## Initial Response to Reviewer Comments

We are providing an initial response to the reviewer comments while the discussion period is still open, which is intended to provide more background on how mock-Walker simulations came to be selected for RCEMIP-II and justification for the experimental design. A line-by-line response to other aspects of the reviewer comments is reserved for after the discussion period has closed. If given the opportunity we would revise the manuscript to clarify the motivations for mock-Walker simulations as the choice for RCEMIP-II as discussed below.

### *Evolution of Phase II:*

The authors of this paper began discussing possible directions for a second phase in Spring 2021, shaped by the initial findings and ongoing analysis of RCEMIP-I by ourselves and other groups. Mock-Walker simulations emerged as our preference, based on their ability to satisfy the four principles discussed below in response to reviewer #2, and the fact that numerous groups in the community were already performing such simulations, with a variety of interesting, yet complex, results. We then inquired with some of the other core RCEMIP participants regarding their thoughts about this proposal. After hearing positive feedback as to the value of this choice and its suitability as a way to move forward, as well as willingness to perform the simulations, we reached out to everyone that contributed simulations to phase I of RCEMIP. Again, we received positive feedback and willingness to contribute, such that by the time A. Wing submitted an NSF proposal in summer 2021 proposing to perform mock-Walker simulations as a phase II of RCEMIP, 18 models had indicated that they would participate.

By summer 2022, Wing's NSF proposal had been funded and mock-Walker simulations were presented at the CFMIP meeting in Seattle and the 2nd Model Hierarchies Workshop at Stanford University as the idea for phase II. Through informal conversations at these meetings and communication with others in the community who had previously or were actively performing their own mock-Walker simulations, the detailed protocol began to take shape. At the end of 2022, mock-Walker simulations were announced to the broader RCEMIP community as the planned phase II, as were other potential opportunities to utilize the RCEMIP set-up to investigate other phenomena, such as tropical cyclones in rotating RCE simulations and cloud-aerosol interactions. The former received interest from a handful of people, while the latter, led by Guy Dagan (Hebrew University) received interest from about 10 CRMs and has subsequently proceeded as an offshoot of RCEMIP (RCEMIP-ACI). Throughout 2023, we tested and optimized the mock-Walker protocol and simulation design. Throughout, we sought to listen to input from the community and address any concerns they brought up. For example, we performed the domain size tests presented in this paper in response to a concern from a community member that there may be sensitivities to both the domain length and width. We switched from prescribing the same wavelength of the SST perturbation across CRMs of slightly different domain lengths to enforcing a wavelength always equal to the domain length, after a RCEMIP participant pointed out some numerical and non-physical artifacts of the former option.

At the Joint CFMIP-GASS Meeting in Paris in July 2023, we held a RCEMIP breakout discussion. This was an opportunity to hear further feedback from the community before the protocol was finalized. The main topics of conversation were (1) exactly what delta-SST values to use; (2) the GCM configuration; and (3) the output request. In response to this, we performed 9 additional sensitivity tests to help us determine the delta-SST values as indicated in this paper. We added additional output variables to meet the analysis needs of the community. We re-considered the GCM configuration, though we ended up back at our original proposal (the one presented in this paper) as the best way to make the CRM and GCM set-ups consistent. As reviewer #2 notes, there was also some interest expressed in rotating RCE experiments, but we elected to stick with mock-Walker simulations as the choice for the official phase II for the reasons discussed below. We have received a much smaller volume of interest in rotating RCE as the framework for RCEMIP II, compared with the mock-Walker simulations. As pointed out by the editor, a MIP is only useful if multiple modeling groups commit to performing the experiments. At the time of the submission of this protocol paper in December 2023, 17 people in addition to the authors of this paper had expressed enthusiasm about the proposed mock-Walker simulations and their willingness to contribute simulations. The 20 models planning to participate are listed in Table 2 of the paper. Since submitting this paper, an additional ~10 people have expressed interest in participating. While the lengthy period of testing and optimization was necessary to ensure a robust protocol and address issues pointed out by the community, further delay in beginning phase II risks losing the momentum we have built.

Reviewer #1

*Maturity of the mock-Walker set-up:*

In designing the protocol for the mock-Walker simulations as RCEMIP-II, we tested different versions of the equation for prescribed SSTs, the choice of delta-SSTs, the CRM domain length and width, maintaining different aspects of the SST pattern across CRMs of slightly different domain length, and how to make the CRM and GCM set-ups consistent. While there are always more sensitivity tests that could be done, the protocol was not decided upon lightly and is instead the result of several years of consideration and optimization. As described above regarding the evolution of phase II, in multiple cases, we performed additional tests and even changed the protocol in response to comments from the RCEMIP community. Therefore, we feel like the testing presented in this paper in combination with the substantial body of literature on mock-Walker simulations over past decades makes it appropriate to do a mock-Walker intercomparison. Further tests in the context of one or a few models could continue to be performed in parallel with the broader intercomparison, as needed.

*SST gradient vs. absolute SSTs:*

Reviewer #1 makes a good point that it is unclear which SST quantity matters most for the climate of mock-Walker simulations. From the perspective of the weak temperature gradient approximation, the absolute SST contrast (maxSST - minSST) and the maximum SST is what ought to matter to the dynamics. The SST gradient (dSST/dx) might plausibly set horizontal flow speeds under Lindzen-Nigam type arguments,  and if that is the case, the SST laplacian ought to matter for vertical motion and precipitation. Ideally all three of these parameters would be kept fixed across the models, but due to computational limitations on domain size this is not possible. This issue is one that we considered at length, discussed with other members of the RCEMIP community, and tested extensively in preparing the RCEMIP-II protocol.

We discussed two possible options in the paper: (1) enforcing that the wavelength equals the domain length and (2) enforcing the same wavelength (6000 km) regardless of domain length. Option (1) keeps the absolute SST contrast (delta-SST), maximum SST, and mean SST the same but leads to slightly different SST gradients. Option (2) leads to slightly different mean SSTs, a discontinuous SST distribution at the boundaries, and the projection of the prescribed SST forcing onto all scales, introducing substantial noise at higher wavenumbers.

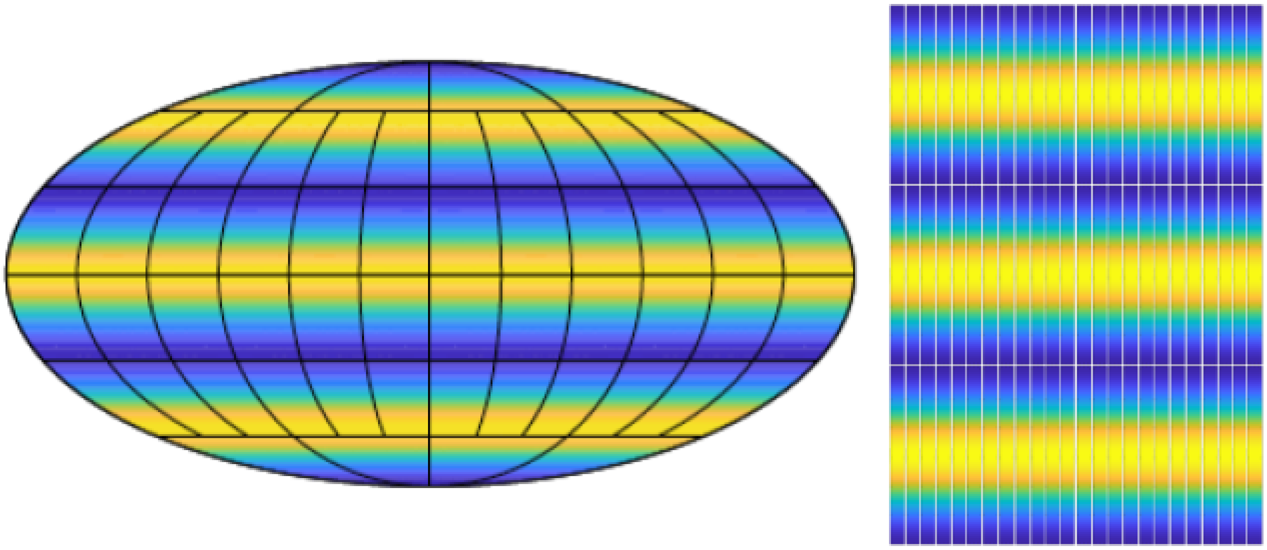
After testing in one CRM, SAM, we elected to go with option (1). Even though this choice could cause differences in the results due to differences in domain length, these differences would at least result from a physical reason (a different SST gradient) which we felt was preferable to the non-physical artifacts present in option (2). Adjusting the value of delta-SST, as the reviewer suggests, would cause differences in the absolute SST contrast and maximum SSTs. We feel that keeping the absolute SST contrast and maximum SSTs consistent (our chosen option) is the most elegant and plausibly what matters most for dynamics, precipitation, clouds, etc… based on weak temperature gradient arguments. It is also the simplest to implement.

The differences in the SST gradient that would result from different domain lengths are small compared to the differences in SST gradient from choosing 0.625 K, 0.75 K, or 1 K as the “weak gradient”, for instance. So, while we cannot rule out that the differences in domain size could contribute to differences across models, we believe it would be a small effect. This is supported by the testing we did in SAM, in which our results did not qualitatively depend on the choice of option (1) or option (2). We could attempt to determine the influence of domain length on the results by assessing if models with a larger domain length behave systematically differently from those with a smaller domain length.  While we acknowledge that it could be difficult to disentangle the relative contributions of the domain length difference and other aspects of model physics and numerics to intermodel differences, this difficulty would also be present if we were to employ any of the other options. The only thing we can do is try to make the set-up as uniform as possible given computational limitations and avoid the imposition of non-physical artifacts.

*GCM set-up:*

Regarding the reviewer’s concern about the GCM set-up, we considered and tested alternate geometries of the SST pattern, which was a subject of discussion in the RCEMIP breakout session at the 2023 Joint CFMIP-GASS Meeting in Paris. We ultimately elected to utilize zonal bands of hot and cold SSTs to ensure the closest possible correspondence to the set-up in the doubly periodic long-channel CRM domain (chosen to be identical to the RCEMIP-I domain), including the mean SST, maximum SST, and SST gradient. Given the double periodicity, the CRM domain should be conceptualized as being infinitely repeated in both dimensions, which would then (other than the sphericity) make it analogous to the GCM set-up. This is demonstrated in the figure below, in which the SSTs in the GCM are shown on the left and the SSTs in the CRM are shown on the right, in which the CRM has been rotated and tiled 24 times in one dimension and 3 times in the other, to emulate the GCM domain. Note that since these simulations are non-rotating, the choice of which direction is x- and which direction is y- does not matter.

Since one of the core principles of RCEMIP is to be able to compare limited area CRMs and GCMs, we chose to confine the warm SSTs in only one direction in both model types. The GCM set-up is similar to that used in Mueller and Hohenegger (2020), which utilized zonally homogenous but meridionally varying SSTs. Convective self-aggregation within the warm latitude bands is indeed possible, as the reviewer suggests. Zonal contraction of convection was seen in Mueller and Hohenegger (2020) and in our test simulations with CAM. While this could complicate interpretation, it will also be interesting to see how the degree to which this "intrinsic" self-aggregation emerges on top of the forced convergence varies across models. The ability to study self-aggregation both in the context of SST gradients and constant SST is one of our motivations for this particular setup and creates an additional connecting point both with the RCEMIP experiments and the tropical oceans of Earth.



<u>Reviewer #2</u>

In selecting an experimental design for phase II of RCEMIP, we sought to follow the following principles, in the spirit of the design of RCEMIP-I:

1) The ability to directly compare CRMs and GCMs
2) Ease of implementation, to encourage the broadest possible participation
3) Permitting continued investigation of the three themes of RCEMIP while moving a step up the model hierarchy
4) Providing some sort of constraint on convection.

Several different possibilities were suggested in the RCEMIP-I protocol paper (Wing et al. 2018) and were considered by the author, as paths forward for a second phase. A mock-Walker configuration, as proposed, meets all of these criteria. It does not prohibit other possible studies but simply represents a practical direction that both our conversations with colleagues and our work over the past few years has naturally taken. Above we reviewed the process by which the proposed mock-Walker protocol emerged. Here, we discuss how mock-Walker simulations satisfy these four principles, whereas other possible options for a second phase of RCEMIP do not.

*Mock-Walker:*

The proposed mock-Walker experimental design maintains an identical set-up to RCEMIP-I with the exception of a simple, prescribed SST pattern. This is easy to implement and allows for direct comparison with RCEMIP-I. Care was taken to maintain consistent values of mean SST, maximum SST, and SST gradient between the CRM and GCM domains, as much as possible. These characteristics satisfy **principles 1 and 2.**

From its inception, RCEMIP has been motivated in part by a desire to better understand how the balance between convection and radiation interacts with large-scale circulations (Wing et al., 2018). However, the only large-scale circulations present in RCEMIP-I are those generated by self-aggregation. One of our motivations for selecting mock-Walker simulations is a desire to broaden the range of dynamical regimes and cloud types that can be simulated by moving one step up the model hierarchy from RCE. Interactions between convection and a large-scale circulation that is forced by SST anomalies have direct analogues on Earth to the ITCZ, the Walker Circulation, and the Hadley Circulation (in a non-rotating context). The presence of subsiding circulations consistently occurring over regions of cooler SST also allows for the possibility of simulations that include stratocumulus clouds. Initial tests with SAM indicate that the mock-Walker simulation does contain optically and geometrically thicker low-clouds than RCEMIP-I (Stauffer 2023, PhD thesis). The SST gradient, combined with an overturning circulation, also allow for the possibility of modeling the transition between shallow and deep convective clouds. Additional dynamical regimes and cloud types could increase the sensitivity of the model results to details of the parameterization schemes, but we think that RCEMIP-I and a mock-Walker based RCEMIP-II will provide an excellent dataset for investigating many interesting question and will serve as a good reference for further study.

The mock-Walker configuration thus allows investigation on the three themes of RCEMIP (robustness of simulated mean state, response of clouds to warming and climate sensitivity, and dependence of convective aggregation on temperature) in a framework focused on cloud-circulation coupling that is one step up the model hierarchy from RCE, satisfying **principle 3.** The characteristic that makes RCE an idealized model configuration is its homogeneous boundary conditions (uniform surface temperature and insolation). Thus to move one step up the model hierarchy from RCE closer to the real world, we need to relax the idealization of the boundary conditions. A prescribed sinusoidal SST pattern as in the proposed mock-Walker set-up achieves this by providing a still idealized but more realistic (i.e., heterogeneous) boundary condition, while maintaining all other aspects of the original RCEMIP set-up. It brings in *one* of the dynamical instabilities that is present in the real world but was not present in the original RCEMIP – a circulation forced by an SST gradient. This provides a clearer tie to observations than the original RCEMIP simulations with uniform SSTs.

The prescribed SST gradient in the mock-Walker simulations provides forcing for low-level convergence towards the warmest SSTs and will drive a large-scale circulation that provides a dynamical constraint on the location and spatial pattern of convection, satisfying **principle 4.** Reviewer #2 argues that the prescribed SST gradient will not provide a constraint on convection and circulation or narrow the intermodel spread. It is possible that we are overly optimistic about the degree to which the prescribed SST gradient will reduce the diversity of simulated climates, but we maintain that a prescribed SST gradient of at least moderate strength will provide a dynamically-forced organizing constraint on the convection and circulation relative to uniform SSTs. The ability of SST gradients to dynamically constrain/organize the convection is clear in our present paper as well as numerous previous studies (e.g. Grabowski et al., 2000; Tompkins, 2001; Bretherton et al., 2006; Lutsko and Cronin, 2018; Silvers and Robinson, 2021). We believe that the extent to which the prescribed SST gradient constrains convection and circulation in an environment of complex interactions between moist convective processes, radiation, and microphysics is a question worthy of investigation across an ensemble of models.

*Rotation:*

We do not feel that it would be possible to satisfy **principle 1** while including rotation. Based on the author's own experience, and the abundant prior literature, f-plane RCE simulations in a limited area CRM domain are quite different from rotating RCE simulations in a GCM (realistic rotation on the sphere with uniform thermal forcing). In our view, f-plane simulations in a limited area domain, while richly explored and yielding valuable insights on many questions, preclude investigation of essential questions about tropical cyclone frequency or genesis rate, since at low-f a single TC is artificially squeezed into the domain size provided and at high-f the number of TCs is controlled by the maximum packing. GCM simulations with uniform rotation have similar issues. Rotating RCE on the sphere would be a promising set-up for an intercomparison about tropical cyclones, but there is no obvious CRM analog other than a global CRM or perhaps a large beta-plane, but the latter would entail a different domain set-up to RCEMIP-I and would be more computationally expensive (in opposition to **principle 2**). While adding rotation does move up the model hierarchy of complexity from non-rotating RCE (supporting **principle 3**), the themes that would be investigated are likely tropical cyclone-focused. Though such questions are of great interest in general, they are different from the current themes of RCEMIP (opposing **principle 3**). Rotation allows for additional dynamical interactions that could provide a constraint on convection (**principle 4**), though it is not clear to what extent. The same sensitivities to microphysical and radiative parameterizations that Reviewer #2 is concerned about in the context of mock-Walker simulations would also likely be present in rotating simulations.

*Interactive SST:*

Performing simulations with interactive SST involves jumping further up the model hierarchy. While an important step towards the real world, RCE simulations with interactive SSTs have been studied in far less detail than mock-Walker simulations. Slab mixed layer oceans of even relatively shallow depth take many hundreds of days to reach equilibrium (Cronin and Emanuel, 2013). This greatly increases the computational expense, particularly for CRMs (in opposition to **principle 2**). To reach our goal of relaxing the idealization of uniform SSTs, we chose mock-Walker simulations over interactive SSTs as the next step partly for this pragmatic reason, and partly because of our interest in the scientific questions that open up once we have a system with a forced circulation. We are open to revisiting the idea of interactive SSTs if we make it to a phase three of RCEMIP :-)

*Simplified physics:*

Simplified radiation/microphysics schemes was another possible direction for a second phase of RCEMIP. Imposing simplified physics schemes can in principle be done in both CRMs and GCMs, satisfying **principle 1.** However, a simplified microphysics scheme would be significantly more complicated to implement in most models, opposing **principle 2**. Simplified physics would provide a further constraint on convection (supporting part of **principle 3**), but would move *down* the model hierarchy towards more idealization, not less, and, depending on the types of simplifications, could remove some phenomena of interest, such as self-aggregation, as topics of investigation (opposing **principle** 3). Furthermore, the more the physics is modified to be simpler, the further the models diverge from their parent models. In the case of GCMs, the ability to learn about the comprehensive version of the model from more idealized configurations was a strength of RCEMIP (e.g., Reed et al. 2021). This would be less likely with the use of simplified physics. Simplified physics would likely provide a constraint on convection (supporting **principle 4**). But while we admit that it is more complicated, we prefer a dynamical constraint on convection (as in the mock-Walker set-up) to a constraint provided by removing physical processes (as in simplified physics), because the former is more consistent with how convection is constrained in the real world.

Our view of the value of simplified physics has evolved since we originally suggested in the RCEMIP-I protocol paper and the response to reviewers there that it would be needed. We believe that simplified physics is valuable when trying to isolate the minimal ingredients necessary for a particular physical mechanism, typically within the context of experiments within an individual model. With an individual model, mechanism denial experiments, in which specific mechanisms are methodically removed through targeted simplifications, are also an excellent tool for determining the role of particular processes. We are sure that assessing the sensitivity to dynamical core, radiation scheme, microphysics scheme, boundary layer scheme, convective scheme, and the sensitivity to various parameters in those schemes would likely lead to both interesting and informative results. However, we feel that it is more suitable and tractable to do this in investigations with a single model or related group of models, rather than a large intercomparison.

In a model intercomparison, simplifying some of the physics could be useful for ruling out particular sources of intermodel spread. However, in our opinion the goal of an intercomparison is NOT to constrain models so much that they are forced to agree. When a robust result emerges from an intercomparison *in spite of* great diversity in model physics, this provides much stronger evidence for this behavior than if it is found when the physics has been constrained to be the same. It indicates that in order for this result to emerge, it must be the result of a very fundamental physical mechanism that is *not* dependent on the details of physics parameterizations. In addition, model diversity in an intercomparison provides an opportunity to explain the intermodel spread, not in terms of a particular model detail, but in terms of robust physical mechanisms and theory. For example, Wing and Singh (2023) used zero-buoyancy plume theory to explain the intermodel spread in stability and humidity in RCEMIP.

That being said, we do recognize the reviewer's point that the mock-Walker simulations may develop odd behaviors (stacked overturning circulations, low-frequency variability, etc…) that likely will differ across models and complicate interpretation. We will consider pairing the proposed simulations with simplified radiation runs to aid understanding, but we have not yet figured out how this could be implemented and still satisfy the four guiding principles outlined above. Compared with simplified microphysics, simplified radiation would be easier to implement and the comparison with fully interactive radiation would be instructive. However, we worry in particular about **principle 2**, as even if they are easy to configure, requesting more simulations may reduce the ability and willingness of groups to participate. We would want to retain the simulations with full physics to satisfy **principle 3** and facilitate direct comparison between RCEMIP-I (uniform SST) and RCEMIP-II (prescribed SST gradient) simulations, and so simplified radiation simulations would be in addition to the full physics simulations, rather than instead of them. It is also not clear to us which prescription of radiation would be most appropriate. A fixed cooling rate is easy and would be consistent across models, but could result in a mean cooling rate that is different than in the simulations with interactive radiation which presents its own issues of interpretation. One way to address this is to spatially homogenize the radiative cooling at each level, or prescribe a profile of radiative cooling determined from the horizontal- and time-mean of the simulations with interactive radiation. These options would decouple the radiative cooling from the condensate and moisture and would be internally consistent, avoiding a mean bias. However, the former would still allow time dependence and both options would yield different radiative cooling in each model and may still contain sharp vertical gradients and thus may not remove stacked-overturning cells. Prescribing radiative flux divergence in temperature coordinates seems a bit more complicated to implement, though we admittedly have not tried this before. With any of these choices, we would need to perform additional test simulations, and it is not possible to complete simulations that are directly comparable with the simulations presented in the paper because the computer on which they were performed (Cheyenne) no longer exists. We are also hesitant to commit to including simplified radiation simulations unless we receive assurance from a majority of the now ~30 groups interested in participating in RCEMIP-II that they would be willing to perform these additional simulations.