate term $f_2 = 1$ is held fixed while f_1 is varied. In **b)**, $f_1 = 0.15$ and $f_2 = 1$ are held fixed while f_3 is varied. The grey lines show the value of the parameter that is used throughout the study.

Choosing the values of the weights f_i is considered a subjective task with many valid solutions (Yang, 2014). In our case, the approach is as follows: first, we fix $f_2 = 1$ and $f_3 = 0$ as we vary f_1, to focus on the tradeoff between heating rates and fluxes. For each value of f_1, we fit linear weights to 100 random subsets of size 32 at every fifth vertical level for each of the 50 training atmospheres. The maximum relative (percent) error across vertical levels is recorded for each of the 100 iterations. The mean of these 100 iterations is plotted, where any random subset that produces an ill-conditioned cost function is ignored. The results of this experiment are shown in Fig. 1a). The grey line indicates the parameter value $f_1 = 0.15$ that is chosen, as it occurs where heating rate errors level off.

In Fig. 1b), the same procedure is repeated while holding $f_1 = 0.15$ and $f_2 = 1$ fixed and varying f_3. We choose parameter $f_3 = 1$ in order to minimize increase in heating rate and flux errors.

3 Accuracy in Flux, Heating Rate, and CO_2 Forcing Prediction

3.1 Boundary Fluxes

We illustrate the method by training the algorithm only on the outgoing atmospheric boundary fluxes: the upwelling flux at the top of the atmosphere and the downwelling flux at the surface. These fluxes can be useful for diagnostic calculations such as radiative forcing or the cloud radiative effect, and can help understand the Earth's energy balance. The spectral subset and the corresponding linear weights are chosen to minimize the cost function

$$C = \sqrt{(F_{est}^{\uparrow TOA} - F_{ref}^{\uparrow TOA})^2 + (F_{est}^{\downarrow surf.} - F_{ref}^{\downarrow surf.})^2}. \tag{2}$$

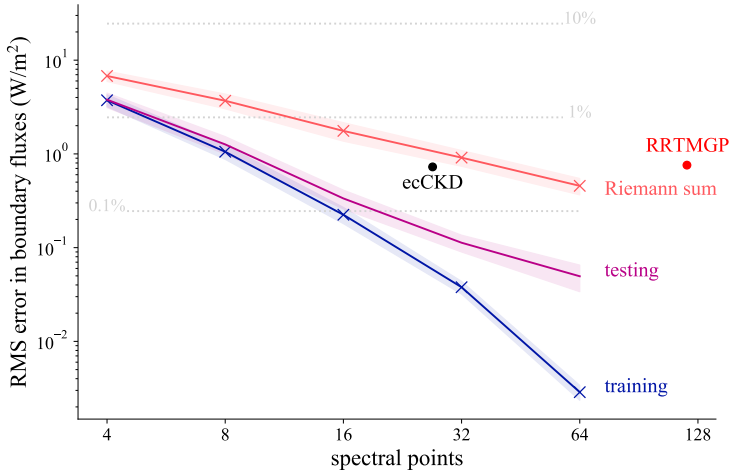

Figure 2. The root mean squared error, in units of W/m^2, across a range of wavenumber subset sizes for boundary fluxes – upward flux at the top of the atmosphere and downward at the bottom. The blue curve represents the training set, and the purple the testing, when trained on the cost function in Eqn. 2; the pink curve represents training error when the integral is performed using a Riemann sum over the chosen spectral points. We omit Riemann testing error as it is nearly identical to the training error. Envelopes denote standard deviation across sixteen independent optimizations, and crosses show where training took place. The same metric is plotted for boundary fluxes predicted by correlated-k models ecCKD (black) and RRTMGP (red). 10%, 1%, and 0.1% of mean boundary fluxes are plotted in grey.

Fluxes at the boundaries of the atmosphere can be computed accurately with relatively sparse spectral information (Fig. 2). For various representative subset sizes, the

best flux and weight configurations are recorded and the root mean squared error away from reference calculations,

$$RMSE = \sqrt{\frac{1}{N}\frac{1}{M}\sum_{m,n}(F_{est}^{m,n} - F_{ref}^{m,n})^2},$$

where n indexes the atmospheric profiles and m indexes vertical levels, is plotted. This is the same as the cost function up to a constant multiplier, which serves to take the average of the error across atmospheric profiles and vertical levels rather than the sum. Additionally, using data submitted to CKDMIP (Hogan & Matricardi, 2020) and available online, our results are plotted alongside metrics from modern correlated k-distributions, the 27 g-point ecCKD (Hogan & Matricardi, 2022), which is trained on the CKDMIP training data, and the reduced 128 g-point RRTMGP (Pincus et al., 2019).

First, to separate the contributions of simulated annealing and linear programming, weights are chosen by finding the Euclidean distance between spectral points – a Riemann sum is performed over the quadrature points chosen by the annealing algorithm. The pink curve shows that for a simple diagnostic calculation, the Riemann sum can perform at around 1 W/m^2 error with 32 spectral points, similarly to correlated-k models, highlighting the effectiveness of simulated annealing at choosing an appropriate representative subset. The purple and blue curves show the training and testing error, respectively, using both simulated annealing and linear weight optimization. Here, eight spectral points allow fluxes at the boundaries of the atmosphere to be computed to within 1 W/m^2 (1% accuracy), comparable to ecCKD and RRTMGP. With more than 16 spectral points, errors in testing data decrease much less rapidly with spectral detail than do errors in training data, indicating overfitting. However, the difference between the Riemann sum error and the RMS testing error of the model that optimizes Eqn. 2 is at close to 1 W/m^2 throughout. This highlights the importance of optimizing both the linear weights and the representative spectral subset.

3.2 Net Flux and Heating Rate Profiles

Boundary fluxes control the total energy exchanged by the Earth and the atmosphere; flux profiles throughout the atmosphere set heating rates and hence the diabatic forcing of circulation. Simultaneous prediction of fluxes and heating rate can be achieved by specifying a cost function that contains both quantities, i.e.

$$C = f_1\|F_{est} - F_{ref}\| + f_2\|H_{est} - H_{ref}\|, \tag{3}$$

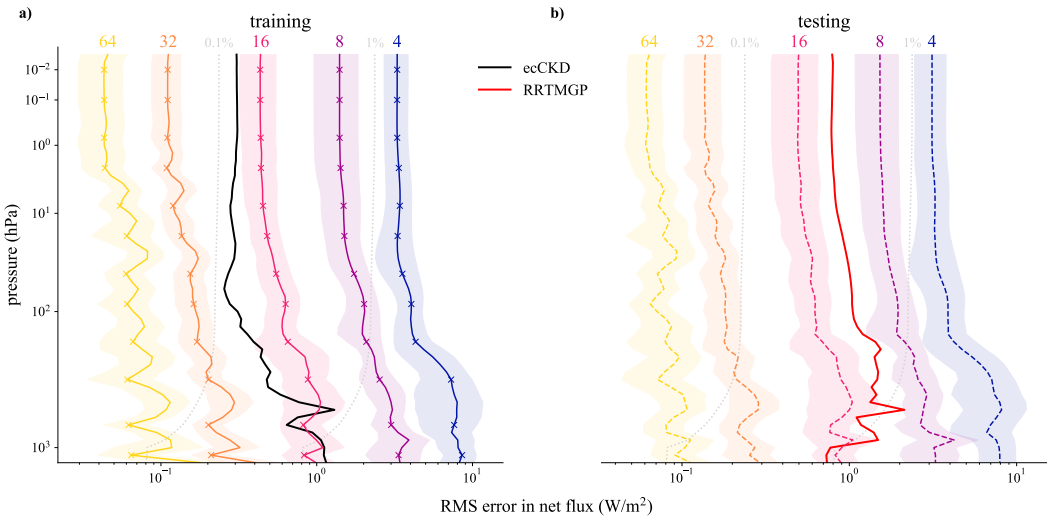

Figure 3. The root mean squared error (W/m^2) for predicted flux values for each pressure level, for a model that minimizes both flux error and heating rate error (Eqn. 3). Subset sizes are shown above the plot. Panel **a)** shows results for the training set, and panel **b)** plots out-of-sample testing data. Since the algorithm is trained only on every fifth vertical level, denoted by crosses on the training plot, in-between levels in the training set are also out-of-sample. Envelopes denote the standard deviation in RMSE across sixteen algorithm runs. We highlight ecCKD on the training plot since it is trained on the CKDMIP data (black line), and show RRTMGP on the testing plot, since it is tested on our training data (red line). 1% and 0.1% of the mean flux at each level is plotted in grey for reference.

which sums the ℓ^2 norm of the differences between estimated and reference heating rate and net flux with weights f_i across all training profiles and vertical levels, formulated similarly to Eqn. 2 in Hogan and Matricardi (2022). F_{est} is defined as the dot product between the weights and the monochromatic fluxes, as in Eqn. 1, and the heating rate is computed using this predicted flux; for each vertical level m in a given training profile:

$$H_{est} = -S\frac{g}{c_p}\frac{dF_{est}}{dp} \approx -S\frac{g}{c_p}\frac{F_{est}^m - F_{est}^{m+1}}{p^m - p^{m+1}},$$

where $S = 3600{\cdot}24$ s/day is a scale factor; $g = 9.81$ m/s is the gravitational constant; $c_p = 1004$ J/kg K is the specific heat at constant pressure for dry air; and p^m is pressure provided by the CKDMIP dataset at vertical level m, given in Pa. The heating rate is obtained in units of K/day. Finally, $f_1 = 0.15$ and $f_2 = 1$ are empirically derived

hyperparameters that balance the optimization of heating rates and fluxes, as discussed in Fig. 1.

We train the model on every fifth vertical level of the CKDMIP dataset to predict vertically-resolved fluxes for these applications. The algorithm requires random access to a large (∼170 GB) dataset, so this is a practical limitation, but also allows us to treat some levels as out-of-sample, which can help us understand the impact of the vertical discretization of the atmosphere on the accuracy of the algorithm.

In Fig. 3, we show the profiles of absolute RMSE (W/m^2) of flux predicted by the minimization of Eqn. 3 for each subset size in order to examine this vertical dependence. We compare to ecCKD on the training plot as it is trained on CKDMIP data, and to RRTMGP on the testing plot, as it is not. Both here and in Fig. 4, pressure is plotted linearly in CKDMIP model level because training levels were chosen linearly in model level. The figure demonstrates good generalizeability of the method in the vertical as well as to out-of-sample atmospheric profiles. Throughout the vertical profile, errors remain of a constant order of magnitude. Extrusions between training levels become more pronounced in larger subset sizes, indicating overfitting. Errors in the testing set are of similar order of magnitude to those in the training set. With 32 spectral points, we achieve errors within 0.3 W/m^2 (about 0.1% relative error), comparable to the correlated-k models.

Calculating heating rates from predicted fluxes with accuracy is difficult, especially at high altitudes, where the small pressure difference will magnify any errors in the flux. Furthermore, since heating rate is a derivative of flux computed by a simple finite difference, the heating rate error will be determined by the difference in flux error at adjoining vertical levels, rather than by the error in the fluxes themselves. Thus, to produce good estimates of heating rate profiles, we include heating rates in the cost function (Eqn. 3). Without this term, the optimization produces large heating rate errors in the stratosphere and above, even for large subset sizes.

Fig. 4 shows the absolute root mean squared error in heating rates, in units of K/day. Similarly as in Fig. 3, levels we trained on are marked with crosses. In panels a) and b), heating RMSE is shown only for the levels in the training set. Though errors are relatively large for small subset sizes, the errors for 32 and 64 wavenumbers are commen-

surate with correlated k-distributions throughout the atmopshere, maintaining errors of < 0.2 K/day (approaching 1%).

In panels c) and d), heating rates computed from predicted fluxes at all levels (as in Fig. 3) are shown. Here, for small subset sizes ($\leq$16 representative wavenumbers), heating rate errors are larger, especially high in the atmosphere; in the layers immediately above the ground, the errors subset sizes 4 and 8 reach as high as 8 and 5 K/day, respectively, due to the fine pressure discretization in these levels. Even for 16 predictors, where flux errors are commensurate with correlated-k models in Fig. 3 at about 1 W/m^2 (1%) error throughout the atmopshere, the heating rate error is large where pressure thickness is small. For 64 predictors, the heating rate errors of our model are smaller than 0.5 K/day for both training and testing throughout most of the atmosphere. However, the dependence of accuracy on vertical discretization is highlighted here. The method extrapolates fluxes to out-of-sample levels well because the relationship between the monochromatic fluxes and the broadband integral remains fixed. This is not true of heating rates when training is performed on certain pressure levels and extrapolation to different pressure levels is attempted.

3.3 Forcing by CO_2

The top-of-the-atmosphere radiative forcing is defined as the difference in outgoing longwave radiation (OLR) between two states. This quantity is crucial for studying how Earth's energy balance will evolve with changing atmospheric composition. As with heating rates, it is not sufficient to predict the OLR in the two states; we must also minimize their difference. We adjust our cost function accordingly:

$$C = f_1\|F_{est} - F_{ref}\| + f_2\|H_{est} - H_{ref}\| + f_3\|\mathcal{F}_{est} - \mathcal{F}_{ref}\|, \tag{4}$$

where the TOA forcing by CO_2 $\mathcal{F}_{est} = OLR_{est}^{present} - OLR_{est}^{perturbed}$, and similarly for $\mathcal{F}_{ref}$. Hyperparameters $f_1 = 0.15$ and $f_2 = 1$ are as in Eqn. 3, and f_3 is set to 1, as described by Fig. 1. Here, we use the difference between 8x preindustrial (PI) CO_2 and present-day conditions in the cost function, and extrapolate to forcing between present-day and the scenarios used in CKDMIP: LGM, PI, 2x PI, and 8x PI (Hogan & Matricardi, 2020). We also extrapolate to 16x PI (4480 ppm), to test whether our model can be useful in extreme paleoclimate scenarios (Bice & Norris, 2002).

Fig. 5 shows the RMSE for different subset sizes across all the CO_2 scenarios. Our method displays highest errors in extrapolations to very low and very high CO_2 scenarios, such as the last glacial maximum and the extreme paleoclimate scenario. With 64 spectral points, we can achieve about 0.2 W/m^2 ($< 10\%$) error across scenarios, which is similar to the correlated-k models. Training on just the 2240 ppm scenario is able to maintain errors within 10% for all climate scenarios for 32 and 64 spectral points, including extreme high and low CO_2 concentrations.

The choice of climate scenario to train on is somewhat arbitrary; training on any perturbation of CO_2 yields similar results. This is demonstrated in Fig. 6, where the RMSE in CO_2 forcing is plotted for three independent optimizations trained on the various climate scenarios, each with 32 spectral points. The 8x PI scenario is the same as Fig. 5. There are minima in the RMS at the scenarios trained on, and scenarios that are further out-of-sample yield higher errors. For example, when training on the LGM or preindustrial CO_2 concentrations, the dramatically-different 16x PI scenario is predicted with up to 1 W/m^2 error. However, the quadrature method remains flexible to different applications, as predictions for all scenarios remain within about 10% error for 32 spectral points.

4 Evaluating Optimization Priorities

One of our aspirations is to create a framework in which optimization goals can be explicitly and transparently specified. For instance, for a weather model, it may be imperative to predict fluxes and heating rates well, but forcing by CO_2 may not be a concern. Adding terms to the cost function increases optimization times. It also results in an error tradeoff, as ensuring that multiple quantities are equally optimized yields larger errors in each individual quantity.

Fig. 7 highlights the importance of including quantities of interest in the cost function. In panel a), while the model trained only on the outgoing fluxes (dark red) outperforms all others, on average, for almost all subset sizes, the others are not far behind. In panel b), the tradeoff of adding extra quantities to the cost function is demonstrated – the model trained only on fluxes is consistently more accurate than all others, and adding heating rates or CO_2 to the cost function decreases slightly accuracy for all subset sizes. As shown above, including the flux derivative in the cost function is imperative in ac-

curately reproducing heating rates. In panel c), the models trained on heating rates (orange and yellow) achieve 10% error, on average, with 32 spectral points, while the model trained only on fluxes (red) exhibits larger errors. Finally, in panel d), we explore forcing by CO_2. Here, it is clear that in order to reproduce this TOA radiative forcing, we must include it in the cost function. Although including CO_2 forcing in the cost function increases errors in the flux profiles and boundary fluxes, these increases are small.

5 Physical Interpretation of Optimization Choices

Though simulated annealing is a stochastic optimization algorithm agnostic to the physics of the Earth system, we hope to draw some physical understanding from the wavenumber subsets and associated weights that are chosen. Fig. 8 shows the location of subset members on an example OLR spectrum for three independent optimizations. Bars are shown below the spectrum noting major features: the atmospheric water vapor window as well as absorption by water vapor, carbon dioxide, and ozone. On these bars, the dots indicate where subset members lay on the spectrum for three 16-member models trained on heating rates and fluxes (Eqn. 3). The size of the dots is proportional to the value of the monochromatic flux at that wavenumber multiplied by its linear weight. This quantifies how much a given subset member contributes to the integrated broadband flux. The colors correspond to the three different models.

The optimization chooses wavenumbers that are distributed across the spectrum, spanning absorption by all three major gases as well as the water vapor window. Where the Planck curve is small, such as the tails of the spectrum, few points are chosen, and those that are tend not to be "important". The largest contributions tend to come from the water vapor window, followed by the main CO_2 and H_2O bands.

Next, the amount of absorption captured by each subset member is quantified through optical depth. For each wavenumber in a given subset S, the pressure level at which optical depth $\tau = 1$ is reached from the top of the atmosphere is recorded. The wavenumbers are then sorted by this pressure level, averaged across training atmospheric profiles and independent optimizations, and then plotted in Fig. 9. The colorful lines correspond to an optimization minimizing error in both heating rates and fluxes (Eqn. 3), while the accompanying grey curves result from optimizing only fluxes.

As the size of the subset increases, the members span more of the atmosphere – the pressure of the lowest optically-thick level decreases. Furthermore, more levels that radiate at higher pressures in the atmosphere are included: the slope of the lines below 100 hPa decreases. Finally, the models trained on only fluxes span less of the atmosphere and contain more wavenumbers that radiate from higher pressures on average. This explains why the model trained only on fluxes cannot predict heating rates at low pressures with accuracy, and further justifies our inclusion of heating rates in the cost function. It also proposes a relationship between the vertical discretization of the training atmospheres and the number of subset members required to accurately predict heating rates.

6 Conclusions

In this paper, we demonstrate that a small set of spectral points can be used to predict flux profiles, heating rate profiles, and CO_2 forcing with accuracy comparable to correlated k-distribution methods. The empirically-optimized quadrature method selects an optimal subset of spectral points at which to compute a weighted sum to estimate the broadband net flux integral. We apply the method to predicting fluxes throughout the vertical column and test its generalizeability to out-of-sample atmospheric profiles. We can achieve accuracy to within 1 W/m^2 for vertically-resolved net fluxes with 16 spectral points while avoiding overfitting.

We show that in order to reproduce heating rates accurately, especially high in the atmosphere, we must optimize not only on fluxes, but also on the difference between fluxes at consecutive height levels. Including heating rates in the cost function yields errors of less than 0.2 K/day throughout the atmospheric column for subsets of 32 points. Similarly, in order to explore applications in climate science, we include forcing by CO_2 in the cost function. We find that by training only on an 8x PI scenario, we can extrapolate to reproduce top-of-the-atmosphere forcing within 10% (< 1 W/m^2) with 32 or 64 spectral points across various scenarios. The ability to optimize quantities of interest lends flexibility to different applications to the method.

While our approach performs well on the present-day clear-sky scenarios provided in CKDMIP, there are several limitations to this initial demonstration. First, we have chosen to focus on clear-sky scenarios, as many k-distributions are initially developed

using only clear skies (e.g., Pincus et al., 2019). The optical properties of clouds vary only weakly with wavenumber so we anticipate that training on cloudy profiles would not change the accuracy or the flexibility of our quadrature method. However, as with choosing clear-sky profiles, choosing an appropriate set of cloudy training profiles is an open question, and so we leave it out of the scope of this work.

We have also limited our training and testing data to the CKDMIP sets, which commits us to certain atmospheric condition and vertical discretization choices. Choosing training atmospheres that will represent diverse conditions and generalize well is an open question for many statistical and machine learning applications, and we use only fifty. Furthermore, memory constraints (the total size of the spectral training dataset is about 170 GB) also prevent us from training on all model levels in the vertical. Though this allows us to highlight the importance of the vertical discretization of the dataset, using a finer grid would certainly improve both flux and heating rate predictions. It may also require more spectral points to parameterize the extra information.

Finally, though we know many greenhouse gases exert a forcing on the climate, we focus only on CO_2. Since we fit one set of weights with a linear regression, there is a limit on how many related quantities we can optimize at once before we formulate an ill-conditioned problem. In order to predict forcing by different gases, or improve our CO_2 forcing estimates across different scenarios, we will have to consider how best to include many different optimization targets.

7 Open Data

Model output data and the Jupyter notebook for creating figures shown in this paper are publicly available at `https://github.com/pczarnecki/quadrature-paper-figures`, doi:10.5281/zenodo.7939119. The line-by-line code and data is freely available from the CKDMIP project (Hogan and Matricardi (2020), `https://confluence.ecmwf.int/display/CKDMIP`).

8 Acknowledgements

We thank Stefan Buehler for many helpful discussions about this work. This work was supported by the US National Science Foundation (award AGS 19-16908), the Na-

tional Oceanic and Atmospheric Administration (award NA200AR4310375), and the Vetlesen Foundation.

References

Agrawal, A., Verschueren, R., Diamond, S., & Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, *5*(1), 42–60.

Bice, K. L., & Norris, R. D. (2002). Possible atmospheric CO_2 extremes of the Middle Cretaceous (late Albian–Turonian). *Paleoceanography*, *17*(4), 22–1.

Buehler, S., John, V., Kottayil, A., Milz, M., & Eriksson, P. (2010). Efficient radiative transfer simulations for a broadband infrared radiometer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *111*(4), 602–615.

Dekkers, A., & Aarts, E. (1991). Global optimization and simulated annealing. *Mathematical Programming*, *50*(1), 367–393.

de Mourgues, M., Emde, C., & Mayer, B. (2023). Optimized wavelength sampling for thermal radiative transfer in numerical weather prediction models. *Atmosphere*, *14*(2), 332.

Diamond, S., & Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, *17*(83), 1–5.

Fu, Q., & Liou, K. (1992). On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of the Atmospheric Sciences*, *49*(22), 2139–2156.

Goody, R., West, R., Chen, L., & Crisp, D. (1989). The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *42*(6), 539–550.

Gordon, I., Rothman, L., Hargreaves, R., Hashemi, R., Karlovets, E., Skinner, F., ... others (2022). The HITRAN2020 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *277*, 107949.

Hogan, R. J., & Matricardi, M. (2020). Evaluating and improving the treatment of gases in radiation schemes: the Correlated K-Distribution Model Intercomparison Project (CKDMIP). *GMD*, *13*(12), 6501–6521.

Hogan, R. J., & Matricardi, M. (2022). A tool for generating fast k-distribution gas-optics models for weather and climate applications. *Journal of Advances in*

394 *Modeling Earth Systems*, e2022MS003033.

395 Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by simulated
396 annealing. *Science*, *220*(4598), 671–680.

397 Lacis, A. A., & Oinas, V. (1991). A description of the correlated k-distribution
398 method for modeling nongray gaseous absorption, thermal emission, and multi-
399 ple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical*
400 *Research: Atmospheres*, *96*(D5), 9027–9063.

401 Manabe, S., & Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere
402 with a given distribution of relative humidity.

403 Moncet, J.-L., Uymin, G., Lipton, A. E., & Snell, H. E. (2008). Infrared radiance
404 modeling by optimal spectral sampling. *Journal of the Atmospheric Sciences*,
405 *65*(12), 3917–3934.

406 Pincus, R., Buehler, S. A., Brath, M., Crevoisier, C., Jamil, O., Franklin Evans,
407 K., ... others (2020). Benchmark calculations of radiative forcing by
408 greenhouse gases. *Journal of Geophysical Research: Atmospheres*, *125*(23),
409 e2020JD033483.

410 Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency,
411 and flexibility in radiation calculations for dynamical models. *Journal of Ad-*
412 *vances in Modeling Earth Systems*, *11*(10), 3074–3089.

413 Yang, X.-S. (2014). Chapter 14 - multi-objective optimization. In X.-S. Yang (Ed.),
414 *Nature-inspired optimization algorithms* (p. 197-211). Oxford: Elsevier. doi:
415 https://doi.org/10.1016/B978-0-12-416743-8.00014-2

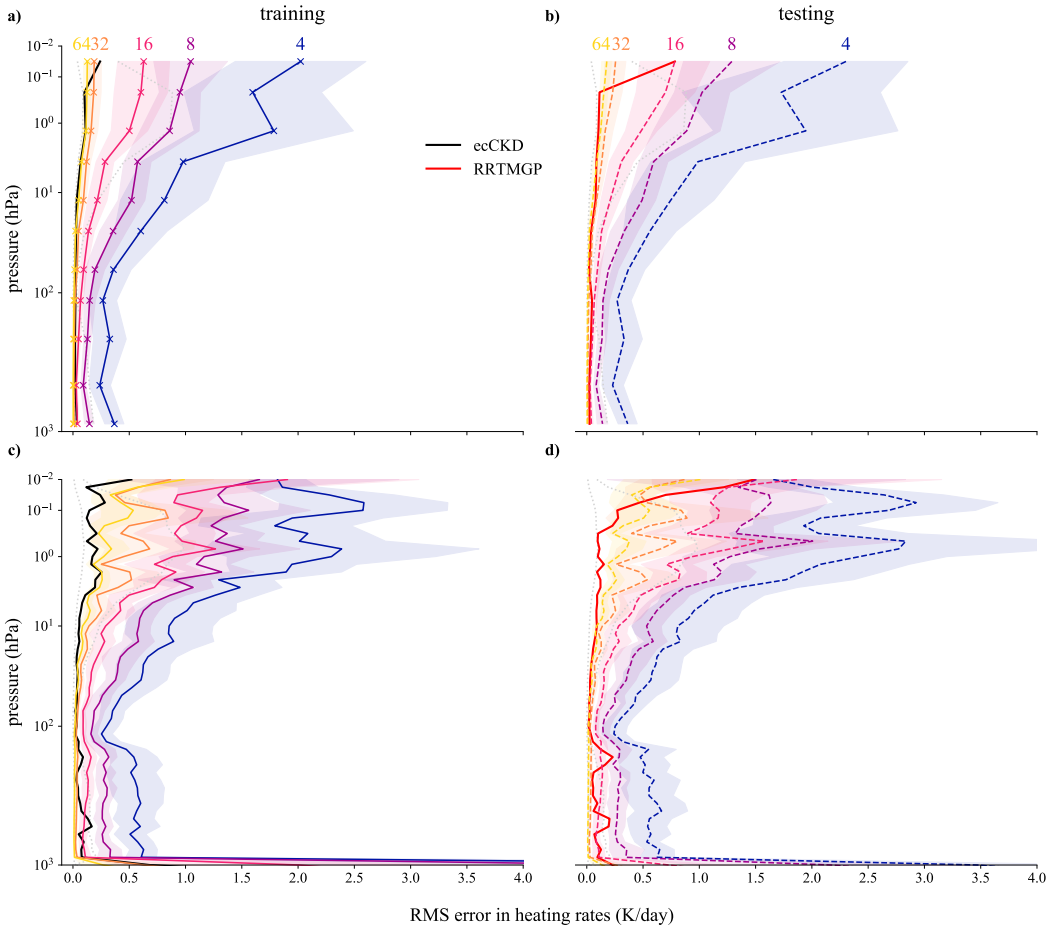

Figure 4. The absolute root mean squared error (in K/day) of the heating rate calculated from predicted fluxes throughout the atmosphere. Subset sizes are noted above the plot. Panel **a)** shows the RMSE of heating rates at only the pressure levels in the training set; panel **b)** shows the associated out-of-sample atmospheric columns; **c)** shows results for the training set using predicted fluxes for all model levels; panel **d)** shows absolute RMSE for out-of-sample atmospheric profiles for all levels. Envelopes denote the standard deviation in absolute RMSE across sixteen algorithm runs. The black and red lines show heating rates predicted by ecCKD and RRTMGP, respectively. 10% and 1% of the mean heating rate profile are plotted in grey dotted lines.

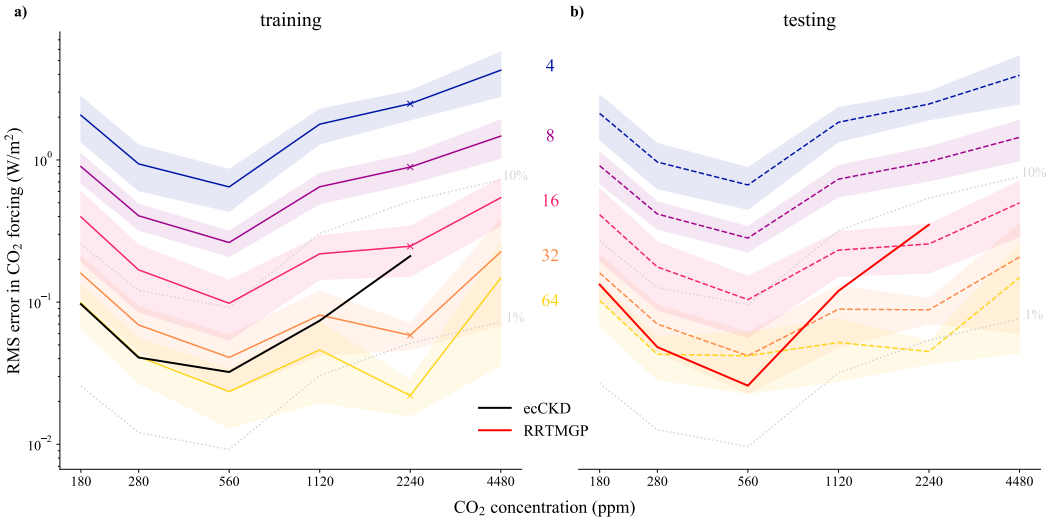

Figure 5. The root mean squared error (in units of W/m^2) of the forcing by CO_2 is plotted for different subset sizes across various climate scenarios. Panel **a)** shows the error from the training set, and testing error on panel **b)**. Crosses indicate that only forcing from the 2240 ppm training scenario is included in the cost function. The results from ecCKD and RRTMGP are shown in black and red, respectively. Clouds show the standard error across sixteen independent optimizations. 10% and 1% of the mean forcing for each scenario are plotted in grey dotted lines.

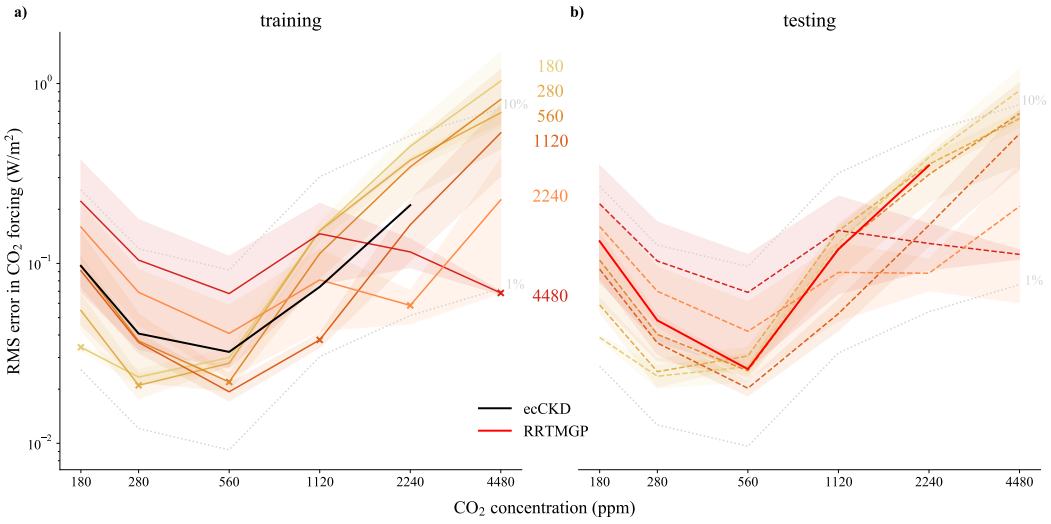

Figure 6. The root mean squared error (in units of W/m^2) of the forcing by CO_2 is plotted for a 32-wavenumber subset trained on the different climate scenarios: the last glacial maximum (LGM), pre-industrial concentrations (PI), and 2-, 4-, 8-, and 16x PI CO_2. Panel **a)** shows the error from the training set, and testing error on panel **b)**. Crosses indicate which training scenario is included in the cost function. The results from ecCKD and RRTMGP are shown in black and red, respectively. 10% and 1% of the mean forcing for each scenario are plotted in grey dotted lines. Clouds show the standard error across just three independent optimizations.

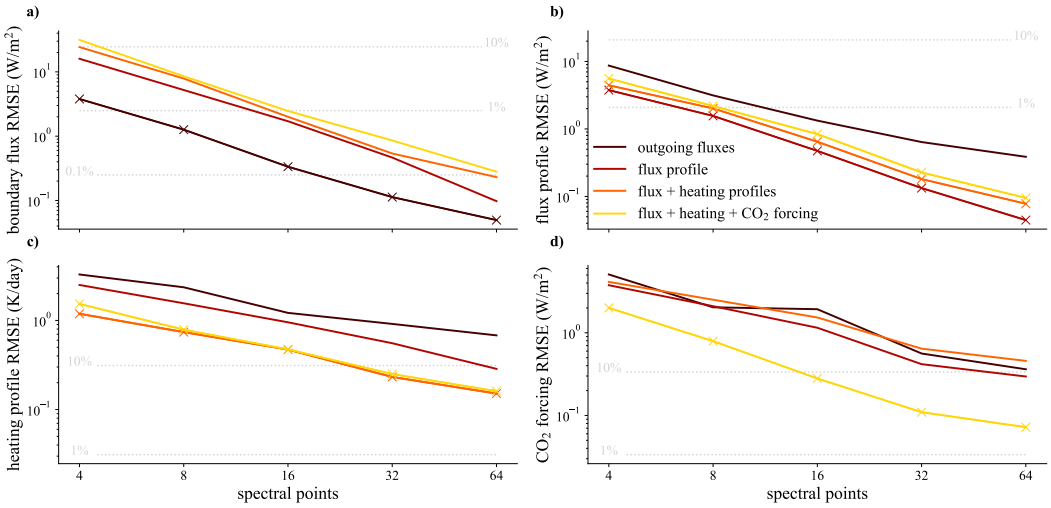

Figure 7. The average errors for models trained on various cost functions: one including only boundary fluxes (Eqn. 2), in dark red; one including every fifth net flux in the vertical, in red; one trained on these fluxes as well as corresponding heating rates (Eqn. 3), in orange; and one where the cost function encompasses fluxes, heating rates, and the CO_2 forcing between present day and 8x PI CO_2 (Eqn. 4), in yellow. The panels show: **a)** the mean RMSE for boundary fluxes; **b)** the mean RMSE for fluxes throughout the atmospheric column; **c)** the average absolute RMSE of heating rates throughout the atmospheric profile and **d)** the average RMSE of forcing by CO_2 across climate scenarios. The crosses indicate where the given quantity was included in the cost function. 10%, 1% and 0.1% of the mean reference values for each quantity are plotted in grey dotted lines.

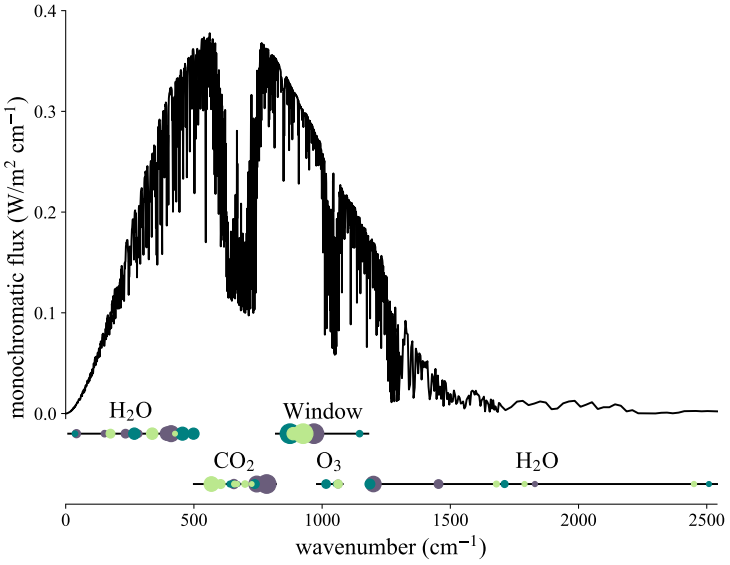

Figure 8. Below an example of a spectrum of OLR, major absorbers are noted. Dots mark where a wavenumber was chosen, and the size of the dots is proportional to the importance of that spectral point to the broadband flux integration. The colors correspond to three independent optimizations.

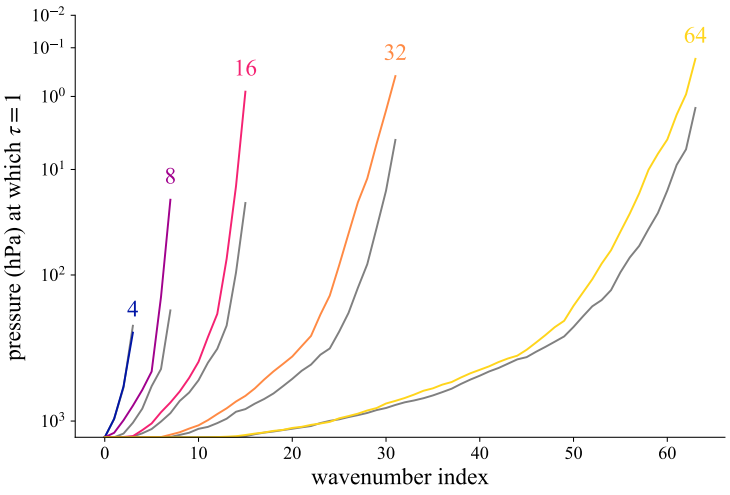

Figure 9. For each wavenumber in the final subset S, the pressure level at which that wavenumber first reaches optical depth $\tau = 1$ from the top of the atmosphere is plotted. The colorful lines correspond to models optimizing both fluxes and heating rates (Eqn. 3), and the accompanying grey lines correspond to models optimizing only fluxes.

Sparse, Empirically Optimized Quadrature for Radiative Calculations

Paulina Czarnecki[1], Robert Pincus[2], Lorenzo Polvani[1,2,3]

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA
[2]Lamont-Doherty Earth Observatory, Palisades, NY, USA
[3]Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA

Key Points:

- Spectrally integrated radiative quantities can be estimated from a small weighted set of monochromatic calculations.
- Tradeoffs between computational cost and accuracy are commensurate with correlated k-distributions.
- Optimization priorities can be tailored to a specific applications.

Abstract

Broadband (spectrally-integrated) radiation calculations are dominated by the expense of spectral integration, and many applications require fast parameterizations for computing radiative flux. Here we describe a novel approach using a linear weighted sum of monochromatic calculations at a small set of optimally-chosen frequencies. The empirically-optimized quadrature method is used to compute atmospheric boundary fluxes, net flux profiles throughout the atmosphere, heating rate profiles, and top-of-the-atmosphere forcing by CO_2. We evaluate the method against two modern correlated k-distributions and find that we can achieve comparable errors with 32 spectral points. We also examine the effect of minimizing different cost functions, and find that in order to accurately represent heating rates and CO_2 forcing, these quantities must be included in the cost function.

Plain Language Summary

Quantifying the way radiation flows through the atmosphere is computationally expensive, and most applications require fast approximation. In this paper, we develop an empirically-optimized quadrature method that can compute quantities of interest in climate science and weather prediction such as radiative flux throughout the atmosphere, heating rates, and forcing by carbon dioxide. The method uses a simple optimization algorithm, combined with a linear model, to achieve accuracy and computational cost of similar order of magnitude to the modern correlated k-distributions. Depending on the focus of the application, the algorithm can easily be adapted to prioritize desired quantities.

1 Approximations for Spectral Integrals in Radiative Transfer Calculations

Radiation is the mechanism by which the Earth system achieves energy balance; radiative heating and cooling within the atmosphere also play a crucial role in setting the temperature profile of the atmosphere (Manabe & Wetherald, 1967). The strength of the interactions between radiation and the atmosphere depends strongly on frequency because the optical properties of gases, fundamentally determined by their molecular structure along with local temperature and pressure, vary rapidly in frequency by many or-

ders of magnitude. (The optical properties of condensed materials such as clouds and aerosols vary much more slowly with frequency and are not considered here.)

The optical properties of the gases in the atmosphere can be determined at given frequencies based on detailed spectroscopic information describing each absorption line for each species (e.g., HITRAN, Gordon et al., 2022). The spectrally-integrated (broadband) exchange of radiation can be determined from this information to near-arbitrary accuracy by direct integration at very high spectral resolution. Results from line-by-line models taking this approach (along with high angular resolution) may be considered reference calculations (Pincus et al., 2020).

Highly detailed calculations are, however, impractical in many contexts, such as when radiation must be computed as a diabatic forcing for a model of the atmosphere. In these contexts the radiative transfer integral is approximated using parameterizations, which reduce both spectral and angular detail. The state-of-the-art method for reducing spectral resolution is the correlated k-distribution (CKD) algorithm (Fu & Liou, 1992; Lacis & Oinas, 1991; Goody et al., 1989), which sorts frequencies according to their absorption coefficient k and determines the average at some set of quadrature points. The sorting and averaging creates a much smoother integral that can be evaluated with just tens or hundreds of quadrature points, as compared to the millions of points required by line-by-line models. These parameterizations seek to minimize the number of these quadrature points, called g-points, to limit computational cost (Hogan & Matricardi, 2022).

There is, however, no theory underlying k-distributions, so developers must make a range of somewhat arbitrary choices (Hogan & Matricardi, 2020). Moreover, the common experience is that the absorption coefficients determined with any algorithm need to be tuned to match reference calculations (Hogan & Matricardi, 2022; Pincus et al., 2019), making the link between the reference and approximate calculations somewhat obscure.

Here we describe an approximation to spectral integration in the longwave consisting of a weighted sum of monochromatic fluxes computed at a small set of discrete wavenumbers. We optimize both the members of this set and the linear weights to find the best estimate of the broadband integrated flux. The idea builds on approaches for computing narrow-band radiances originally developed for remote sensing applications (Buehler et al., 2010; Moncet et al., 2008) to sample the spectrum sparsely rather than averag-

ing absorption coefficients. Our work is in the same spirit as de Mourgues et al. (2023), though we take a more general approach and treat a wider range of applications including heating rates, boundary fluxes, and CO_2 forcing. We explore the relationship between the quantities targeted in the optimization and individual error profiles, test the effect of the vertical discretization of the training dataset, and compare the performance of the empirically-optimized quadrature method to state-of-the-art correlated-k methods. We find that weighted, sampled monochomatic calculations can predict fluxes, heating rates, and CO_2 forcing with error-cost relationships commensurate with modern correlated-k models, while being flexible to the needs of a given application.

2 Broadband Radiative Integral Calculation

The net broadband radiative flux at height z in the atmosphere, $F(z)$, is given by integrating monochromatic fluxes across all wavenumbers:

$$F(z) = \int F_\nu(z)\, d\nu \approx \sum_i F_{\nu_i}(z)\Delta\nu_i,$$

where ν denotes wavenumber and $F_\nu(z)$ is monochromatic net flux. We develop the method for the longwave in clear skies, so we disregard scattering.

Rather than calculating the line-by-line integral directly, we seek to compute the broadband net flux from a small subset of representative wavenumbers. This breaks the problem into two parts:

- Optimally select the representative wavenumbers, and
- Compute the net integrated flux from this subset.

Inspired by approaches originally developed for remote sensing applications (Buehler et al., 2010; Moncet et al., 2008), we select the representative subset via a combinatorial optimization algorithm called simulated annealing (Kirkpatrick et al., 1983), and calculate the net flux by performing a weighted sum of monochromatic fluxes found at the representative wavenumbers. In other words, we rewrite our problem as:

$$F(z) \approx \sum_{\nu \in S} w_\nu F_\nu(z), \tag{1}$$

where w_ν is the weight corresponding to wavenumber $\nu \in S$, and $F_\nu(z)$ is the monochromatic flux found at wavenumber ν at height z. Though this is similar to the formula-

tion of k-distributions (for example, Eqn. 3 in Lacis and Oinas (1991)), here we are sampling monochromatic fluxes rather than averaging absorption coefficients.

2.1 Line-by-Line Spectra and Correlated-K Data

In order to spectrally sample $F_{\nu \in S}$ and choose w_ν, we need accurate, finely-resolved line-by-line spectra at arbitrary frequencies, as well as corresponding integrated fluxes and heating rates throughout the atmosphere. Atmospheric profiles used for training must be independent and provide a representative sample of conditions for a given applications.

We use the set of atmospheric conditions developed for the Correlated K-Distribution Model Intercomparison Project (CKDMIP; see Hogan & Matricardi, 2020). The CKDMIP profiles are chosen from reanalysis data to represent the mean state as well as minimum and maximum extremes in water vapor concentration, ozone concentration, and temperature for the present day climate. We use software provided by CKDMIP to generate 100 independent present-day, clear-sky reference flux profiles, half for training and half for testing, integrated in 4 hemispheric angles. Heating rates are computed from fluxes using a simple finite difference. Finally, we perturb the CO_2 concentration of the 100 profiles to generate reference CO_2 forcing values for the Last Glacial Maximum (180 ppm), preindustrial (280 ppm), present day ($\sim$415 ppm), 2x PI (560 ppm), 4x PI (1120 ppm), 8x PI (2240 ppm) (Hogan & Matricardi, 2020), and 16x PI (4480 ppm) for extreme paleoclimate scenarios (Bice & Norris, 2002).

2.2 Predicting Net Flux from a Subset of Spectral Points

Given a subset of spectral points S of predetermined size $|S|$, we must obtain a set of weights such that the weighted sum of fluxes at those points predicts the broadband net flux across the spectrum. The weights w_ν are found by minimizing a linear cost function subject to several constraints.

First, all $w_\nu \geq 0$, since a negative weight would indicate redundant information (Buehler et al., 2010; Moncet et al., 2008). Also, the weights are normalized so that they sum to the length of the spectral domain, a general property of integration:

$$\sum w_\nu = \sum_i \Delta\nu_i \approx 3260 \text{ cm}^{-1},$$

as we focus on the longwave. This is achieved by formulating the cost function as a linear program and solving using the convex optimization package `cvxpy` in Python (Agrawal et al., 2018; Diamond & Boyd, 2016).

2.3 Wavenumber Subset Optimization Via Simulated Annealing

Next, we must identify an optimal subset at which to compute the predicted flux. We use simulated annealing, a stochastic optimization algorithm that can combinatorially evaluate different subsets in order to find a global minimum in a given cost function (Kirkpatrick et al., 1983). The algorithm is guaranteed to converge asymptotically to a global minimum in cost function with probability 1 (Dekkers & Aarts, 1991), and can do so at much lower computational cost than a brute force approach for a large state space.

Buehler et al. (2010) provides a useful tutorial on algorithm implementation. Annealing temperature is initialized such that 99% of moves will be accepted (Buehler et al., 2010). The optimization is organized in blocks of moves that the cost function value is averaged over (Buehler et al., 2010). Each block consists of 100 moves that must all be successful; otherwise, the algorithm is considered to have converged to a minimum and is terminated. Finally, when a block has a lower average cost function value than the previous block, the annealing temperature is decreased by a factor of 0.9 (Buehler et al., 2010).

Because the optimization process is stochastic, we perform 16 independent optimizations for each subset size and plot the standard deviation away from the mean error in our results. We refer to a given set of wavenumbers and corresponding weights as a model.

2.4 Cost Function Formulation

As we are interested in optimizing multiple radiative quantities – fluxes, heating rates, and top-of-the-atmosphere CO_2 forcing – we take inspiration from multi-objective optimization in the formulation of our cost function (Yang, 2014). Given multiple optimization targets F, H, and $\mathcal{F}$, the simplest way to minimize their respective cost functions c simultaneously is to minimize their weighted sum:

$$C = f_1\mathbf{c}(F) + f_2\mathbf{c}(H) + f_3\mathbf{c}(\mathcal{F}).$$

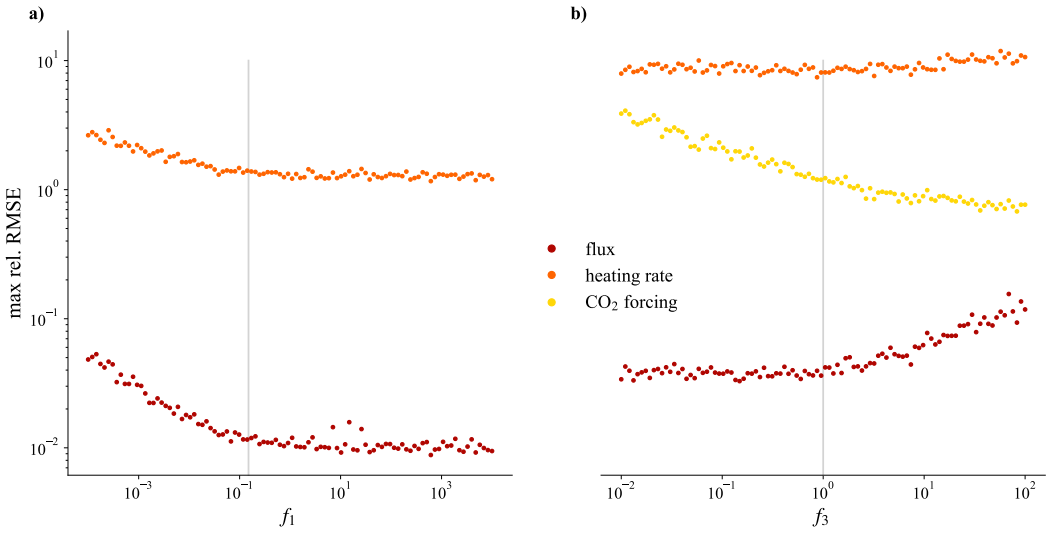

Figure 1. The relative root mean squared error (RMSE) of each radiative term of interest is plotted as weighting parameters vary. In **a)**, the weight on the heating rate term $f_2 = 1$ is held fixed while f_1 is varied. In **b)**, $f_1 = 0.15$ and $f_2 = 1$ are held fixed while f_3 is varied. The grey lines show the value of the parameter that is used throughout the study.

Choosing the values of the weights f_i is considered a subjective task with many valid solutions (Yang, 2014). In our case, the approach is as follows: first, we fix $f_2 = 1$ and $f_3 = 0$ as we vary f_1, to focus on the tradeoff between heating rates and fluxes. For each value of f_1, we fit linear weights to 100 random subsets of size 32 at every fifth vertical level for each of the 50 training atmospheres. The maximum relative (percent) error across vertical levels is recorded for each of the 100 iterations. The mean of these 100 iterations is plotted, where any random subset that produces an ill-conditioned cost function is ignored. The results of this experiment are shown in Fig. 1a). The grey line indicates the parameter value $f_1 = 0.15$ that is chosen, as it occurs where heating rate errors level off.

In Fig. 1b), the same procedure is repeated while holding $f_1 = 0.15$ and $f_2 = 1$ fixed and varying f_3. We choose parameter $f_3 = 1$ in order to minimize increase in heating rate and flux errors.

3 Accuracy in Flux, Heating Rate, and CO_2 Forcing Prediction

3.1 Boundary Fluxes

We illustrate the method by training the algorithm only on the outgoing atmospheric boundary fluxes: the upwelling flux at the top of the atmosphere and the downwelling flux at the surface. These fluxes can be useful for diagnostic calculations such as radiative forcing or the cloud radiative effect, and can help understand the Earth's energy balance. The spectral subset and the corresponding linear weights are chosen to minimize the cost function

$$C = \sqrt{(F_{est}^{\uparrow TOA} - F_{ref}^{\uparrow TOA})^2 + (F_{est}^{\downarrow surf.} - F_{ref}^{\downarrow surf.})^2}. \tag{2}$$

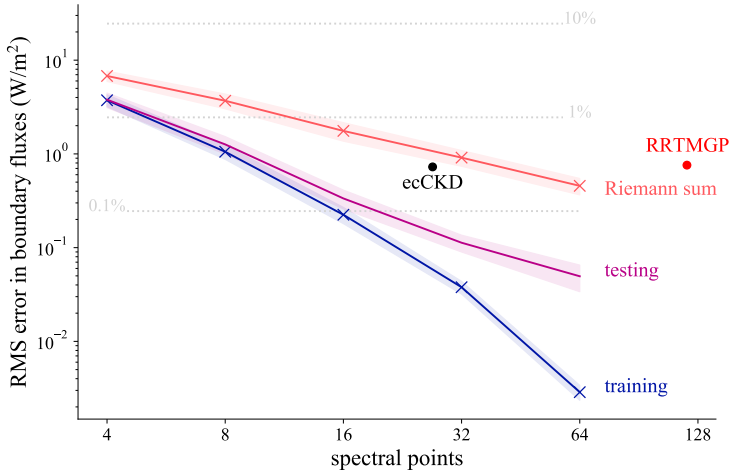

Figure 2. The root mean squared error, in units of W/m^2, across a range of wavenumber subset sizes for boundary fluxes – upward flux at the top of the atmosphere and downward at the bottom. The blue curve represents the training set, and the purple the testing, when trained on the cost function in Eqn. 2; the pink curve represents training error when the integral is performed using a Riemann sum over the chosen spectral points. We omit Riemann testing error as it is nearly identical to the training error. Envelopes denote standard deviation across sixteen independent optimizations, and crosses show where training took place. The same metric is plotted for boundary fluxes predicted by correlated-k models ecCKD (black) and RRTMGP (red). 10%, 1%, and 0.1% of mean boundary fluxes are plotted in grey.

Fluxes at the boundaries of the atmosphere can be computed accurately with relatively sparse spectral information (Fig. 2). For various representative subset sizes, the

best flux and weight configurations are recorded and the root mean squared error away from reference calculations,

$$RMSE = \sqrt{\frac{1}{N}\frac{1}{M}\sum_{m,n}(F_{est}^{m,n} - F_{ref}^{m,n})^2},$$

where n indexes the atmospheric profiles and m indexes vertical levels, is plotted. This is the same as the cost function up to a constant multiplier, which serves to take the average of the error across atmospheric profiles and vertical levels rather than the sum. Additionally, using data submitted to CKDMIP (Hogan & Matricardi, 2020) and available online, our results are plotted alongside metrics from modern correlated k-distributions, the 27 g-point ecCKD (Hogan & Matricardi, 2022), which is trained on the CKDMIP training data, and the reduced 128 g-point RRTMGP (Pincus et al., 2019).

First, to separate the contributions of simulated annealing and linear programming, weights are chosen by finding the Euclidean distance between spectral points – a Riemann sum is performed over the quadrature points chosen by the annealing algorithm. The pink curve shows that for a simple diagnostic calculation, the Riemann sum can perform at around 1 W/m^2 error with 32 spectral points, similarly to correlated-k models, highlighting the effectiveness of simulated annealing at choosing an appropriate representative subset. The purple and blue curves show the training and testing error, respectively, using both simulated annealing and linear weight optimization. Here, eight spectral points allow fluxes at the boundaries of the atmosphere to be computed to within 1 W/m^2 (1% accuracy), comparable to ecCKD and RRTMGP. With more than 16 spectral points, errors in testing data decrease much less rapidly with spectral detail than do errors in training data, indicating overfitting. However, the difference between the Riemann sum error and the RMS testing error of the model that optimizes Eqn. 2 is at close to 1 W/m^2 throughout. This highlights the importance of optimizing both the linear weights and the representative spectral subset.

3.2 Net Flux and Heating Rate Profiles

Boundary fluxes control the total energy exchanged by the Earth and the atmosphere; flux profiles throughout the atmosphere set heating rates and hence the diabatic forcing of circulation. Simultaneous prediction of fluxes and heating rate can be achieved by specifying a cost function that contains both quantities, i.e.

$$C = f_1\|F_{est} - F_{ref}\| + f_2\|H_{est} - H_{ref}\|, \tag{3}$$

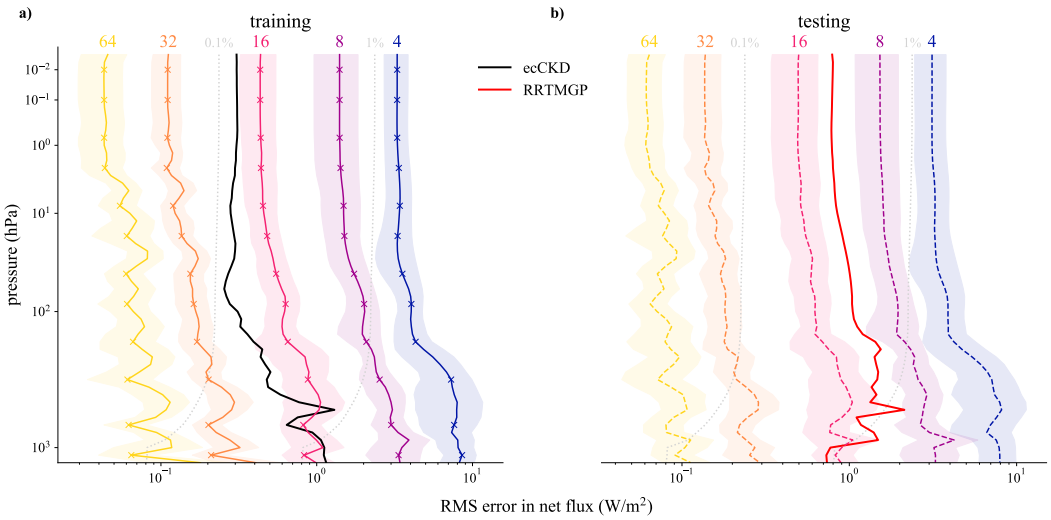

Figure 3. The root mean squared error (W/m^2) for predicted flux values for each pressure level, for a model that minimizes both flux error and heating rate error (Eqn. 3). Subset sizes are shown above the plot. Panel **a)** shows results for the training set, and panel **b)** plots out-of-sample testing data. Since the algorithm is trained only on every fifth vertical level, denoted by crosses on the training plot, in-between levels in the training set are also out-of-sample. Envelopes denote the standard deviation in RMSE across sixteen algorithm runs. We highlight ecCKD on the training plot since it is trained on the CKDMIP data (black line), and show RRTMGP on the testing plot, since it is tested on our training data (red line). 1% and 0.1% of the mean flux at each level is plotted in grey for reference.

which sums the ℓ^2 norm of the differences between estimated and reference heating rate and net flux with weights f_i across all training profiles and vertical levels, formulated similarly to Eqn. 2 in Hogan and Matricardi (2022). F_{est} is defined as the dot product between the weights and the monochromatic fluxes, as in Eqn. 1, and the heating rate is computed using this predicted flux; for each vertical level m in a given training profile:

$$H_{est} = -S\frac{g}{c_p}\frac{dF_{est}}{dp} \approx -S\frac{g}{c_p}\frac{F_{est}^m - F_{est}^{m+1}}{p^m - p^{m+1}},$$

where $S = 3600{\cdot}24$ s/day is a scale factor; $g = 9.81$ m/s is the gravitational constant; $c_p = 1004$ J/kg K is the specific heat at constant pressure for dry air; and p^m is pressure provided by the CKDMIP dataset at vertical level m, given in Pa. The heating rate is obtained in units of K/day. Finally, $f_1 = 0.15$ and $f_2 = 1$ are empirically derived

hyperparameters that balance the optimization of heating rates and fluxes, as discussed in Fig. 1.

We train the model on every fifth vertical level of the CKDMIP dataset to predict vertically-resolved fluxes for these applications. The algorithm requires random access to a large (∼170 GB) dataset, so this is a practical limitation, but also allows us to treat some levels as out-of-sample, which can help us understand the impact of the vertical discretization of the atmosphere on the accuracy of the algorithm.

In Fig. 3, we show the profiles of absolute RMSE (W/m^2) of flux predicted by the minimization of Eqn. 3 for each subset size in order to examine this vertical dependence. We compare to ecCKD on the training plot as it is trained on CKDMIP data, and to RRTMGP on the testing plot, as it is not. Both here and in Fig. 4, pressure is plotted linearly in CKDMIP model level because training levels were chosen linearly in model level. The figure demonstrates good generalizeability of the method in the vertical as well as to out-of-sample atmospheric profiles. Throughout the vertical profile, errors remain of a constant order of magnitude. Extrusions between training levels become more pronounced in larger subset sizes, indicating overfitting. Errors in the testing set are of similar order of magnitude to those in the training set. With 32 spectral points, we achieve errors within 0.3 W/m^2 (about 0.1% relative error), comparable to the correlated-k models.

Calculating heating rates from predicted fluxes with accuracy is difficult, especially at high altitudes, where the small pressure difference will magnify any errors in the flux. Furthermore, since heating rate is a derivative of flux computed by a simple finite difference, the heating rate error will be determined by the difference in flux error at adjoining vertical levels, rather than by the error in the fluxes themselves. Thus, to produce good estimates of heating rate profiles, we include heating rates in the cost function (Eqn. 3). Without this term, the optimization produces large heating rate errors in the stratosphere and above, even for large subset sizes.

Fig. 4 shows the absolute root mean squared error in heating rates, in units of K/day. Similarly as in Fig. 3, levels we trained on are marked with crosses. In panels a) and b), heating RMSE is shown only for the levels in the training set. Though errors are relatively large for small subset sizes, the errors for 32 and 64 wavenumbers are commen-

surate with correlated k-distributions throughout the atmopshere, maintaining errors of < 0.2 K/day (approaching 1%).

In panels c) and d), heating rates computed from predicted fluxes at all levels (as in Fig. 3) are shown. Here, for small subset sizes ($\leq$16 representative wavenumbers), heating rate errors are larger, especially high in the atmosphere; in the layers immediately above the ground, the errors subset sizes 4 and 8 reach as high as 8 and 5 K/day, respectively, due to the fine pressure discretization in these levels. Even for 16 predictors, where flux errors are commensurate with correlated-k models in Fig. 3 at about 1 W/m^2 (1%) error throughout the atmopshere, the heating rate error is large where pressure thickness is small. For 64 predictors, the heating rate errors of our model are smaller than 0.5 K/day for both training and testing throughout most of the atmosphere. However, the dependence of accuracy on vertical discretization is highlighted here. The method extrapolates fluxes to out-of-sample levels well because the relationship between the monochromatic fluxes and the broadband integral remains fixed. This is not true of heating rates when training is performed on certain pressure levels and extrapolation to different pressure levels is attempted.

3.3 Forcing by CO_2

The top-of-the-atmosphere radiative forcing is defined as the difference in outgoing longwave radiation (OLR) between two states. This quantity is crucial for studying how Earth's energy balance will evolve with changing atmospheric composition. As with heating rates, it is not sufficient to predict the OLR in the two states; we must also minimize their difference. We adjust our cost function accordingly:

$$C = f_1\|F_{est} - F_{ref}\| + f_2\|H_{est} - H_{ref}\| + f_3\|\mathcal{F}_{est} - \mathcal{F}_{ref}\|, \tag{4}$$

where the TOA forcing by CO_2 $\mathcal{F}_{est} = OLR_{est}^{present} - OLR_{est}^{perturbed}$, and similarly for $\mathcal{F}_{ref}$. Hyperparameters $f_1 = 0.15$ and $f_2 = 1$ are as in Eqn. 3, and f_3 is set to 1, as described by Fig. 1. Here, we use the difference between 8x preindustrial (PI) CO_2 and present-day conditions in the cost function, and extrapolate to forcing between present-day and the scenarios used in CKDMIP: LGM, PI, 2x PI, and 8x PI (Hogan & Matricardi, 2020). We also extrapolate to 16x PI (4480 ppm), to test whether our model can be useful in extreme paleoclimate scenarios (Bice & Norris, 2002).

Fig. 5 shows the RMSE for different subset sizes across all the CO_2 scenarios. Our method displays highest errors in extrapolations to very low and very high CO_2 scenarios, such as the last glacial maximum and the extreme paleoclimate scenario. With 64 spectral points, we can achieve about 0.2 W/m^2 ($< 10\%$) error across scenarios, which is similar to the correlated-k models. Training on just the 2240 ppm scenario is able to maintain errors within 10% for all climate scenarios for 32 and 64 spectral points, including extreme high and low CO_2 concentrations.

The choice of climate scenario to train on is somewhat arbitrary; training on any perturbation of CO_2 yields similar results. This is demonstrated in Fig. 6, where the RMSE in CO_2 forcing is plotted for three independent optimizations trained on the various climate scenarios, each with 32 spectral points. The 8x PI scenario is the same as Fig. 5. There are minima in the RMS at the scenarios trained on, and scenarios that are further out-of-sample yield higher errors. For example, when training on the LGM or preindustrial CO_2 concentrations, the dramatically-different 16x PI scenario is predicted with up to 1 W/m^2 error. However, the quadrature method remains flexible to different applications, as predictions for all scenarios remain within about 10% error for 32 spectral points.

4 Evaluating Optimization Priorities

One of our aspirations is to create a framework in which optimization goals can be explicitly and transparently specified. For instance, for a weather model, it may be imperative to predict fluxes and heating rates well, but forcing by CO_2 may not be a concern. Adding terms to the cost function increases optimization times. It also results in an error tradeoff, as ensuring that multiple quantities are equally optimized yields larger errors in each individual quantity.

Fig. 7 highlights the importance of including quantities of interest in the cost function. In panel a), while the model trained only on the outgoing fluxes (dark red) outperforms all others, on average, for almost all subset sizes, the others are not far behind. In panel b), the tradeoff of adding extra quantities to the cost function is demonstrated – the model trained only on fluxes is consistently more accurate than all others, and adding heating rates or CO_2 to the cost function decreases slightly accuracy for all subset sizes. As shown above, including the flux derivative in the cost function is imperative in ac-

curately reproducing heating rates. In panel c), the models trained on heating rates (orange and yellow) achieve 10% error, on average, with 32 spectral points, while the model trained only on fluxes (red) exhibits larger errors. Finally, in panel d), we explore forcing by CO_2. Here, it is clear that in order to reproduce this TOA radiative forcing, we must include it in the cost function. Although including CO_2 forcing in the cost function increases errors in the flux profiles and boundary fluxes, these increases are small.

5 Physical Interpretation of Optimization Choices

Though simulated annealing is a stochastic optimization algorithm agnostic to the physics of the Earth system, we hope to draw some physical understanding from the wavenumber subsets and associated weights that are chosen. Fig. 8 shows the location of subset members on an example OLR spectrum for three independent optimizations. Bars are shown below the spectrum noting major features: the atmospheric water vapor window as well as absorption by water vapor, carbon dioxide, and ozone. On these bars, the dots indicate where subset members lay on the spectrum for three 16-member models trained on heating rates and fluxes (Eqn. 3). The size of the dots is proportional to the value of the monochromatic flux at that wavenumber multiplied by its linear weight. This quantifies how much a given subset member contributes to the integrated broadband flux. The colors correspond to the three different models.

The optimization chooses wavenumbers that are distributed across the spectrum, spanning absorption by all three major gases as well as the water vapor window. Where the Planck curve is small, such as the tails of the spectrum, few points are chosen, and those that are tend not to be "important". The largest contributions tend to come from the water vapor window, followed by the main CO_2 and H_2O bands.

Next, the amount of absorption captured by each subset member is quantified through optical depth. For each wavenumber in a given subset S, the pressure level at which optical depth $\tau = 1$ is reached from the top of the atmosphere is recorded. The wavenumbers are then sorted by this pressure level, averaged across training atmospheric profiles and independent optimizations, and then plotted in Fig. 9. The colorful lines correspond to an optimization minimizing error in both heating rates and fluxes (Eqn. 3), while the accompanying grey curves result from optimizing only fluxes.

As the size of the subset increases, the members span more of the atmosphere – the pressure of the lowest optically-thick level decreases. Furthermore, more levels that radiate at higher pressures in the atmosphere are included: the slope of the lines below 100 hPa decreases. Finally, the models trained on only fluxes span less of the atmosphere and contain more wavenumbers that radiate from higher pressures on average. This explains why the model trained only on fluxes cannot predict heating rates at low pressures with accuracy, and further justifies our inclusion of heating rates in the cost function. It also proposes a relationship between the vertical discretization of the training atmospheres and the number of subset members required to accurately predict heating rates.

6 Conclusions

In this paper, we demonstrate that a small set of spectral points can be used to predict flux profiles, heating rate profiles, and CO_2 forcing with accuracy comparable to correlated k-distribution methods. The empirically-optimized quadrature method selects an optimal subset of spectral points at which to compute a weighted sum to estimate the broadband net flux integral. We apply the method to predicting fluxes throughout the vertical column and test its generalizeability to out-of-sample atmospheric profiles. We can achieve accuracy to within 1 W/m^2 for vertically-resolved net fluxes with 16 spectral points while avoiding overfitting.

We show that in order to reproduce heating rates accurately, especially high in the atmosphere, we must optimize not only on fluxes, but also on the difference between fluxes at consecutive height levels. Including heating rates in the cost function yields errors of less than 0.2 K/day throughout the atmospheric column for subsets of 32 points. Similarly, in order to explore applications in climate science, we include forcing by CO_2 in the cost function. We find that by training only on an 8x PI scenario, we can extrapolate to reproduce top-of-the-atmosphere forcing within 10% (< 1 W/m^2) with 32 or 64 spectral points across various scenarios. The ability to optimize quantities of interest lends flexibility to different applications to the method.

While our approach performs well on the present-day clear-sky scenarios provided in CKDMIP, there are several limitations to this initial demonstration. First, we have chosen to focus on clear-sky scenarios, as many k-distributions are initially developed

using only clear skies (e.g., Pincus et al., 2019). The optical properties of clouds vary only weakly with wavenumber so we anticipate that training on cloudy profiles would not change the accuracy or the flexibility of our quadrature method. However, as with choosing clear-sky profiles, choosing an appropriate set of cloudy training profiles is an open question, and so we leave it out of the scope of this work.

We have also limited our training and testing data to the CKDMIP sets, which commits us to certain atmospheric condition and vertical discretization choices. Choosing training atmospheres that will represent diverse conditions and generalize well is an open question for many statistical and machine learning applications, and we use only fifty. Furthermore, memory constraints (the total size of the spectral training dataset is about 170 GB) also prevent us from training on all model levels in the vertical. Though this allows us to highlight the importance of the vertical discretization of the dataset, using a finer grid would certainly improve both flux and heating rate predictions. It may also require more spectral points to parameterize the extra information.

Finally, though we know many greenhouse gases exert a forcing on the climate, we focus only on CO_2. Since we fit one set of weights with a linear regression, there is a limit on how many related quantities we can optimize at once before we formulate an ill-conditioned problem. In order to predict forcing by different gases, or improve our CO_2 forcing estimates across different scenarios, we will have to consider how best to include many different optimization targets.

7 Open Data

Model output data and the Jupyter notebook for creating figures shown in this paper are publicly available at `https://github.com/pczarnecki/quadrature-paper-figures`, doi:10.5281/zenodo.7939119. The line-by-line code and data is freely available from the CKDMIP project (Hogan and Matricardi (2020), `https://confluence.ecmwf.int/display/CKDMIP`).

8 Acknowledgements

We thank Stefan Buehler for many helpful discussions about this work. This work was supported by the US National Science Foundation (award AGS 19-16908), the Na-

tional Oceanic and Atmospheric Administration (award NA200AR4310375), and the Vetlesen Foundation.

References

Agrawal, A., Verschueren, R., Diamond, S., & Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, *5*(1), 42–60.

Bice, K. L., & Norris, R. D. (2002). Possible atmospheric CO_2 extremes of the Middle Cretaceous (late Albian–Turonian). *Paleoceanography*, *17*(4), 22–1.

Buehler, S., John, V., Kottayil, A., Milz, M., & Eriksson, P. (2010). Efficient radiative transfer simulations for a broadband infrared radiometer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *111*(4), 602–615.

Dekkers, A., & Aarts, E. (1991). Global optimization and simulated annealing. *Mathematical Programming*, *50*(1), 367–393.

de Mourgues, M., Emde, C., & Mayer, B. (2023). Optimized wavelength sampling for thermal radiative transfer in numerical weather prediction models. *Atmosphere*, *14*(2), 332.

Diamond, S., & Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, *17*(83), 1–5.

Fu, Q., & Liou, K. (1992). On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of the Atmospheric Sciences*, *49*(22), 2139–2156.

Goody, R., West, R., Chen, L., & Crisp, D. (1989). The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *42*(6), 539–550.

Gordon, I., Rothman, L., Hargreaves, R., Hashemi, R., Karlovets, E., Skinner, F., ... others (2022). The HITRAN2020 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *277*, 107949.

Hogan, R. J., & Matricardi, M. (2020). Evaluating and improving the treatment of gases in radiation schemes: the Correlated K-Distribution Model Intercomparison Project (CKDMIP). *GMD*, *13*(12), 6501–6521.

Hogan, R. J., & Matricardi, M. (2022). A tool for generating fast k-distribution gas-optics models for weather and climate applications. *Journal of Advances in*

394 *Modeling Earth Systems*, e2022MS003033.

395 Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by simulated
396 annealing. *Science*, *220*(4598), 671–680.

397 Lacis, A. A., & Oinas, V. (1991). A description of the correlated k-distribution
398 method for modeling nongray gaseous absorption, thermal emission, and multi-
399 ple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical*
400 *Research: Atmospheres*, *96*(D5), 9027–9063.

401 Manabe, S., & Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere
402 with a given distribution of relative humidity.

403 Moncet, J.-L., Uymin, G., Lipton, A. E., & Snell, H. E. (2008). Infrared radiance
404 modeling by optimal spectral sampling. *Journal of the Atmospheric Sciences*,
405 *65*(12), 3917–3934.

406 Pincus, R., Buehler, S. A., Brath, M., Crevoisier, C., Jamil, O., Franklin Evans,
407 K., ... others (2020). Benchmark calculations of radiative forcing by
408 greenhouse gases. *Journal of Geophysical Research: Atmospheres*, *125*(23),
409 e2020JD033483.

410 Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency,
411 and flexibility in radiation calculations for dynamical models. *Journal of Ad-*
412 *vances in Modeling Earth Systems*, *11*(10), 3074–3089.

413 Yang, X.-S. (2014). Chapter 14 - multi-objective optimization. In X.-S. Yang (Ed.),
414 *Nature-inspired optimization algorithms* (p. 197-211). Oxford: Elsevier. doi:
415 https://doi.org/10.1016/B978-0-12-416743-8.00014-2

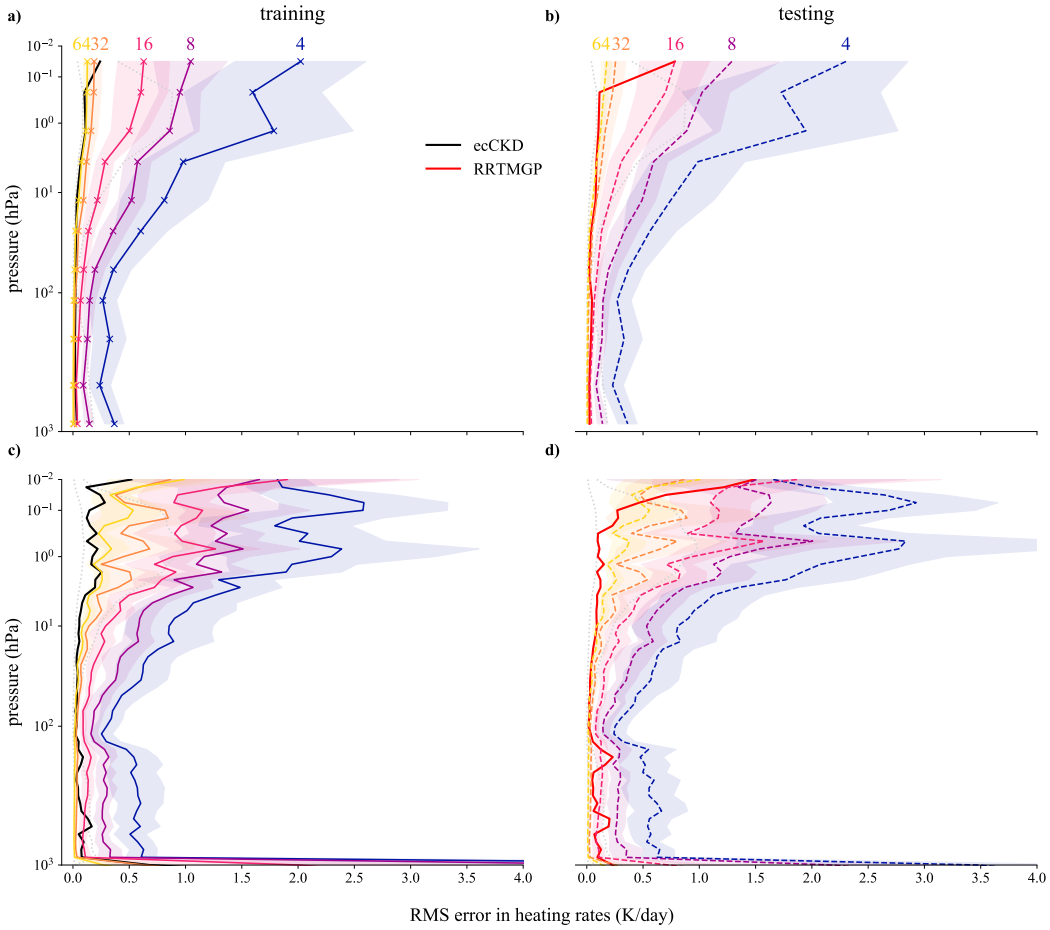

Figure 4. The absolute root mean squared error (in K/day) of the heating rate calculated from predicted fluxes throughout the atmosphere. Subset sizes are noted above the plot. Panel **a)** shows the RMSE of heating rates at only the pressure levels in the training set; panel **b)** shows the associated out-of-sample atmospheric columns; **c)** shows results for the training set using predicted fluxes for all model levels; panel **d)** shows absolute RMSE for out-of-sample atmospheric profiles for all levels. Envelopes denote the standard deviation in absolute RMSE across sixteen algorithm runs. The black and red lines show heating rates predicted by ecCKD and RRTMGP, respectively. 10% and 1% of the mean heating rate profile are plotted in grey dotted lines.

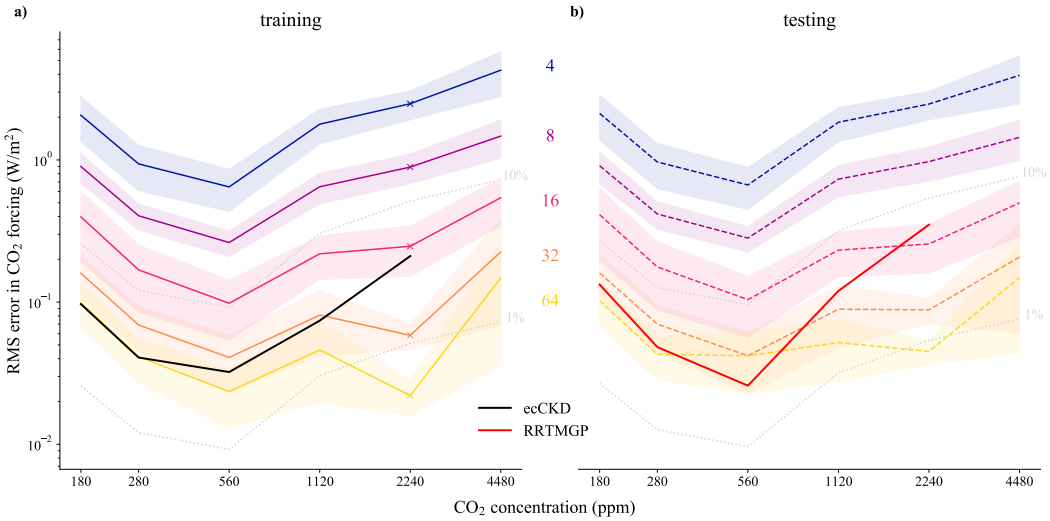

Figure 5. The root mean squared error (in units of W/m^2) of the forcing by CO_2 is plotted for different subset sizes across various climate scenarios. Panel **a)** shows the error from the training set, and testing error on panel **b)**. Crosses indicate that only forcing from the 2240 ppm training scenario is included in the cost function. The results from ecCKD and RRTMGP are shown in black and red, respectively. Clouds show the standard error across sixteen independent optimizations. 10% and 1% of the mean forcing for each scenario are plotted in grey dotted lines.

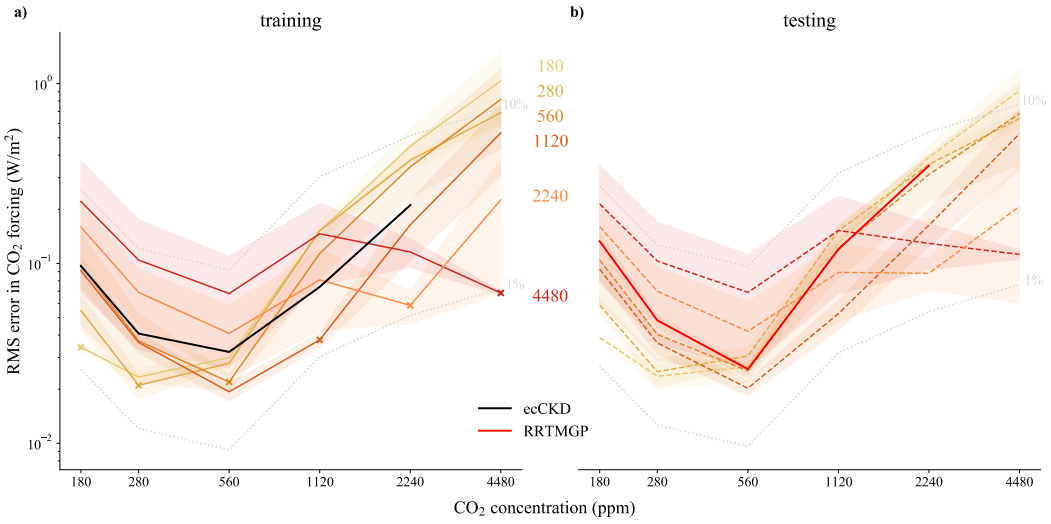

Figure 6. The root mean squared error (in units of W/m^2) of the forcing by CO_2 is plotted for a 32-wavenumber subset trained on the different climate scenarios: the last glacial maximum (LGM), pre-industrial concentrations (PI), and 2-, 4-, 8-, and 16x PI CO_2. Panel **a)** shows the error from the training set, and testing error on panel **b)**. Crosses indicate which training scenario is included in the cost function. The results from ecCKD and RRTMGP are shown in black and red, respectively. 10% and 1% of the mean forcing for each scenario are plotted in grey dotted lines. Clouds show the standard error across just three independent optimizations.

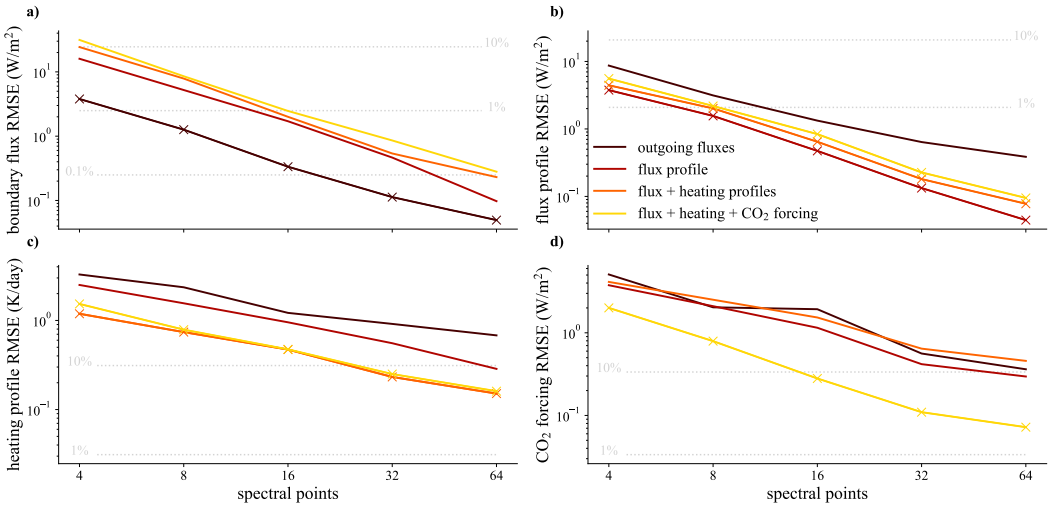

Figure 7. The average errors for models trained on various cost functions: one including only boundary fluxes (Eqn. 2), in dark red; one including every fifth net flux in the vertical, in red; one trained on these fluxes as well as corresponding heating rates (Eqn. 3), in orange; and one where the cost function encompasses fluxes, heating rates, and the CO_2 forcing between present day and 8x PI CO_2 (Eqn. 4), in yellow. The panels show: **a)** the mean RMSE for boundary fluxes; **b)** the mean RMSE for fluxes throughout the atmospheric column; **c)** the average absolute RMSE of heating rates throughout the atmospheric profile and **d)** the average RMSE of forcing by CO_2 across climate scenarios. The crosses indicate where the given quantity was included in the cost function. 10%, 1% and 0.1% of the mean reference values for each quantity are plotted in grey dotted lines.

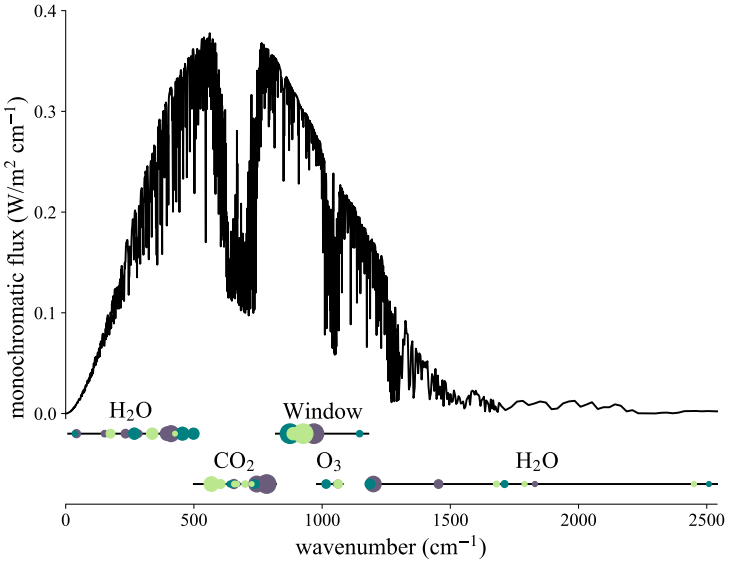

Figure 8. Below an example of a spectrum of OLR, major absorbers are noted. Dots mark where a wavenumber was chosen, and the size of the dots is proportional to the importance of that spectral point to the broadband flux integration. The colors correspond to three independent optimizations.

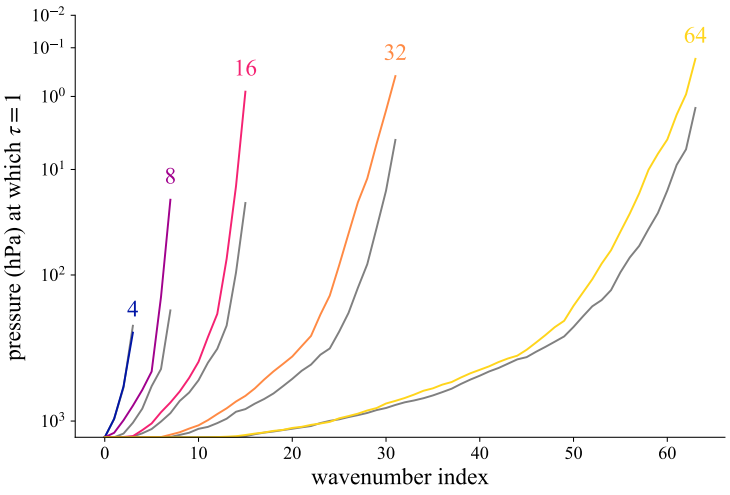

Figure 9. For each wavenumber in the final subset S, the pressure level at which that wavenumber first reaches optical depth $\tau = 1$ from the top of the atmosphere is plotted. The colorful lines correspond to models optimizing both fluxes and heating rates (Eqn. 3), and the accompanying grey lines correspond to models optimizing only fluxes.