Wing et al. have suggested a protocol for the second phase of the Radiative Convective Equilibrium Model Intercomparison Project (RCEMIP), following on from the Phase One experiments which were motivated in their 2017 paper, also in GMD. My general thoughts are that the protocol is poorly motivated, and neglects many of the (arguably more promising) avenues in which a Phase II of RCEMIP could explore in favor of a relatively untested setup, the 'mock-Walker' simulation.

In my opinion, the following issues mean that the current protocol is unfit for publication in its current form.

Issue #1: approach to constraining model diversity

The 2017 protocol paper (RCEMIP-I) outlined a series of small and large domain CRM/LES simulations, in addition to suggested runs with global models and single-column models. One of the key takeaways from RCEMIP-I is that there is substantial diversity in the simulated RCE state even in the absence of convective self-aggregation.

This point is acknowledged in the current manuscript, which notes that: "While several robust results emerged across the spectrum of models that participated in the first phase of RCEMIP (RCEMIP-I), two points that stand out are (1) the strikingly large diversity in simulated climate states and (2) the strong imprint of convective self-aggregation on the climate state." This diversity in RCEMIP-I is also acknowledged to be driven by "representations of convection, microphysics, turbulence, and dynamical cores".

However, the current manuscript does little to tackle the question of inter-model diversity in the RCE state and instead primarily focuses on point (2) stating that "...the wide range in the degree of self-aggregation and the lack of consensus in its temperature dependence is a barrier to understanding."

This emphasis is counter to the vision laid out in the RCEMIP-I protocol (and contrary to the response given to the RCEMIP-I reviewers, quotations from which will be cited below). Originally, RCEMIP-I was presented as an opportunity to explore inter-model diversity, with RCEMIP-II being an opportunity to try to understand/narrow that diversity through the use of simplified radiation/microphysics schemes. For example, many of the reviewers for RCEMIP-I strongly suggested the use of simplified radiation and/or microphysics schemes as a way to better understand the diversity of RCE states and their response to warming. A list of these is presented below:

(Isaac Held): "*I would strongly encourage you to reconsider and ask groups to run with a standard microphysical mechanism in addition to their model's microphysics. Otherwise, there is a good chance that the diversity of simulations will be dominated by the diversity of microphysical assumptions. (For example, we know that in RCEs assumptions about ice fall-speeds will exert a strong control on the cirrus climate.)*"

(Levi Silvers): "*It is not essential, but I think it would be useful to be more precise about a second set of non-required (Tier 2) experiments. This could be written in such a way that modeling centers wishing to participate with minimal effort are not thus discouraged from participating, but that more ambitious modeling centers or individuals could clearly push farther into the project in a coordinated way. My suggestions for further experiments would be: 1. Rotating RCE 2. GCMs in RCE mode with convective parameterization turned off 3. RCE with cloud RCE off (COOKIE type experiments) 4. Kessler physics across the hierarchy of models*"

(Anonymous reviewer IV): "*...the resulting equilibrium state and the clustering may look very different in the different CRMs. It will thus be very difficult to compare the different* *models and to identify the root for the differences (radiation scheme, microphysics parametrizations, . . .). An even simpler setup for the models could therefore be useful to identify, which schemes are responsible for the differences. As suggested by the authors and brought forward by Jeevanjee et al. (2017) a simplified microphysics scheme could be one option. A further option could be to simplify the longwave radiative cooling, as e.g. described in Muller and Bony (2015).*"

(Nadir Jeevanjee): "*As advocated for by Isaac and other reviewers, there is also interest here in using simplified (Kessler) microphysics in our RCE setup. Such a scheme already exists in development branches of our code.*"

Furthermore, this was explicitly stated as a potential target for the second phase of RCEMIP (Sec 6.2 of 2017 protocol paper):

"*Additional simulations could be performed to assess the sensitivity to dynamical core, radiation scheme, microphysics scheme, boundary layer scheme, convective scheme (in the case of models with parameterized convection), and the sensitivity to various parameters in those schemes (such as the entrainment parameter in a convective scheme).*"

Additionally, in the reply to Reviewer IV of RCEMIP-I (similar comments exist in the reply to other reviewers):

"*We agree that large differences could result from differences in physical parameterizations. However, we think that it is useful to first determine the full range of RCE simulations and then proceed to test the parameterization sensitivity by imposing a simple microphysics scheme on all models in the second phase of RCEMIP.*"

From these statements it is clear that simplified experiments were anticipated to be necessary/useful by reviewers for understanding inter-model diversity in the simulated RCE state. This was also acknowledged by the authors and explicitly suggested as a route forward for Phase II of RCEMIP. The need for idealized experiments seems even more necessary now that the RCEMIP-I simulations have demonstrated such an extreme diversity in the simulated RCE state (perhaps even larger than expected).

Furthermore, since RCEMIP-I it has been fairly well-established that relevant quantities such as high cloud feedbacks are extremely sensitive to microphysics. For example, both Wing et al. (2020) and Stauffer and Wing (2022) demonstrated that about one-third of models do not exhibit an 'iris' effect, and in fact have an *increase* in high cloud fraction under warming despite a decrease in clear-sky divergence at the anvil level. The disconnect between radiatively driven divergence and high-cloud fraction explored in Seeley et al (2019), Beydoun et al. (2021) and Jeevanjee (2023) (among others), and one key takeaway from those papers is that the lifetime of detrained cloud condensate is a crucial determinant of high-cloud fraction and the anvil cloud feedback. It is thus extremely confusing that the authors do not explore the possibility of using simplified microphysics schemes, or at least require that all models output microphysical process rates and (where relevant) particle size distributions, which would be trivial for most models, and allow for a better understanding of this crucial feedback.

While implementing idealized microphysical schemes is a potential burden on modeling centers (though some models such as FV3 already have the option to use a Kessler-style microphysics package), it is quite simple to replace interactive radiative transfer with a prescribed radiative cooling rate. For example, Paulius and Garner (2006) use a simple 1.5K/day radiative cooling rate up to a fixed tropopause temperature. Such an approach is already widely used in the RCE literature. Additionally, a slightly more realistic approach, which better captures the effect of warming on radiative fluxes, is to prescribe the radiative flux divergence in temperature coordinates ($\partial_T F$) as being linear in temperature (following Eq. 2 of Jeevanjee and Zhou (JAMES, 2022)). Such an approach was also used by Seeley and Wordsworth (2023).

My point here is not simply that we should revisit Phase One experiments, for which simplified runs would be helpful, but also that runs with simplified physics will likely be *even more necessary* in RCEMIP-II if the authors do intend on using 'mock-Walker' simulations. This is because it is now well-established (although not cited in the protocol paper…) that the 'mock-Walker' simulations are prone to develop stacked overturning circulations (e.g. Grabowski et al., 2000; Yano et al., 2002; Larson and Hartmann, 2003; Liu and Moncrieff , 2008; Silvers and Robinson, 2021), due to the interactions between radiation and detrained condensate/water vapor (e.g. Nuijens and Emanuel, 2018; Sokol and Hartmann, 2022; Lutsko and Cronin, 2023). An example of which, from experiments with the ICON model, is shown below (domain is ~3000x100km, with dx=2km, but similar results are obtained for the standard RCE_large setup):

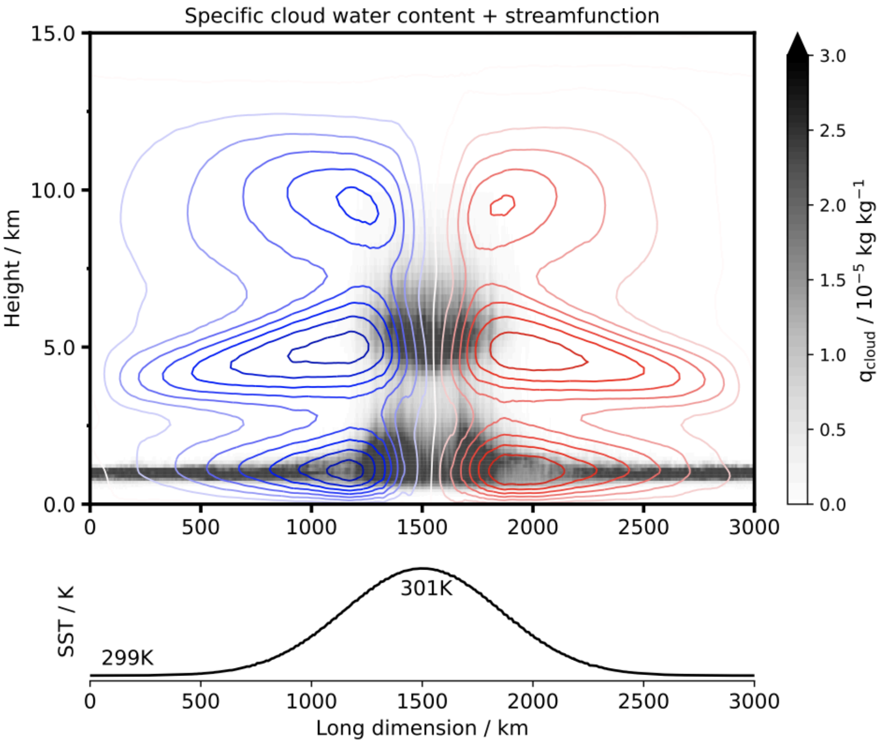

Figure 1: Time-averaged streamfunction (red=clockwise motion; blue=anticlockwise motion) and cloud liquid water content in our control simulation, driven by a zonal SST gradient between the centre and edges of the domain.

As shown in Fig. 1, the ICON model exhibits a pronounced 'triple-cell' structure, with three stacked overturning circulations. This structure differs between models, with some simulating a triple-cell structure (ICON), others a double-cell (e.g. Silvers and Robinson, 2021). Furthermore, the onset of these stacked overturning structures is itself dependent on temperature (Lutsko and Cronin, 2018, 2023). In both their 2018 experiments (2D CRM) and in their 2023 experiments (3D CRM), Lutsko and Cronin found that the circulation transitions from a single- to double-cell structure at mean SSTs of ~300K. Although this number is likely model-dependent, it is worrying that in their SAM simulations the transition occurred right in the middle of the SSTs considered as part of the Phase Two RCEMIP protocol.

This model- and SST-dependent transition between different stacked circulation structures will certainly complicate analysis and interpretation of the proposed RCEMIP-II experiments. In light of this I strongly disagree with the statement on L103 that mock-Walker simulations will "...reduce the diversity of simulated climates and provide a clearer tie to observations…"

Previous studies have shown that the stacked-overturning cells can be effectively suppressed by prescribing uniform (or otherwise simplified) radiative cooling profiles (e.g., Grabowski et al., 2000; Wofsy and Kuang, 2012; Sokol and Hartmann, 2022; Lutsko and Cronin, 2023). This is effective at suppressing these features because they are largely driven by sharp vertical gradients in the longwave cooling rate, caused by sharp gradients in moisture near the melting line. These moisture gradients are themselves associated with complex interactions between convection, microphysics and radiation; near the melting line, evaporation of detrained condensate and melting of ice form a stable layer which promotes further detrainment of condensate, whose enhanced radiative appears to draw out more condensate (Nuijens and Emanuel, 2018; Sokol and Hartmann, 2022).

Overall, one of the most salient features of mock-Walker circulations is their stacked overturning circulations and the elevated moist layers associated with them. Both of these appear to be governed by complex interactions between microphysics (with both the liquid and ice phase playing a key role), convection and radiation. Hence, it seems highly unlikely that these mock-Walker simulations will "provide a constraint on convection and circulation" or "narrow the intermodel spread", as hypothesized in the protocol.

I find it somewhat unlikely the authors will give up on mock-Walker simulations altogether, but to claim (as they do in the manuscript) that they will narrow model diversity is quite naïve. If they do insist on using mock-Walker simulations (see point #2, below for more on this), then it is highly recommended that they pair such simulations with simplified radiation/microphysics runs to help narrow some of the model diversity and aid interpretation and understanding of the results. If it is deemed too complex to impose a uniform microphysics scheme, then models should instead be required to output comprehensive microphysical diagnostics.

Issue #2: the choice of mock-Walker simulations over other options

The RCEMIP-I protocol named a number of possible routes forward for Phase Two. For example, introducing a mixed-layer ocean to ensure a closed surface energy budget, or introducing rotation (f-plane / beta-plane / full spherical geometry) in order to simulate tropical cyclones and equatorial waves. Both of these options have been explored in single-model studies prior to and following on from RCEMIP-I, and thus it is confusing why the authors decided to use the mock-Walker setup instead of these alternative approaches. For example, investigating tropical cyclones in f-plane simulations has been richly explored in a number of single-model papers, but each with differing setups. This seems like a natural avenue for RCEMIP, especially given that tropical cyclones are an incredibly impactful phenomena which we frequently observe in the real-world, as opposed to the "stacked" overturning circulations of mock-Walker cells (which have no obvious observational analogue).

During the RCEMIP breakout session at CFMIP 2023, a number of participants expressed interest in rotating RCE experiments. It is thus extremely strange to me that the current protocol paper does not justify its emphasis on mock-Walker simulations and argue for their benefit *in comparison to* rotating RCE (or indeed other configurations suggested by the RCEMIP Phase One paper).