Partial validation of a lossy compression approach to computing radiative transfer in cloud system-resolving models

Howard W. Barker 1 | Zhipeng Qu 1 | Varun Dhanraj 2 | Jason N. S. Cole 1

1 Environment and Climate Change Canada, Toronto, Ontario Canada
2 University of Waterloo, Waterloo, Ontario Canada

Correspondence
H. Barker, Environment and Climate Change Canada, 4950 Dufferin Street, Toronto, ON M3H 5T4, Canada.
Email: howard.barker@canada.ca

Abstract
Cloud system-resolving models (CSRM)s routinely calculate radiative flux profiles via the Independent Column Approximation (ICA). The ICA applies 1D radiative transfer models (RTMs) to all N columns in a CSRM’s domain. For this study, the Partitioned Gauss–Legendre Quadrature (PGLQ) method replaced the ICA in a CSRM. The PGLQ applies RTMs to \( n_G = N \) columns, identified with GLQ rules, and distributes their flux profiles to the other \( N - n_G \) columns. The PGLQ approach is likened to a lossy compression algorithm that trades information for efficiency. While verification and validation of an audio compression algorithm rest, respectively, on file size reduction and sound quality according to listeners, for the PGLQ they rest on increasing \( f_{RT} = N/n_G \) while maintaining, according to experimenters, the integrity of CSRM simulations. A CSRM was run for 80 days in radiative-convective equilibrium (1,024 × 1,024 columns and horizontal grid-spacing of 0.25 km) for sea-surface temperature \( SST = 295 \) and 300 K; the last 40 days were analysed. Simulations using the ICA represent the control; experiments used the PGLQ calling the RTMs \( f_{RT} = 200 \), 5,000 and 50,000 fewer times than the control. For \( f_{RT} = 50,000 \), several key variables spanning time/domain-averaged cloud and radiation properties, a measure of cloud (convection) aggregation, and horizontal fluctuations of cloud and radiation fields, differ significantly from the control. In contrast, corresponding differences between the control and PGLQ with \( f_{RT} \leq 5,000 \), for both a given \( SST \) and differences between \( SSTs \), are often minor, in all respects, and appear to be drawn from a single population. While these results partially validate the PGLQ method for \( f_{RT} \leq 5,000 \), they also indicate that overly large reductions in radiative information are detrimental.

Keywords
3. physical phenomenon: clouds, 3. physical phenomenon: radiation, 4. geophysical sphere: atmosphere
1 INTRODUCTION

Limited-area numerical weather prediction models, cloud system-resolving models (CSRMs), and large-eddy simulations simulate cloudy atmospheres for domains that often exceed $10^4$ km$^2$ with horizontal grid-spacings $\Delta x \leq 1$ km (Stevens et al., 2002; Bretherton and Khairoutdinov, 2015; Milbrandt et al., 2016). With atmospheres partitioned into numerous layers, these simulations often contain over $10^8$ cells, and require substantial computing resources to run efficiently. Computation of atmospheric radiative transfer (RT) is particularly taxing and achieved, almost always, by the Independent Column Approximation (ICA) in which one-dimensional (1D) RT models get applied to every column (e.g. Barker et al., 2017). Executing the ICA can become burdensome when the number of cells is large. To lighten the load, modellers often allow several dynamical time steps to pass between RT calls. Still, however, the RT models account for 10–40% of CSRMs’ total computing times (Manners et al., 2009; S. Krueger, M. Khairoutdinov, personal communication, 2019).

Barker and Li (2019) developed an algorithm that aims to reduce the ICA’s computational burden whilst minimally impacting the overall character of CSRMs simulations. Their method, which still employs 1D RT models, begins by partitioning columns according to whether they contain clouds or not. Cloudy columns are further partitioned into classes based on cloud-top altitude, since that is where most radiative heating and cooling occurs. Then, based on a variable that is important for RT, each class’s columns are sorted from smallest to largest. The initial version sorts cloudy and cloudless columns according to cloud water path and precipitable water, respectively. Then, for each sorted sequence containing $N$ columns, identify $n_G \ll N$ of them according to Gauss–Legendre Quadrature (GLQ) rules (Li and Barker, 2018), apply 1D RT models to the $n_G$ columns, allocate the resulting flux profiles to neighbouring columns in the sorted sequences, and finally return the profiles to their proper locations in the CSRM domain. Hereinafter, this algorithm is referred to as Partitioned Gauss–Legendre Quadrature (PGLQ).

As PGLQ calculates $n_G \ll N$ flux profiles in order to represent all $N$, it could be referred to as a lossy computation algorithm, given its resemblance to lossy compression algorithms that encode data files into relatively small, but usable, files for ease of transfer and storage (e.g. the ~10:1 compression of audio WAV files into MP3). Developers and users of lossy compression algorithms subjectively weigh algorithmic losses of information against the benefits of having small files. By analogy, the encoder portion of PGLQ involves partitioning, sorting and identification of $n_G \ll N$ columns. The RT models, which transform physical information into radiative information, are akin to transferring files. The decoder portion of PGLQ, which usually only applies to lossless compression algorithms, involves populating all $N$ columns with the $n_G$ flux profiles. In contrast, the ICA is analogous to transfer and storage of uncompressed files. See Figure 1 for a summary.

Validation of a lossy compression algorithm is not straightforward as it entails pitting the subjectively established benefits of dealing with smaller files against the willingness of users to accept errors incurred in the encoding process. For instance, if most listeners were satisfied with the sound quality of relatively small MP3 files, despite quantifiable errors relative to their larger WAV progenitors, the algorithm’s developers would consider their product to be validated (i.e. its intended purpose realized). Hence, analyses of PGLQ by Barker and Li (2019) and Barker et al. (2019) amounted only to verification of the methodology; they assessed how well it reproduces ICA radiative fluxes for CSRM cloudy atmospheres.

Validation of PGLQ requires CSRM simulations that use the ICA and PGLQ interactively. Assessing differences between these simulations is tantamount to listeners’ subjective judgment of reduced size and quality of MP3 files. For PGLQ to be validated, at some $f_{RT} = N/n_G$, not only must the underlying motivation to reduce RT computation time be realized satisfactorily, but CSRM simulations using the ICA and PGLQ must display sufficiently small differences in order that user-defined experimental outcomes do not differ (e.g. Cole et al., 2005). This means satisfying

$$1 \ll \frac{CPU(T(x_N))}{CPU(T(x_{N})) + CPU(T(x_{n_G}))} + CPU(D(F_{n_G}))$$

$$\approx \frac{CPU(T(x_N))}{CPU(T(x_{n_G}))} - \epsilon,$$

where $CPU(p)$ represents CPU time required for process $p$ (see Figure 1 for an explanation), and $\epsilon$ being acceptably small according to the experimenter. On the contrary, should use of PGLQ, at some $f_{RT}$, have insignificant impacts on CSRM simulations, yet fail to satisfactorily reduce $CPU(T)$ relative to the ICA, or vice versa, that would amount to its invalidation.

The objective of this study was to begin validation of the PGLQ algorithm. This was achieved by interactive applications of both the ICA (control) and the PGLQ (experiment) methodologies in the System for Atmospheric Modeling (Khairoutdinov and Randall, 2006) CSRM, and assessing differences in predicted cloud and radiation properties. Conditions of radiative-convective equilibrium were used with sea-surface temperatures SST of 295 and 300 K (Wing et al., 2018).

The following section reiterates the PGLQ algorithm. It is followed by discussion of the CSRM and the
experiments. The fourth section presents results, and the fifth provides a summary.

2 PARTITIONED GAUSS–LEGENDRE QUADRATURE (PGLQ) ALGORITHM

As the PGLQ methodology was proposed only recently, it is recapped briefly here. To begin, it is assumed that PGLQ gets applied once to a CSRM’s entire domain, which consists of N columns and has vertically projected cloud fraction $A_c$. The initial stage of PGLQ partitions the $N_{cd}=A_cN$ cloudy columns into $M \geq 1$ classes. Thus far, cloud classes have been defined for ranges of cloud-top altitudes $h_{cd}$ (cf. section 3.2 of Barker and Li, 2019) with each class having near-constant $N_m$ cloudy columns (cf. orthogonal sampling: Owen, 1992). Letting $F(n)$ be a broadband flux profile for the $n$th column, domain-averaged ICA flux can be expressed as

$\langle F \rangle = \frac{1}{N} \sum_{n=1}^{N} F(n) = (1 - A_c) \left[ \frac{1}{N - N_{cd}} \sum_{n=1}^{N - N_{cd}} F_{cd}(n) \right] + A_c \left[ \frac{1}{N_{cd}} \sum_{n=1}^{N_{cd}} F_{cd}(n) \right]$

where $\langle F_{cd} \rangle$ is mean cloudy-sky flux profile, and $\langle F_{cd}(n) \rangle$ is mean cloudy-sky flux profile for the $m$th cloud class. Cloudy columns within a class are then sorted from smallest to largest cloud water path $W$ (Li and Barker, 2018) and sequentially numbered as

$x_l = \frac{(i^* - 1)}{N_m}$,

with $x_{l^*} = 0$ and $x_{l^*} = N_m = 1$ corresponding to min{$W$} and max{$W$}, respectively. Now, mean flux (profile) for the $m$th class $\langle F_{cd}(n) \rangle$ can be approximated by GLQ of order $N_G = N_m$ as

$\langle F_{cd}(n) \rangle = \frac{1}{N_m} \sum_{n=1}^{N_m} F_{cd}(n|m)$

with RT models applied to $N_G$ profiles at $(x_i + 1)/2$ (Abramowitz and Stegun, 1972). Finally, ranges of $i^*$ either side of a GLQ profile receive its $F_{cd}(x_i + 1)/2$ (see equation 6 of Barker and Li, 2019).

Cloudless columns were treated as just described, except that precipitable water, rather than $W$, is used to sort the $(1 - A_c)N$ cloudless columns. For this study, the number of GLQ points allocated for cloudless columns was $n_{clr} = 1$. This extreme setting differs from Barker and Li’s (2019) use of $n_{clr} \geq 2$, but for RCE with uniform $SST$
and \( A_c \gg 0 \), use of a single column to represent all cloudless columns generated little error (only slightly less than \( n_{clr} = 30 \)). This is because distributions of cloudless-sky profiles of temperature and water vapour are narrow and near-symmetric, with almost identical mean and median values; the latter being used by PGLQ when \( n_{clr} = 1 \).

When using the PGLQ method, it is straightforward to specify, at the outset, the ratio \( f_{RT} \) between numbers of calls to cloudy RT codes made by the ICA and PGLQ. This governs the magnitudes of the PGLQ’s (largely random) errors and computational saving compared to the ICA. It is expected that \( f_{RT} \gg 1 \). As described above, the version of the PGLQ method used here makes \( M \cdot n_G \) calls to cloudy-sky RT models, and so for each cloud class

\[
n_G = \frac{N_{cl}d}{M \cdot f_{RT}}.
\]

Correspondingly, for cloudless columns with \( n_{clr} = 1 \), the ratio of ICA-to-PGLQ calls to clear-sky RT models is simply \((1 - A_c)N\), which can be enormous when \( N \) is large; even for large \( A_c \).

### 3 | EXPERIMENTAL DESIGN

To realize this study’s objective, PGLQ was implemented in the System for Atmospheric Modeling (SAM v6.11) CSRM (e.g. Khairoutdinov and Randall, 2006; Khairoutdinov and Emanuel, 2013). Only conditions of radiative-convective equilibrium (RCE) were considered. Numerical simulation of RCE has a long history (e.g. Manabe and Strickler, 1964; Ramanathan and Coakley, 1978; Bretherton et al., 2005) that rests on the argument that RCE captures some essential features of Earth’s climate system, and provides a tractable and responsive way to study radiative sensitivities and cloud feedback processes (e.g. Bony et al., 2015; Mauritsen and Stevens, 2015).

Boundary conditions followed Wing et al. (2018). This included cyclic horizontal boundary conditions over uniform ocean with \( SST = 295 \) K for 200 days for a domain of \( 128 \times 128 \) columns at horizontal grid-spacing \( \Delta x = 1 \) km. To refine RCE conditions, the simulation was restarted for an additional 30-day simulation using \( 1024 \times 1024 \) columns and \( \Delta x = 1 \) km. This simulation’s term provided initial conditions for multiple simulations using \( 1024 \times 1024 \) columns and \( \Delta x = 0.25 \) km; each lasted for 80 days, with the final 40 used in the analyses. Dynamical time step was 4 s and SAM ran on 256 parallel-processors. The control simulation used the ICA (SAM’s default). PGLQ experimental simulations used \( f_{RT} = 200 \), 5,000 and 50,000. This arbitrary range of \( f_{RT} \) spans: very small PGLQ errors while saving, by most standards, much computation time (i.e. \( f_{RT} = 200 \)); to absurdly large reductions in computation time (i.e. \( f_{RT} = 50,000 \)) with enormous localized flux errors for all but the most stratiform cloud fields.

Restarting off the 200-day \( SST = 295 \) K run mentioned above, the ICA and all PGLQ simulations were repeated using \( SST = 300 \) K. In order to equilibrate, they ran for 160 days with the final 40 days used for analyses. This affords the ability to establish if PGLQ maintains the ICA’s radiative feedbacks when going from \( SST = 295 \) to 300 K.

For the simulations done here, typical values of \( A_c \) were \(-0.8\), meaning \( N_{cl}d \approx 840,000 \). As the number of cloud-top altitude classes \( M \), settled on by the partitioning algorithm (see equation 7 in Barker and Li, 2019), was typically 9 for \( SST = 295 \) K and 12 for \( SST = 300 \) K, each cloud class used \( n_G \approx 420, 168 \) and 17 for \( f_{RT} = 200 \), 5,000 and 50,000, respectively. As documented by Barker and Li (2019) (their figures 4 and 7), even \( n_G \approx 17 \) per \( h_{cl}d \) subdomain yields excellent estimates of subdomain average fluxes \( \langle F_{cl}d \rangle_m \), though localized errors can be large. Table 1 summarizes the simulations. Following Wing et al.’s (2018) tables 5 and 7, 2D arrays are hourly averages, and 3D arrays are instantaneous samples saved every 6 hr.

As an alternative, full domains can be partitioned into \( N_{CPU} \) subdomains and the PGLQ applied to each independently (e.g. a subdomain per processor) (Barker and Li, 2019). Often with this approach, however, radiation fields exhibit weak discontinuities at subdomain boundaries, especially LW fluxes, because sorting on \( W \) and GLQ-point selection are best suited to SW RT. While simulations using this approach were performed for \( N_{CPU} = 256 \), results are not shown as they resembled very closely those applied to the entire domain (i.e. \( N_{CPU} = 1 \)).

### 4 | RESULTS

Results include straightforward comparisons between temporal/spatial-integrated mean values of both
### TABLE 1 Summary of CSRM experiments conducted for SST = 295 and 300 K

| Simulation    | Description                                                                 |
|---------------|-----------------------------------------------------------------------------|
| Control       | SAM’s default configuration in which RRTM is applied to all 1024 × 1024 columns of a domain. |
| Experiment 1  | For the (1024 × 1024) Ac, cloudy columns, where Ac is total cloud fraction, PGLQ makes \( f_{RT} = 200 \) times fewer calls to RRTM than in the control experiments. The remaining cloudless columns use PGLQ with \( n_{fl} = 1 \). |
| Experiment 2  | As in Experiment 1 except \( f_{RT} = 5,000 \).                             |
| Experiment 3  | As in Experiment 1 except \( f_{RT} = 50,000 \).                           |

### TABLE 2 Temporal–spatial integral values of \( T_{sfc} \), \( p_w \), \( P_{sfc} \), and fractional area of surface precipitation for ICA and PGLQ simulations as functions of SST

| Variable                          | Screen | SST (K) | ICA | \( f_{RT} = 200 \) | \( f_{RT} = 5,000 \) | \( f_{RT} = 50,000 \) |
|-----------------------------------|--------|---------|-----|---------------------|----------------------|-----------------------|
| Near-surface air temperature \( T_{sfc} \) (K) | none   | 295     | 291.51 | 291.48 | 291.47 | 291.53 |
|                                   |        | 300     | 296.70 | 296.64 | 296.65 | 296.73 |
|                                   |        | 300–295 | 5.19  | 5.17  | 5.18  | 5.20   |
| Precipitable water \( p_w \) (mm) | none   | 295     | 25.75 | 25.59 | 25.62 | 25.89 |
|                                   |        | 300     | 42.38 | 42.07 | 42.13 | 42.21 |
|                                   |        | 300–295 | 16.63 | 16.48 | 16.51 | 16.32 |
| Surface precipitation rate \( P_{sfc} \) (mm-day\(^{-1}\)) | none   | 295     | 2.254 | 2.28  | 2.27  | 2.24  |
|                                   |        | 300     | 2.594 | 2.63  | 2.62  | 2.62  |
|                                   |        | 300–295 | 0.340 | 0.35  | 0.35  | 0.38  |
|                                   | \( P_{sfc} > 0.0 \) | 295     | 9.86  | 10.12 | 10.05 | 10.05 |
|                                   |        | 300     | 8.45  | 8.61  | 8.54  | 9.39  |
|                                   |        | 300–295 | −1.41 | −1.51 | −1.51 | −0.66 |
| Fractional area of surface precipitation | none   | 295     | 0.23  | 0.23  | 0.23  | 0.22  |
|                                   |        | 300     | 0.31  | 0.31  | 0.31  | 0.28  |
|                                   |        | 300–295 | 0.08  | 0.08  | 0.08  | 0.06  |

Note: Bold-face numbers indicate that PGLQ values, for three values of \( f_{RT} \), are statistically indistinguishable from ICA values, at the 95% confidence level.

vertically integrated and single-layer fields as simulated by control (ICA) and experiment (PGLQ) simulations. Likewise, horizontal structure of cloud and radiation fields, with an emphasis on convective aggregation and structure function analysis, were considered, too.

When validating PGLQ within SAM, the simplest assessments involve temporal integrals of domain-average fields. In lieu of performing expensive multi-member ensemble experiments, single simulations (as described in Section 3) were performed. This raises the issue of assessing the statistical significance of differences between simulations. The approach used here to perform standard tests of null hypotheses \( H_0 \), at a defined confidence level (c.l.), is discussed in the Appendix.

4.1 | Basic variables of state

Two basic atmospheric variables are near-surface temperature \( T_{sfc} \) and vertically integrated precipitable water \( p_w \). Table 2 shows that there are almost no differences between all PGLQ simulations and their ICA counterparts; not too surprising given fixed SST. The same goes for their associated profiles (not shown). Note, however, that despite the tiny differences in mean \( p_w \), all of which are <1%, \( H_0 \) was never accepted. Obviously this is due to the temporal stability of \( p_w \), but it does raise the question of what an experimenter recognizes as a noteworthy impact. On the contrary, differences between PGLQ and ICA surface precipitation rates \( P_{sfc} \) are almost entirely negligible despite differences often being just 1 to 2%; the exception being \( f_{RT} = 50,000 \) for SST = 300 K,
whose overestimate, relative to ICA, is \(\sim 11\%\). Similarly, fractional areas of domains with \(P_{sfc} > 0\) are replicated almost perfectly by the PGLQ simulations; especially for \(f_{RT} \leq 5,000\).

### Cloud properties

Table 3 lists values, and their differences, pertaining to cloud properties. For \(SST = 295\) K, ICA values of total cloud fraction \(A_c\) are overestimated significantly at the 95\% c.l. by PGLQ values; the more so the larger \(f_{RT}\). For \(SST = 300\) K, however, \(A_c\) for all \(f_{RT}\) agree extremely well with the ICA’s. Figure 2 shows corresponding mean profiles of layer cloud fractions \(c(z)\). For \(f_{RT} = 200\) and 5,000 at \(SST = 295\) K, three layers near \(z = 12\) km show \(c(z)\) exceeding the ICA’s by \(\sim 0.02\). This in itself could account for their overestimations of \(A_c\) listed in Table 3. At all other altitudes, agreements are excellent. At \(SST = 300\) K, however, most \(c(z)\) agree extremely well with the ICA’s, but those that do not are slightly small, yet \(A_c\), as listed in Table 3, are fine.

On the contrary, for the extreme case of \(f_{RT} = 50,000\), \(c(z)\) exceed the ICA’s by 0.05 to 0.15 for several layers, yet their \(A_c\) at \(SST = 300\) K match well. To explain this, consider how clouds overlap vertically. Assume that total fractional cloud cover for two layers at altitudes \(z_1\) and \(z_2\) can be explained by the maximum/random model as

\[
c_{12} = \alpha_{12} \max [c(z_1), c(z_2)] + (1 - \alpha_{12}) [c(z_1) + c(z_2) - c(z_1) c(z_2)],
\]

where

\[
\alpha_{12} = \frac{c(z_1) c(z_2)}{c(z_1) + c(z_2) - c(z_1) c(z_2)}.
\]
where

\[ a_{12} = \exp \left[ -\int_{z_i}^{z_0} \frac{dz}{L_c(z)} \right] \]  

and \( L_c(z) \) is overlap decorrelation length (e.g. Hogan and Illingworth, 2000). Following Barker and Räisänen (2005), define the columnar effective overlap decorrelation length \( L'_c \) as the root of

\[ A'_c(L'_c) - A_c = 0, \]  

where \( A_c \) is actual total cloud fraction, and \( A'_c \) is cloud fraction produced by Räisänen et al’s (2004) stochastic cloudy-column generator operating on \( c(z) \) with vertically constant \( L_c \).

Figure 3 shows cumulative distributions of \( L'_c \) for the 6-hourly-saved 3D arrays of cloud water content CWC. Distributions of \( L'_c \) for \( f_{RT} = 200 \) and 5,000 differ very little from the ICA’s; differences between their median \( L'_c \) are <0.2 km, which are subgrid-scale for, as Figure 2 shows, most clouds occur between 5 and 15 km altitude where layer thicknesses are 0.5 km. There are, however, notable deviations from the ICA’s distributions of \( L'_c \) for \( f_{RT} = 50,000 \), the latter exhibiting overly large dispersions for both values of SST. The tendency for \( f_{RT} = 50,000 \) at \( SST = 300 \) K to have too many large values of \( L'_c \) explains why its mean \( A_c \) agrees fairly well with the ICA’s (see Table 3) despite this rendition of PGLQ overestimating \( c(z) \) for several layers (Figure 2); as \( L'_c \) increases, for a given \( c(z) \), \( A_c \) decreases. Likewise, \( f_{RT} = 50,000 \) at \( SST = 295 \) K has too many small \( L'_c \) relative to the ICA, which when coupled with several excessive values of \( c(z) \) results in substantial overestimation of mean \( A_c \) by the PGLQ.

As a final point regarding cloud fractions, all PGLQ experiments underestimate total cloud fraction sensitivities \( \partial A_c / \partial SST \approx [A_c(300 \) K) – \( A_c(295 \) K)]/5 by 15% to 40% regardless of whether all cells, or only those with \( W > 0.001 \) mm, are used to define \( A_c \). These seem to be both large and deserving of more testing with a wider range of SST values.

Following cloud fraction, the most influential variable governing RT for cloudy atmospheres is cloud liquid \( L \) and ice \( I \) water paths (\( L + I = W \)), and their corresponding profiles of CWC. Table 3 indicates that the main outlier is unscreened \( W \) (i.e. when cloudless columns are included in the average) for \( f_{RT} = 50,000 \); Figure 4 shows the culprit to be ice clouds above ~8 km (where temperatures are <230 K). For \( f_{RT} \leq 5,000 \), Figure 4 shows that profiles of means and standard deviations of CWC agree almost perfectly with those for the ICA simulations, while Table 3 shows that corresponding \( W \) do equally well, too. As is also clear from Table 3, approximate sensitivities \( \partial W / \partial SST \) for all \( f_{RT} \) are essentially indistinguishable from the ICA’s.

The last entries in Table 3 are mean cloud-top altitude \( z_{top} \) and temperature \( T_{top} \) for columns with \( W > 0.001 \) mm. Generally, PGLQ estimates of \( z_{top} \) are systematically less than the ICA’s by no more than 0.15 km. As with \( L'_c \), these differences are subgrid-scale. Accordingly, all PGLQ estimates of \( T_{top} \) overestimate ICA values for a given \( SST \), by as much as 5 K for \( f_{RT} = 50,000 \) at \( SST = 300 \) K.

### 4.3 Radiative fluxes

Table 4 lists time-domain averages for TOA and surface net radiative fluxes. At \( SST = 295 \) K, differences between ICA and \( f_{RT} = 200 \) and 5,000 are insignificant for almost all fluxes. The exception is LW at surface, which is a very stable field in RCE conditions; hence, differences of only \(~0.5 \) W m\(^{-2}\) or <1% register as being significant at the 95% c.l. At \( SST = 300 \) K, on the other hand, all domain-average flux fields vary weaker in time than at 295 K, and so even minor mean differences of less than 1% often lead to rejections of \( H_0 \). Screening to include only cloudy columns with...
(a,b) Time/domain-averaged profiles of layer mean CWC (cells with CWC > 0) for the ICA and PGLQ simulations for three values of $f_{RT}$ at SST = 295 K and 300 K. (c,d) Same as (a,b) except these are standard deviations of CWC.

$W > 0.001$ mm does little to change the acceptance rate of $H_0$.

The most notable errors listed in Table 4 are for $f_{RT} = 50,000$ where only half of the entries have $H_0$ accepted at the 95% c.l. The poorest performances are for columns with $W > 0.001$ mm where overestimates relative to the ICA reach 10 W·m$^{-2}$, or $\sim 5\%$, for SW at surface. This is because as $f_{RT}$ decreases, and numbers of GLQ points decrease, tails of the (sorted) distributions of fluxes, that come about via sorted $W$, can be quite steep, and thus substantially under-represented by GLQ (Li and Barker, 2018; Barker and Li, 2019). This is made clear in Figure 5 which shows distributions of domain-wide minima of net SW and LW fluxes at TOA. These quantities correspond to the two columns in each domain that reflect most and emit least. For $f_{RT} = 200$ (i.e. large $n_G$), all distributions agree nicely with those for the ICA simulations. By $f_{RT} = 5,000$, however, these extremes become difficult to capture, and would be so even if clouds were simulated perfectly. The modes of the outgoing LW radiation (OLR) distributions for SST = 300 K correspond to effective radiating temperatures of 196 K for ICA and 199 K for $f_{RT} = 5,000$, which, in turn, corresponds to altitudes of $\sim 13.7$ and 13.0 km, respectively.

The inset plots in Figure 5 show that for $f_{RT} = 50,000$ the situation worsens substantially. At SST = 300 K the mode of the OLR distribution corresponds to 210 K, or $\sim 12$ km. For net SW at TOA the impact is even more dramatic; while modes of the distributions for ICA, $f_{RT} = 200$ and $f_{RT} = 5,000$, correspond to TOA albedos of $\sim 0.8$, for $f_{RT} = 50,000$ they correspond to albedos of only $\sim 0.6$. Again, this is a limitation of PGLQ when using small $n_G$. Fortunately, however, impacts on simulated dynamics and clouds are squelched much, as shown in Figures 2 and 4.

Figure 6 shows time/domain-averaged all-sky radiative heating rate (HR) profiles for SST = 300 K (plots for 295 K are very similar and not shown). In concert with prior results, both SW and LW profiles for $f_{RT} = 200$ and 5,000 agree extremely well with the ICA profiles, save for minor underestimates in SW HR between $\sim$ 8 to 11 km on account of slightly too few clouds with tops exposed to direct irradiance (Figure 2b). As expected by now, there are obvious problems with $f_{RT} = 50,000$ as SW and LW HR biases align with errors in $c(z)$ (cf. Figure 2b and 5).

Figure 7a,c show time/domain-averaged HR profiles for the cloudy-sky components of all-sky profiles shown in Figure 6. Generally, the absolute value of PGLQ values are, often much, smaller than corresponding ICA values for both SW and LW. For the cloudiest layers between $\sim 9$ and 14 km, differences are not too severe, but for liquid clouds below $\sim 5$ km, they are pronounced; but as Figure 7b shows, $c(z)$ at these altitudes are $<0.03$ (i.e. sparse Cu), and so all-sky HRs are fine. As Figure 4 shows, many of these low clouds have high densities and thus large $W$. Therefore, if and when clouds at these altitudes have little cloud above them, they have the potential to either absorb much SW direct-beam radiation or emit LW radiation freely to space. With $c(z)$ being so small, the columns satisfying these conditions can be missed by PGLQ (Figure 5); the upshot being too many small values of HR, associated with clouds below 5 km with much cloud above them, get replicated in columns that should have large HRs at their exposed tops.

Related directly to this is PGLQ’s suppression of variations in horizontal fluctuations of HR as $f_{RT}$ increases. Figure 7b,d show profiles of time-averaged standard deviations of cloudy-sky SW and LW HR for SST = 300 K. For $f_{RT} = 200$ above 5 km, variability of HRs is extremely good. Performance slips for $f_{RT} = 5,000$, and becomes seriously flawed for $f_{RT} = 50,000$. Below 5 km, all PGLQ simulations show underestimated variances; the more so as $f_{RT}$ increases and GLQ sampling degrades (Figure 5; Barker and Li, 2019).

4.4 Convective aggregation

Convective aggregation has received much attention recently because of its possible co-dependence with climatic change (e.g. Muller and Bony, 2015; Holloway
Table 4  As in Table 2 except these are for several different radiation variables

| Variable            | Screen | SST (K) | ICA  | $f_{RT} = 200$ | $f_{RT} = 5,000$ | $f_{RT} = 50,000$ |
|---------------------|--------|---------|------|---------------|-----------------|------------------|
| Net SW at TOA (W·m$^{-2}$) | None   | 295     | 346.76 | 348.04        | 347.65          | 349.38           |
|                     |        | 300     | 341.91 | 343.56        | 343.189         | 346.19           |
|                     |        | 300–295 | −4.85  | −4.48         | −4.46           | −3.20            |
| $W > 0.001$ mm      | 295    | 317.77  | 318.44 | 317.69        | 327.20          |                  |
|                     | 300    | 320.07  | 320.88 | 320.80        | 329.57          |                  |
|                     | 300–295| 2.30    | 2.44   | 3.11          | 2.38            |                  |
| Net LW at TOA (W·m$^{-2}$) | None   | 295     | 236.59 | 238.19        | 238.15          | 236.27           |
|                     |        | 300     | 236.24 | 238.37        | 238.29          | 237.03           |
|                     |        | 300–295 | −0.35  | 0.18          | 0.14            | 0.77             |
| $W > 0.001$ mm      | 295    | 213.72  | 215.37 | 215.29        | 217.57          |                  |
|                     | 300    | 218.40  | 220.46 | 220.79        | 222.47          |                  |
|                     | 300–295| 4.68    | 5.09   | 5.34          | 4.90            |                  |
| Net SW at SFC (W·m$^{-2}$) | None   | 295     | 266.74 | 268.08        | 267.67          | 269.15           |
|                     |        | 300     | 253.75 | 255.44        | 255.04          | 257.86           |
|                     |        | 300–295 | −12.99 | −12.63        | −12.63          | −11.30           |
| $W > 0.001$ mm      | 295    | 237.04  | 237.78 | 237.00        | 246.29          |                  |
|                     | 300    | 231.90  | 232.75 | 232.61        | 241.09          |                  |
|                     | 300–295| −5.15   | −5.04  | −4.39         | −5.21           |                  |
| Net LW at SFC (W·m$^{-2}$) | None   | 295     | 83.67  | 84.14         | 84.10           | 83.84            |
|                     |        | 300     | 65.92  | 66.66         | 66.52           | 66.64            |
|                     |        | 300–295 | −17.74 | −17.49        | −17.58          | −17.21           |
| $W > 0.001$ mm      | 295    | 77.39   | 77.89  | 77.78         | 79.09           |                  |
|                     | 300    | 62.61   | 63.25  | 63.14         | 64.07           |                  |
|                     | 300–295| −14.78  | −14.64 | −14.64        | −15.03          |                  |

and Woolnough, 2016; Wing et al., 2017). It is not clear if convective systems self-organize in the real world as, and for the same reasons, they do in models. While the simulations done here were configured to avoid extreme, and visually obvious, aggregation (e.g. figure 1b in Wing et al., 2017), it was present. The question addressed in this subsection is: does replacing the ICA with PGLQ impact convective aggregation?

To establish the impact of PGLQ on cloud/convection aggregation, a satisfactory metric is needed. Tompkins and Semie’s (2017) is based on distribution of nearest neighbour distances between “convective entities”, but a different measure was used here. While any of a number of variables could be used as an “observable proxy” for convection, use was made of $W$. First, using all columns in a domain, sort $W$, from smallest to largest, set a value $p$, and identify $W^*$ such that

$$Pr[W \geq W^*] = p. \quad (9)$$

Next, define a binary-threshold version of $W$ as

$$W'(i,j) = \begin{cases} 1 ; & W(i,j) > W^* \\ 0 ; & W(i,j) \leq W^* \end{cases}. \quad (10)$$

Then, in a manner akin to degrading resolution and box-filling, divide the full domain, of $N$ columns, into $N/2^n$ equal-area sub-regions $S_n$, each measuring $(2^n \cdot \Delta x) \times (2^n \cdot \Delta x)$ km. If there exists at least one instance of $W'(i,j) = 1$ in the $m$th sub-region $S_n(m)$, simply count $S_n(m)$ as housing a convective entity, and compute the scale-dependent fractional area $C(2^n \cdot \Delta x)$ of sub-regions containing at least one such entity. This is expressed symbolically as

$$C(2^n \cdot \Delta x) = \frac{2^n N/2^n}{N} \sum_{m=1}^{N} \Phi[\exists W'(i,j) = 1 \in S_n(m)]. \quad (11)$$
Figure 5  (a) Normalized density functions of domain-wide minimum outgoing LW radiation (OLR) for \( \text{SST} = 295 \text{ K} \) simulations using the ICA and PGLQ with \( f_{RT} = 200 \) and 5,000. Inset log-linear plot shows the same but for \( f_{RT} = 50,000 \). (b) As in (a) except this is for simulations using \( \text{SST} = 300 \text{ K} \). (c,d) are the same as (a,b) except they show density functions for net SW radiation at TOA; the smaller the value the more reflective the clouds.

\[
\Phi(c) = \begin{cases} 
1; & \text{c is true} \\
0; & \text{c is false}.
\end{cases}
\]  

(12)

As expected with box-filling, \( C(2^n \cdot \Delta x) \) increases monotonically with \( n \). A sequence of \( C(2^n \cdot \Delta x) \) can be computed, but for this study, only the value at \( n = 7 \), which corresponds to a resolution of 32 km, was analysed. Thus, while all domains begin at \( C(0.25 \text{ km}) = p \), those that exhibit highly aggregated \( W \) tend to fill boxes slower than those with more scattered \( W \), and so \( C(32 \text{ km}) \) will tend to decrease with enhancing aggregation (at scales <32 km). The smaller \( p \), the more the routine isolates and works with the most massive clouds and, ostensibly, areas of most vigorous convection. Conversely, if \( p \) is set too large, it begins with increasingly “routine” clouds and not so well-defined areas of intense convection, and often yields many instances of \( C(32 \text{ km}) = 1 \) and a useless metric. For this study, \( p = .01 \).

Figure 8 shows sample domains having typical and extreme values of \( C(32 \text{ km}) \). As evident from the leftmost columns of \( W \), going from \( C(32 \text{ km}) > 0.9 \) to \( C(32 \text{ km}) < 0.2 \) covers a wide range of aggregation. While their associated fields of SW and LW radiation at TOA exhibit differences in clustering that are visually obvious, ranges of \( C(32 \text{ km}) \) for these fields are less dynamic than those for \( W \).

Figure 9 shows cumulative frequency distributions of \( C(32 \text{ km}) \) derived from 960 domains of \( W \) for ICA simulations at \( \text{SST} = 295 \) and 300 K. While all domains started with \( C(0.25 \text{ km}) = 0.01 \), median values of \( C(32 \text{ km}) \) are 0.52 and 0.84 for 295 and 300 K, respectively. As
experiments. While distributions for \( f_{RT} = 200 \) and 5,000 resemble closely the ICA’s (i.e. \( D_{n,n} < 0.062 \)), it is clear from Figure 9 and Table 5 that distributions of C(32 km) for \( f_{RT} = 50,000 \) do not resemble the ICA’s. It is not at all obvious why the PGLQ’s aggregation bias should reverse sign with \( SST \). Likewise, why, for \( SST = 300 \text{ K} \), when PGLQ is pushed to the extreme should aggregation become so enhanced? Presumably this has to do with PGLQ’s severe squelching of radiative heating/cooling differentials between very high cloud tops right above intense convective cores and the large areas of nearby slightly lower (detrained?) cloud?

4.5 Structure function analyses

For a 2D field \( x(i, j; t) \) at time \( t \), its domain-average 1D structure function can be defined as

\[
S_q(L) = \frac{1}{N} \sum_{j=1}^{N} \left[ \frac{1}{N - L} \sum_{i=1}^{N-L} |x(i, j; t) - x(i + L, j; t)|^q \right];
\]

\[
1 \leq L \leq N - 1; \quad q \geq 0, \tag{13}
\]

where \( N \) is the number of grid-cells across the length of a square domain, and \( L \) represents spatial separation (e.g. Marshak and Davis, 2005). Applying Equation (13) to each domain in a time series and arithmetically averaging the results yields \( \langle S_q(L) \rangle \). This differs slightly from the application of structure functions in the Appendix. If \( x(i, j; t) \) is scale-invariant over a range of \( L \),

\[
\langle S_q(L) \rangle \propto L^{\zeta(q)}, \tag{14}
\]

and when it exhibits weak autocorrelation, \( \langle S_q(L) \rangle \approx \) const. with \( \zeta(q) \approx 0 \). If \( \zeta(q) \) is nonlinear, \( x(i, j; t) \) is said to be multi-scaling or multi-fractal (Parisi and Frisch, 1985).

Figure 10 shows \( \langle S_q(L) \rangle \) for several key variables for simulations at \( SST = 295 \text{ K} \) (results for 300 K are similar and not shown). Encouragingly, differences between PGLQ and ICA, in form and magnitude, are trivial. The only notable differences are for the radiation fields at \( f_{RT} = 50,000 \), but as above, they are due directly to PGLQ sampling, not spatial fluctuations in cloud properties. Basically, as \( f_{RT} \) increases, mid- to long-range differences in PGLQ fluxes get suppressed, as do corresponding \( \langle S_q(L) \rangle \), yet short-range differences tend to become more pronounced via realization of step-like functions. This combination affects fields that look increasingly binary with smaller \( \zeta(q) \).

This last point can be seen in Figure 11a–d where \( \zeta(q) \) for radiation fields at \( f_{RT} = 50,000 \) are underestimated greatly, yet for \( W \) they are too large. The PGLQ simulations for \( f_{RT} \leq 5,000 \), however, exhibit excellent agreement.
FIGURE 8  Leftmost columns, of each set of three columns (left grouping coming from the SST = 295 K ICA simulation and right grouping from its 300 K counterpart), are binary maps of thresholded cloud water paths W’ (see Equation (10)) for three sample domains. Next to them are their associated fields of outgoing SW radiation (OSR) and OLR; value of the cloud aggregation metric C(32 km) (see Equation (11)) is listed above each row. Note that C(32 km) decreases going down columns. The small square in the lower-left map represents 32 × 32 km

FIGURE 9  Cumulative frequency distributions of cloud aggregation metric C(32 km) for all hourly domains of the ICA and PGLQ, using f_{RT} of 200 and 50,000, simulations for SST = 295 and 300 K as indicated. Dashed horizontal lines denote the 0.16 and 0.84 percentiles, which correspond approximately to one standard deviation, with shaded areas signifying the standard deviation for the ICA simulations

with their ICA correspondents, even for large q which suggests that extreme, yet rare, fluctuations across domains are captured well, despite some notable, but evidently

unimportant, differences in the magnitudes of extreme TOA fluxes (Figure 5).

TABLE 5  Two-sample Kolmogorov–Smirnov test statistics D_{n,n} for distributions of C(32 km) from ICA and various PGLQ experiments denoted by f_{RT}

| SST (K) | f_{RT} | D_{n,n} |
|--------|--------|---------|
| 295    | 200    | 0.050   |
|        | 5,000  | 0.035   |
|        | 50,000 | 0.068   |
| 300    | 200    | 0.077   |
|        | 5,000  | 0.075   |
|        | 50,000 | 0.251   |

Note: For sample-sizes n = 960, at α = 0.05, it can be assumed that distributions were drawn from the same population if D_{n,n} < 0.059.

5 | SUMMARY AND CONCLUSION

Virtually all cloud system-resolving models (CSRM) apply 1D radiative transfer (RT) models to all N columns in their domains. This is the Independent Column Approximation (ICA), and it becomes computationally burdensome for large N. Barker and Li (2019) presented the Partitioned Gauss–Legendre Quadrature
(PGLQ) algorithm for calculating RT in CSRMss. Its aim is to use a tiny fraction of the computing resources required by the ICA. In essence, PGLQ involves sorting columns according to variables that are important for RT, identifying $n_c \ll N$ “key” columns, applying RT models to them, replicating those flux profiles in nearby columns in sorted-space, and finally repositioning them back across the domain.

Barker and Li (2019) and Barker et al. (2019) verified the accuracy of PGLQ’s flux and heating rate estimates diagnostically by applying it and the ICA to the same cloud fields. To validate PGLQ is, however, more demanding as it requires using PGLQ in a CSRM and realizing “satisfactorily large” reductions in RT computation time and simulations that “satisfactorily” resemble those that are used for the ICA. The objective of the present study was to begin validation of PGLQ.

Use of the word “satisfactorily” highlights that PGLQ’s validity rests with subjective experimenters. Given the chaotic nature of CSRMs, alterations to them, no matter how trivial, will produce time series that eventually become entirely uncorrelated despite identical initial conditions. This might be important for some experiments, but most CSRM experiments, including those performed here, are concerned with temporal and spatial integrals of cloud and radiative variables. If null hypotheses involving such quantities cannot be rejected, then it can be assumed that alterations to a CSRM, including replacing the ICA with PGLQ, do not impact conclusions drawn from, and about, experiments.

In the Introduction, PGLQ was likened to a *lossy compression* algorithm; a familiar example of the latter is transformation of WAV audio files into relatively small MP3 counterparts, where potentially redundant information (i.e. data) gets permanently deleted to ease transmission and storage rates. If listeners are content with the quality and size of MP3 files, then the developers would deem their algorithm as having been “validated” by subjective users. Analogously, by calling RT models $n_c \ll N$ times, PGLQ represents a kind of *lossy computation* algorithm, whose validation, like MP3’s, resides with a subjective experimenter’s assessment of a CSRM’s response to much

---

**FIGURE 10** (a–f) Time-averages of second-order, domain-averaged 1D structure functions (see Equation (13)) as functions of separation distance $L$ for several variables, as listed in plot titles, for the SST = 295 K simulations using the ICA and PGLQ at three different values of $f_{RT}$ as listed on the plots. Heavy dashed lines are for reference indicate $\zeta(2) = 2/3$ (see Equation (14)) which is often found for satellite imagery (e.g. Cahalan and Snider, 1989; Barker, 1996)
reduced radiative information for the sake of numerical efficiency.

The PGLQ algorithm was installed in the System for Atmospheric Modeling (SAM) CSRM and run for conditions of radiative-convective equilibrium (RCE) at $SST = 295$ and $300$ K (Wing et al., 2018). Benchmark simulations used the ICA, which is SAM’s default. The PGLQ experiments used $f_{RT} = N/n_G = 200, 5,000$ and $50,000$; the latter setting is unlikely to be used routinely, but illustrates what happens when PGLQ is pushed to the extreme (the ICA is tantamount to $f_{RT} = 1$). Moreover, as $f_{RT}$ approaches $N$, PGLQ’s efficiency asymptotes, and by $f_{RT} \approx 50,000$ it is well on its way.

With much consistency, time/domain-averaged cloud properties and radiative fluxes simulated using $f_{RT} = 200$ and $5,000$ differ insignificantly, at the 95% confidence level, from their ICA-based counterparts. The exceptions are overestimates of total cloud fractions, due to slight reductions in the vertical overlap of fractional clouds, and too little radiative exchange taking place within the cloudy portions of layers that contain small amounts of cloud to begin with.

On the other hand, numerous variables exhibited large differences between ICA- and PGLQ-based simulations for $f_{RT} = 50,000$. Barker and Li (2019) already showed that for large $f_{RT}$, extreme values of radiative fluxes are often sampled poorly. Evidently, losses of radiative information at $f_{RT} = 50,000$ were so severe they affected undesirable localized anomalous forcings (cf. Cole et al., 2005). As such, it would appear that for RCE conditions, the PGLQ approximation eventually becomes untenable, but only for $f_{RT} > 5,000$. If this holds for other atmosphere–surface conditions, PGLQ should affect enough savings in computation time to allow relinquishment of some of those savings in order to reduce, and possibly eliminate, the need to skip time steps between calls to the RT models (e.g. Manners et al., 2009), and still be more efficient than the ICA. This reallocation of computation time and radiative flux errors could be very beneficial.

---

**FIGURE 11** (a) Multi-scale functions $\zeta(q)$ (see Equation (14)) as functions of $q$ for net SW flux at TOA for $SST = 295$ K when using the ICA and PGLQ for three values of $f_{RT}$ as listed on the plot. $\zeta(q)$ were computed by linear least-square regression using data from $L$ between 0.25 and 4 km (i.e. the extreme left of the curves shown in Figure 10). (b) As in (a) except this is for $SST = 300$ K. (c,d) are as in (a,b) except these are for net LW flux at TOA. (e,f) are as in (a,b) except these are for total cloud water path $W$. 
In addition to time/domain-averaged differences, horizontal geometric properties of simulated clouds and radiative fluxes were examined, too. For the cases considered, convection and associated clouds exhibited less aggregation as SST increased. Using cloud water path as a proxy for convective aggregation, the basic character of convective aggregation was impacted only weakly across a wide range of \( f_{RT} \). While this represents a positive outcome for the PGLQ method, it should not be interpreted as a substantive comment on the underlying nature of aggregation of atmospheric energy and moisture.

On a similar note, structure function analyses of several key cloud and radiative variables provided clear indications that horizontal arrangements of cloud water are impacted insignificantly for \( f_{RT} \leq 5,000 \). At \( f_{RT} = 50,000 \), however, simulated radiation fields showed obvious signs of PGLQ-induced information loss; that is, variance reduction akin to over-compressed JPEG images, with signs of subsequent corruption to the rest of the CSRM’s simulations.

In conclusion, the main finding of this study can be summarized as: large amounts of radiative information produced by the ICA in high-resolution CSRMs can be pared down strategically, via the PGLQ methodology, to reduce the computational burden of doing RT calculations in CSRMs, with minimal impact on the integrity of the overall simulation. If sizable values of \( f_{RT} \) can be established that facilitate sound and efficient CSRM simulations for all cloud regimes, the PGLQ method could be used with ease and advantage in high-resolution numerical weather prediction models as well as in local and global CSRMs.

ACKNOWLEDGEMENTS

Data used for this study can be obtained by contacting the corresponding author. The authors wish to thank Marat Khairoutdinov for helpful discussions pertaining to the use of the SAM CSRM. They also thank the European Space Agency for financial assistance with V. Dhanraj’s undergraduate work term at ECCC.

ORCID
Howard W. Barker https://orcid.org/0000-0003-1397-9505

REFERENCES

Abrahamowitz, M. and Stegun, I. (1972) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. NBS Applied Mathematics Series 55. Washington, DC: National Bureau of Standards.

Barker, H.W. (1996) A parameterization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds.

Part I: Methodology and homogeneous biases. Journal of the Atmospheric Sciences, 53, 2289–2303.

Barker, H.W. and Räisänen, P. (2005) Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. Quarterly Journal of the Royal Meteorological Society, 131(612), 3103–3122.

Barker, H.W., Qu, Z., Belair, S., Leroyer, S., Milbrandt, J.A. and Vaillancourt, P.A. (2017) Scaling properties of observed and simulated satellite visible radiances. Journal of Geophysical Research: Atmospheres, 122, 9413–9429. https://doi.org/10.1002/2017JD027146.

Barker, H.W. and Li, J. (2019) Accelerating radiative transfer calculations for high-resolution atmospheric models. Quarterly Journal of the Royal Meteorological Society, 145(722), 2046–2069.

Barker, H.W., Qu, Z. and Cole, J.N.S. (2019) Error analysis for an algorithm that reduces radiative transfer calculations in high-resolution atmospheric models. Journal of Advances in Modeling Earth Systems, 11, 4576–4596. https://doi.org/10.1029/2019MS001862.

Bony, S., Stevens, B., Frierson, B.D.M.W., Jakob, C., Kageyam, M., Fincus, R., Shepherd, T.G., Sherwood, S.C., Siebesma, A.P., Sobel, A.H., Watanabe, M. and Webb, M.J. (2015) Clouds, circulation and climate sensitivity. Nature Geoscience, 8, 261–268. https://doi.org/10.1038/ngeo2398.

Bretherton, C.S., Blossey, P.N. and Khairoutdinov, M.F. (2005) An energy-balance analysis of deep convective self-aggregation above uniform SST. Journal of the Atmospheric Sciences, 62, 4237–4292. https://doi.org/10.1175/JAS3614.1.

Bretherton, C.S. and Khairoutdinov, M. (2015) Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. Journal of Advances in Modeling Earth Systems, 7, 1765–1787. https://doi.org/10.1002/2015MS000499.

Cahalan, R.F. and Snider, J.B. (1989) Marine stratocumulus structure during FIRE. Remote Sensing of Environment, 28, 95–107.

Clough, S.A., Shephard, M.W., Mlawer, E.J., Iacono, M.J., Cady-Pereira, K., Boukabara, S. and Brown, P.D. (2005) Atmospheric radiative transfer modeling: a summary of the AER codes. Journal of Quantitative Spectroscopy and Radiative Transfer, 91, 233–244.

Cole, J.N.S., Barker, H.W., Randall, D.A., Khairoutdinov, M.F. and Clothiaux, E.E. (2005) Global consequences of interactions between clouds and radiation at scales unresolved by global climate models. Geophysical Research Letters, 32, L06703. https://doi.org/10.1029/2004GL020945.

Cronin, T.W. (2014) On the choice of average solar zenith angle. Journal of the Atmospheric Sciences, 71, 2994–3003. https://doi.org/10.1175/JAS-D-13-0392.1.

Hogan, R.J. and Illingworth, A.J. (2000) Deriving cloud overlap statistics from radar. Quarterly Journal of the Royal Meteorological Society, 126(569), 2903–2909. https://doi.org/10.1002/qj.49712656914.

Holloway, C.E. and Woolnough, S.J. (2016) The sensitivity of convective aggregation to diabatic processes in idealized radiative-convective equilibrium simulations. Journal of Advances in Modeling Earth Systems, 8, 166–195. https://doi.org/10.1002/2015MS000511.

Khairoutdinov, M.F. and Randall, D.A. (2006) High-resolution simulation of shallow-to-deep convection transition over land. Journal of the Atmospheric Sciences, 63, 3421–3436.
Wing, A.A., Emanuel, K., Holloway, C.E. and Muller, C. (2017) Convective self-aggregation in numerical simulations: a review. Surveys in Geophysics, 38, 1173–1197.

Wing, A.A., Reed, K.A., Satoh, M., Stevens, B., Bony, S. and Ohno, T. (2018) Radiative-convective equilibrium model intercomparison project. Geoscientific Model Development, 11, 793–813. https://doi.org/10.5194/gmd-11-793-2018.

How to cite this article: Barker HW, Qu Z, Dhanraj V, Cole JNS. Partial validation of a lossy compression approach to computing radiative transfer in cloud system-resolving models. QJR Meteorol Soc. 2021;147:363–381. https://doi.org/10.1002/qj.3922

APPENDIX A. HYPOTHESIS TESTING FOR TWO TIME SERIES

This appendix explains the method used in Sections 4.1 to 4.3 to test hypotheses which state that a quantity computed for two independent time series were drawn from a common population.

Let \(X(i, j, k; t)\) be a 3D field either sampled at time \(t\) or accumulated and averaged over period \(\Delta t\) ending at \(t\). For this study, most analyses were done for 2D horizontal subsets of \(X(i, j, k; t)\) or their vertical integrals; both denoted as \(x(i, j; t)\). Let \(\mu_x(t)\) and \(\sigma_x(t)\) be domain-wide mean and standard deviation of \(x(i, j; t)\), and let \(\langle \mu_x \rangle\) and \(\langle \sigma_x \rangle\) be (arithmetic) temporal-averages of \(\mu_x(t)\) and \(\sigma_x(t)\), respectively. To test null hypotheses regarding the significance, at confidence level \(1 - \alpha\), of differences between \(\langle \mu_x \rangle\) and \(\langle \sigma_x \rangle\) based on use of PGLQ and the ICA could be performed by applying \(t\)-tests to an ensemble of CSRM simulations; but this could be extremely taxing on resources. An alternate, more economical, approach was adopted and explained here.

Figure A1 shows time series of \(N = 960\) hourly \(\mu_x(t)\) and \(\sigma_x(t)\) for net SW fluxes at TOA for ICA and PGLQ at \(SST = 295\) K. The difference between their \(\mu_x(t)\) is 1.28 W m\(^{-2}\), or <0.5%, and standard deviations of \(\mu_x(t)\) for ICA and PGLQ are 5.9 and 6.3 W m\(^{-2}\), respectively; note that these are not \(\langle \mu_x \rangle\) but rather \(\sigma[\mu_x]\). Likewise, \(\langle \sigma_x \rangle\) differ by just 0.01 W m\(^{-2}\), and have standard deviations of 6.7 and 7.4 W m\(^{-2}\) for ICA and PGLQ, respectively. These quantities cannot, however, be used for \(t\)-tests as the samples are not independent. This can be appreciated by looking at structure functions of the series shown in Figure A1.
FIGURE A1  
(a) Time series of hourly- and 24-hourly-integrated domain-averaged net SW flux at TOA for the ICA simulation at \( \text{SST} = 295\, \text{K} \). Horizontal line indicates time-averaged mean value. (b) as in (a) except this is for the PGLQ simulation using \( f_{\text{RT}} = 200 \). (c,d) are as in (a,b) except these time series are standard deviations of net SW at TOA. Boxes between plots indicate temporal mean and standard deviation, while bars denote extremes.

FIGURE A2  
Second-order structure function \( S_2(T) \) (see Equation (A1)) of net SW at TOA’s time series shown in Figure A1a (for \( \text{SST} = 295\, \text{K} \)), as well as corresponding \( S_2(T) \) for net LW at TOA, surface precipitation (sfc precip), total cloud fraction (\( A_c \)), cloud water path (W)

The 1D structure function of order-\( q \) for \( \mu_x(t) \) is defined as

\[
S_q(T) = \frac{1}{N - T} \sum_{i=1}^{N-T} |\mu_x(t) - \mu_x(t + T)|^q; \\
1 \leq T \leq N - 1; \quad q \geq 0,
\]

(A1)

where \( T \) is separation period, and \( N \) the number of samples (cf. Marshak et al., 1997). When \( \mu_x(t) \) exhibits weak autocorrelation, \( dS_q(T)/dT \approx 0 \). Figure A2 shows \( S_2(T) \) for \( \mu_x(t) \) and \( \sigma_x(t) \) for several variables at \( \text{SST} = 295\, \text{K} \) for the ICA, including the series shown in Figure A1. For \( T \) smaller than \( \sim 24\, \text{hr} \), \( dS_q(T)/dT > 0 \), but for larger \( T \), all variables exhibit \( dS_q(T)/dT \approx 0 \). These results resemble
closely those for other \( q \) and \( \text{SST} = 300 \text{K} \). This suggests that if sequences, like those shown in Figure A1, are sampled at least every 24 hr, those subsets will often be pseudo-independent samples suitable for \( t \)-tests.

Let \( \mu_s(t; N_W) \) denote the subset of samples extracted from \( \mu_s(t) \), where \( N_W \) is number of hours between samples (for \( \sigma_s(t) \) it would be \( \sigma_s(t; N_W) \)). To get an approximate indication that the \( N/N_W \) members of \( \mu_s(t; N_W) \) are indeed pseudo-independent samples, consider the Wald–Wolfowitz runs-test (Wald and Wolfowitz, 1940), which tests if elements of a sequence are mutually independent. Letting \( \bar{x} \) be the median value of \( \mu_s(t; N_W) \), convert \( \mu_s(t; N_W) \) to a binary sequence as

\[
B(t) = \begin{cases} 
1 & \mu_s(t; N_W) \geq \bar{x} \\
-1 & \mu_s(t; N_W) < \bar{x}.
\end{cases} (A2)
\]

When \( B(m) \) flips sign, a “run” terminates and a new one starts. Let \( R, n_+, \) and \( n_- \) be, respectively, numbers of runs, and numbers of 1’s and −1’s in \( B(t) \). If \( B(t) \) displays uncorrelated fluctuations, the expected distribution of \( R \), for sequence length \( N/N_W \), is approximately normal with mean, variance, and test statistic given by

\[
\begin{align*}
\mu_R &= 1 + \frac{2n_+n_-}{N/N_W}; \\
\sigma_R^2 &= \frac{(\mu_R - 1)(\mu_R - 2)}{N/N_W - 1}; \\
Z_R &\approx \frac{|R - \mu_R|}{\sigma_R}.
\end{align*} (A3)
\]

These expressions are accurate when \( N/N_W = n_+ + n_- > 20 \); which it is for \( N_W = 24 \). If \( Z_R < 1.96 \), one cannot reject the null hypothesis, at \( \alpha = 0.05 \), which states that \( \mu_s(t; N_W) \) exhibits excursions above and below its \( \bar{x} \) that resemble closely those expected for an uncorrelated sequence.

Figure A3 shows \( Z_R \) for \( \mu_s(t; N_W) \) and \( \sigma_s(t; N_W) \) as functions of \( N_W \) for ICA at \( \text{SST} = 295 \text{ and } 300 \text{K} \). These are mean values using all possible sequences formed with \( N_W \). In almost all instances, \( Z_R \approx 1.96 \) for \( N_W > 24 \text{ hr} \) meaning that when \( \mu_s(t) \) and \( \sigma_s(t) \) are sampled less frequently than every 24 hr, the majority of the resulting \( 960/N_W \)-member sequences resemble independent samples drawn from a random process, such as Gaussian noise (Barker et al., 1999). Hence, considering results of structure function analyses and Wald–Wolfowitz runs-tests, it appears that \( N_W = 24 \) can be used safely to assess differences between ICA and PGLQ experiments.

In principle, standard null-hypothesis tests require that several conditions be met. To avoid violating them unwittingly, a bootstrap approach was applied using all possible \( 960/N_W \)-member sequences \( \mu_s(t; N_W) \) or \( \sigma_s(t; N_W) \) (Von Storch and Zwiers, 1999).

This requires drawing \( 960/N_W \) samples from them at random, with replacement, computing their means, and the quantity \( \tau \) that one wishes to test, which for a given \( \text{SST} \) are

\[
\begin{align*}
\tau &= \langle \mu_s^{\text{PGLQ}}(t; N_W) \rangle - \langle \mu_s^{\text{ICA}}(t; N_W) \rangle, \\
\tau &= \langle \sigma_s^{\text{PGLQ}}(t; N_W) \rangle - \langle \sigma_s^{\text{ICA}}(t; N_W) \rangle. (A4)
\end{align*}
\]

where a simulation’s RT treatment is indicated by the superscript. This process of sampling with replacement to generate “mock-datasets” is repeated \( N_{bs} \) times yielding \( N_{bs} \) values of \( \tau \), which are then sorted from smallest to largest and designated as \( \tau^* (N_{bs}) \). The lower and upper values at confidence level \( 1 - \alpha \) are defined as

\[
t_{\frac{\alpha}{2}} \equiv \tau^* \left( \frac{N_{bs}}{2} \right) \quad \text{and} \quad t_{1-\alpha} \equiv \tau^* \left( \frac{1 - \alpha}{2} N_{bs} \right). (A5)
\]

If \( t_{\frac{\alpha}{2}} \cdot t_{1-\alpha} < 0, \) \( \tau = 0 \) resides inside the confidence interval and \( H_0 \) cannot be rejected at the \( 1 - \alpha \) confidence level. For all tests performed here, \( N_{bs} = 10,000 \) for each of the possible \( N_W \) sampled sequences.

Figure A4 shows \( \tau \) for the following three hypotheses for domain-average net SW TOA fluxes:

\[
\begin{align*}
H_0(295 \text{ K}) : \langle \mu_s^{\text{PGLQ}(295)} \rangle - \langle \mu_s^{\text{ICA}(295)} \rangle &= 0 \\
H_0(300 \text{ K}) : \langle \mu_s^{\text{PGLQ}(300)} \rangle - \langle \mu_s^{\text{ICA}(300)} \rangle &= 0 \\
H_0(300 \text{ K} - 295 \text{ K}) : [\langle \mu_s^{\text{PGLQ}(300)} \rangle - \langle \mu_s^{\text{ICA}(295)} \rangle] - [\langle \mu_s^{\text{PGLQ}(300)} \rangle - \langle \mu_s^{\text{ICA}(295)} \rangle] &= 0. (A7)
\end{align*}
\]

The first two state that PGLQ and ICA are the same when \( \text{SST} = 295 \text{ K} \) and then \( 300 \text{ K} \), while the third states that the difference between PGLQ at \( 300 \text{ and } 295 \text{ K} \) is the same as the corresponding difference for ICA. \( H_0(295) \) is not rejected at \( \alpha = 0.05 \) as the condition in Equation (A6) is met for most sample sequences. Conversely, \( H_0(300) \) is roundly rejected as Equation (A6) is rarely satisfied. For \( H_0(300 - 295) \), however, where mean differences for PGLQ and ICA are −4.48 and −4.85 W·m⁻², which differ by just 0.37 W·m⁻² or ~8%, is not rejected at \( \alpha = 0.05 \) as Equation (A6) is always satisfied.

Distributions of \( \tau^* \) shown in Figure A4 are almost symmetric. This suggests that conventional \( t \)-tests would suffice. Indeed, standard \( t \)-statistics are 1.02, 3.47 and 0.66 for \( H_0(295) \), \( H_0(300) \) and \( H_0(300 - 295) \), respectively, implying rejection of \( H_0(300) \) but not the other two. This agrees with the bootstrap results. At the outset, however, one does not know if the bootstrap method will be needed, so it was used for all tests reported in this study, thereby erring on the side of caution.
**FIGURE A3**  (a) Z-scores for the Wald–Wolfowitz runs-test (see $Z_R$ in Equation (A3)) as functions of hours between samples $N_{NW}$ for ICA at $SST = 295 \text{ K}$ for the same time series of domain-averages used to compute $S_2(T)$ in Figure A2a. Horizontal dashed line denotes critical z-score of 1.96 for $\alpha = 0.05$; when $Z_R$ is below it, values making up the time series are indistinguishable from that expected for an uncorrelated sequence. (b) As in (a) except these are for time series of standard deviations of the listed variables. (c,d) as (a,b) except they apply to simulations using $SST = 300 \text{ K}$.

**FIGURE A4** Cumulative frequency distributions of $\tau$ (see Equation (A4)) for all possible sequences of time series of net SW at TOA constructed from samples using $N_{NW} = 24$. Each “curve” was created using $N_{nb} = 1,000$ bootstrap sampled mock datasets. If a “curve” has values at the 0.025 and 0.975 percentiles on either side of 0, the null hypothesis $H_0$ stated in Equation (A7), and as listed across the top, is not rejected; that is, the PGLQ and ICA series can be assumed to have been drawn from a common population.