Reply to Reviewer comments on
"Extending the Modular Earth Submodel System (MESSy v2.55) model hierarchy: The ECHAM/MESSy idealized (EMIL) model set-up" by Hella Garny et al.

We thank all reviewers for their very valuable reviews of our manuscript, which led to a major revision of the paper, as detailed in our response (blue) to the individual reviewer comments (black, italic) below.

In the process of the revision, we discovered an inadvertently introduced modification to the original equilibrium temperature set-up. In the formulation of the „Held-Suarez" tropospheric equilibrium temperature function, used in all the simulations presented in the manuscript, the modified formulation led to a weaker vertical temperature gradient in the tropics. We corrected the formulation to the standard set-up in the model implementation, and repeated a number of simulations. In particular the benchmark simulations presented in Section 3 are now performed with the standard set-up to ensure comparability to earlier studies. To study the sensitivity of the results on the modified versus standard set-up, a number of additional simulations was performed, and the simulations with the differing set-ups are presented in concert in Section 4.
Most of the results are qualitatively unchanged with the standard versus modified model set-up. One major difference of the two set-ups is the strength of the response of the tropospheric jet to stratospheric polar vortex changes, and we added a new subsection to discuss this result (see new Section 3.3).

Edwin Gerber (Referee)

The authors document a new idealized model configuration within the ECHAM/MESSy modeling framework, and demonstrate how it can be used to investigate open questions in the climate sciences, namely chemistry-transport interactions and the monsoonal circulation. I believe that this work is timely and important, and would be of interest to GMD readers. I therefore recommend publication pending consideration of the comments/suggestions below. As my identify might be obvious given my familiarity with the system, I'm signing this review. Ed Gerber

Thank you, Ed, for your very helpful suggestions to improve our manuscript, that we greatly appreciate.

General comments

1) The authors compare the performance of EMIL against a number of benchmark cases that are in the literature. It would be ideal, however, if we could move beyond the "picture norm" for these comparisons – at least in the future. Could you publish the data for these results (or incorporate it within the ECHAM-MESSy distribution), so that in the future, other groups could check their models against yours? The best standard would be to determine whether your integrations are consistent/inconsistent with other benchmark integrations, within the sampling uncertainty. I believe that data can be archived through Zenodo.org, or other structures. You could just provide the zonal mean time mean data needed for the figures. Another option would be to include the key benchmarks as test cases within ECHAM/MESSy, something that could easily be reproduce by another group. Could you provide a citable link to the model and the required parameter scripts? (That is, a frozen version of the model, as was used to produce this paper, ideally with the same run scripts that you used.) I appreciate that the supplement provides all the parameters, but it would still involve a lot of work (and hence many chances to make a mistake) to reproduce this exactly.

Thank you for this suggestion, we agree that it would be very beneficial to be able to do quantitative comparisons of model simulations in the future, rather than the comparison to published Figures, as we have done in our paper. Indeed, having data for comparison available would have likely prevented us from using the inadvertently modified model set-up for the last 3.5 years, as we likely would have discovered the above mentioned differences much quicker.

Thus, we decided to provide the simulation data as freely available data set (via Zenodo), to enable future users to reproduce the analyses presented in the paper.
The doi to the data set is inserted in the paper in the “data availability” section.

The ECHAM/MESSy code is available upon registration as MESSy user, and the next MESSy release (v2.55) will include the here described implementation of the idealized set-up, together with a sub-set of the namelists for the simulations presented here. Thus, the user will be able to repeat the simulations. Further, to ensure reproducibility, we include now a table with all simulations and information on the set-up (new Table B1), and further the supplement was improved to make the transition from the parameters in the presented Equations to the namelist parameters easier for future users (see improved Tabels 1-3 in supplement, see also answer to comment 3 below).

2) I appreciate that the authors have striven to find a balance between detailing a new model set-up for others to use, and presenting new results. I felt that the test cases that were shown at the end in section 5 were very interesting, but could have been more developed. To provide more space, perhaps the earlier sections could be condensed? (The reader might also be a bit exhausted by the time they reach these really interesting results!) For example, there are a lot of equations and parameters defined in this study, many which are specified in other papers (but also many of which are new). I think some of this detail could best be put in an appendix (e.g., in sections 2.1.1, 2.1.2 and 2.1.3), allowing you to move more quickly to the results.

Thank you for this excellent suggestion, we followed your advice and moved all equations to a new Appendix (Appendix A), and only kept a short description in Section 2, in which we particularly stress which formulations are new.
As we have expanded the results on troposphere-stratosphere coupling, we refrained from adding additional material to Section 5, also because we plan future publications for example on the idealized monsoon circulations.

3) It would help the reader to have a table that defines all the parameters in one place. It would also help you catch any parameters that are multiply defined. One example is k_max, which appears in the equations (8) and (9) with distinct values. k_damp is also defined inconsistently between these two equations (though any reasonable reader would understand what is meant). The parameter \delta\phi also appears in multiple equations, e.g., (5) and (16).
Finally, I noticed that \sigma is sometimes used to refer to a vertical coordinate (p/ps), and at other times used width (where I appreciate the motivation is to connect it to the variance of a Gaussian). It might be good to adopt a consistent notation, where \delta is always used for width parameters – but again watching out to make sure all parameters are uniquely defined. (This said, I know that these parameters came from multiple papers in the literature, where the other authors were not consistent with each other!)

Thank you for spotting the inconsistent naming of the parameters, we realized the parameters were also not assigned to the namelist variables in a straightforward way. To address the comment and clarify parameters, we:
- made sure to name parameters consistently and uniquely by following your advice to use sigma for the vertical coordinate and delta for width parameters, and by adding superscripts to variables used in multiple equations (see equations A1-A16).
- added to the Tables in the Supplement (Table 1 to 3) detailed descriptions of the namelist parameters, including default values and the corresponding symbol in the defining equations A1-A16. Thus, the information on the default values is removed from Appendix A, enhancing the readability of this section.
- added a table with all simulation details as Appendix B.

4) The paragraph spanning from page 3 line 28 to page 4 line 2 is very interesting, but seems out of place in the introduction. I would consider pushing this to final section, where you could present it as the next step in your research program.

Good suggestion, the paragraph is moved to the final paragraph of the paper.

5) Finally, the topic of regimes comes up quite prominently in section 4. I think this is a very interesting (albeit sometimes frustrating) result that could be mentioned in the abstract and introduction. I think these regimes have simmering in idealized models for sometime: as detailed by Gerber and Polvani (2009), the original PK02 result is so dramatic precisely because of a regime switch between their \gamma 2 and 4 integrations. Chan and Plumb (2009, DOI: 10.1175/2009JAS2937.1) and Wang et al. (2012) discuss this in more detail. The presence of regimes is interesting: if such a thing existed in our atmosphere, we could be in for surprises with global warming (or perhaps when the planet enters an ice age). If it is an artifact of these idealized models, however, it's something that the dynamics community should be wary of. It could lead to unphysical parameter sensitivity or results that are qualitatively disconnected from the real atmosphere, breaking the link we'd hope to establish through model hierarchies.

We agree that the regime behavior in the idealized models is a very interesting result, and we agree that it requires careful evaluation whether those regimes are at all relevant for our real atmosphere. Inspired by your comment, we analyzed the regime behavior in our model simulations more closely, which led to the discovery of the modified implementation in the equilibrium temperature (see above). The modified equilibrium temperature, with lower tropical upper tropospheric temperatures, led to an equatorward shift of the tropospheric jet (in agreement with studies prescribing diabatic heating in the upper tropical troposphere), and interestingly the response of the jet location to stratospheric forcing appears to be strongly damped in this set-up compared to the stndard equilibrium temperature (see new Fig. 6 (left)). We added a new subsection (Section 3.3) to the paper to discuss the stratosphere-troposphere coupling and the regime behavior of the tropospheric jet location. The changed tropospheric state in the simulations with modified set-up appears to inhibit the regime-like behavior of the tropospheric jet, possibly because the jet is located further equatorward in the basic state. Thus, while the general result that the tropospheric response is sensitive to the prescribed tropospheric equilibrium temperatures is in line with the study by Chan and Plumb (2009), the reason for the damped response seems to be a different one (in the simulations by Chan and Plumb (2009), the jet was rather located further poleward in the basic state). However, the dynamical reasons for this behavior remain to be analyzed in detail, which is beyond the scope of the present paper.

Moreover, we want to point out that next to the regimes in the location of the near-surface jet, the regimes we had addressed so far in the paper are regimes in the polar vortex strength. We added more discussion on the polar vortex regimes in Sections 4.1-4.2 (including appended Figures of probability distributions, new Figs. C1-3). More detailed analysis of the polar vortex regimes is part of ongoing work that we plan to publish in a follow-on paper.

Overall, the following changes with respect to discussion of the regimes and of stratosphere-troposphere coupling were made:
- added statement to Abstract
- added paragraph to Introduction
- new Section 3.3 and new Fig. 6
- added new Fig. 12, showing relation of tropospheric jet location and stratospheric polar vortex strength, and Figures C1-3 with probability distribution functions of polar vortex strength and free tropospheric and near-surface jet location, and discussion thereof in Sections 4.2 and 4.3
- added paragraph with discussion to Section 6

Specific comments (largely typographical) by page:line number
1:1 Consider "As models of the Earth system grow in complexity, a need emerges to connect them with simplified systems through model hierarchies in order to improve process understanding."
Done
1:3 consider cutting "with the aim"
Done

1:6 Would you consider ECHAM/MESSy a "model", or rather a "framework" which allows you to build many different models.
Good point, MESSy is definitely a framework, ECHAM/MESSy is one instance of this framework. We decided to change „model“ to „framework“ here.
1:10 Consider "Test similations with EMIL reproduce benchmarks provided by earlier dry dynamical core studies."
Done (and changed title of Section 3 accordingly to „Model benchmark tests“).
1:19 What do you mean by "the ability to simulate dynamical systems"? Dynamical systems in the broadest sense is a whole field in mathematics. Perhaps you mean "the ability to simulate qualitatively realistic dynamical variability of the circulation"
True, this was a misnomer, we rephrased to „circulation systems“.
1:22 Consider something like "Earth system models continue to incorporate more processes to enable a more complete simulation of the climate system, and thus produce the best possible climate projections. In practice, this increases the complexity of model codes as new compartments are added to represent new processes." I'm not sure if you need that second sentence; my thought was that the goal is to increase the range of processes that are simulated, and this is effected in practice by adding more compartments, modules, etc..
Done, thanks for the suggestion, we decided to keep the addition on the compartments.
2:9 stray space: "hereafter) ."
Done.
2:13 I think the upper level drag is only in the PK02 set up, and not a part of the original HS94 configuation.
True, removed „and upper level“ here.
2:16 consider a paragraph break before "The functions…"
Done.
2:21 "to idealized heating that mimics the thermal response to CO2 increase" I think "climate change" is the response, not the forcing!
True, and changed.
2:26 "motivates one to include"
Changed to „motivates the expansion of ...“
2:29 Jucker and Gerber (2017) were not the first/only one to do this. Consider also referencing: Merlis, T. M., T. Schneider, S. Bordoni, and I. Eisenman, 2013: Hadley circulation response to orbital precession. Part I: Aquaplanets. J. Climate, 26, 740–753,doi:10.1175/ JCLI-D-11-00716.1. Tan, Z., T. A. Shaw, and O. Lachmy, 2019: The sensitivity of the jet stream response to climate change to radiative assumptions, J. Advan. Mod. Earth Sys.,10.1029/2018MS001492.
Thanks for pointing those references out, we added them here.
2:35 Here and throughout the text, the quotes seem to be reversed. Perhaps this is set by the journal, but I am used to "hello" as opposed to "hello"
Thanks for spotting this, and corrected.
3:19 consider "allows the creation of model hierarchies"
Done.
3:20 consider "Earth-system model. Any developments…"
Done.
5:9 I found "idealzied localize contrained" to be awkward. Consider just "forced by a simple, localized heating that…"
Done.
eqn (1) In HS94 and other papers, it's usually just T_{eq}
Done, replaced throughout the paper.
5:30 This was a point where I feel you've lost the balance on providing enough technical advice without making the paper too long. Do you need to describe an option that "physicaly of little use" eqns (204) To make the paper more concise, you could refer the reader to HS94. I appreciate that equation (2) is modified by the inclusion of the \epsilon sin(\phi) term; this was documented by equations A3 and A4 in PK02. A happy medium might be to reference past work in the paper, highlighting your modifications, and including equations in an appendix.
As detailed above, the section has been reworked by moving equations and details to the appendix, while only a general description is left in the main part of the paper. Thanks for the suggestions.

5:19 T_{US} isn't defined in the paper. The reference is: U.S. Standard Atmosphere, U.S. Government Printing Office, Washington, D.C., 1976. (Which I appreciate isn't so easy to find!)
Thanks, and added.
eqn (6) Aditi Sheshadri did something like this in her 2015 paper, https://doi.org/10.1175/JAS-D-14-0191.1. There she lowered the start of the vortex to 200 hPa. That said, I appreciate the more thorough investigation of the transition height in this study!
Thanks for pointing this out, we added a sentence in section 4.2.
Figure 3 and surrounding discussion. It is interesting that the jets shift equatorward when you move from the T63L19 to the T42L90 integrations. I suspect the vertical resolution plays a more important role her than you might suspect. This is consistent with the behavior of GFDL's spectral core, where the jets also shift equatorward when the vertical resolution is increased. See Fig. 4 of Gerber et al. (2008), https://doi.org/10.1175/2007MWR2211.1. This doesn't seem to happen in finite difference or finite volume based cores. [This said, I don't mean for you to add another citation; I think you've already been very generous in referencing my past work.]
Interesting that there is a similar behavior in the GFDL model. Newer results of ours also indicate that the vertical resolution might be more important here than we had assumed. We changed the text to „The jets are shifted equatorward in the T42L90MA resolution, and eddy variance is generally reduced. This is likely a combined effect of lower horizontal and higher vertical resolution, in agreement with Wan2008.".
11:13 consider a paragraph break after PK02.
We chose to keep the text in one paragraph as still the same Figure is discussed.
11:14 (namely GFDL's spectral dynamica core)
Done.
Figure 5: the caption on this figure could be expanded to help a reader who's skimming the paper, for instance, defining the key parameters p_{Tw} and \gamma that are being used. I'll admit I had to remind myself what p_{Tw} represented.
Done.
13:16 Along the lines of my general comment on the "picture norm", it would be ideal to be more precise about what you meant by negligible. I think you mean that it is small relative to uncertainties in the cliamatology with resolution (i.e., T63L19 vs. T42L90), but you could also define it relative to sampling uncertainty (i.e., it would take inordinately long integrations for the difference to be significant above sampling noise.)
True, the statement on the differences due to the different sponge set-up was rather vague so far, and not based on a statistical evaluation. We added a significance test on the differences, to clarify whether the differences are negligible with respect to sampling uncertainty. We based the t-test on slices of 30-day means, assuming that there is no correlation between those 30-day time-slices (given a decorrelation time-scale in those simulations of about 30 days, this should be about right). The addition of the significance test revealed a weak (significant) downward extension of zonal wind differences into the troposphere, that we had not noticed so far. We rewrote the description of the differences to be more precise and quantitative.
Figure 6 and discussion. I appreciated this portion of the paper, but a a quick question: is one month of austral hemisphere gravity wave drag enough to nail down the effective damping rate in models? I don't have a good sense how much this rate varies. I assume this includes both orographic and non-orographic drag? Would the effective rate be much different in the boreal hemisphere during winter? I think it would help to expand the caption, to explain that GWD/u provides an effective damping time scale of the winds when using a full gravity wave drag scheme.
True, only one month of data for one hemisphere was a very thin data basis to argue with. The „effective damping time-scales" for both the NH and SH are added to the Figure now, including values from 50 winters in each hemisphere. There is considerable variability between different winters, and the damping is on average stronger in the SH (possibly because of the lower planetary wave activity?). Overall, the chosen damping time-scales in the new sponge layer implementation lie well within the variability. We also added text to the legend to be more clear.
15:4 "cannot"
Done.
Figure 8 and following figures. You could possible color the dashed curves which show the equilibrium profiles, to make the comparison with their respective \gamma's easier. For Figure 8 specifically, please specify the location of this profile. Is it right at the pole?

Done (colored dashed lines for Figures with multiple T_eq values).

16:7 consider a paragraph break after \gamma.

Obsolete due to re-writing of paragraph (emphasizing on vortex regimes).

17:3 In Wang et al. (2012), I think we had to grapple with this same regime behavior. The model switch abruptly from a state with active stratospheric variability and a strong residual mean circulation (which allows the temperature to deviate substantially from T_eq) to a state with an very cold, stable vortex near "radiative" equilibrium. In Wang et al., this regime change was associated with a substantial change in the position of the tropospheric jet. Does that happen here?

Yes, and the tropospheric jet shift is discussed in more detail now in Section 4.2 (new Fig. 12 and discussion thereof). Whether the stratospheric polar vortex regime shift and the tropospheric jet location regimes are (necessarily) connected is, to my understanding, a question to be clarified.

17:28 "these two simulations"

Obsolete, as we removed old Fig. 10 (climatologies), as we felt that they do not add much value, and wanted to compensate for the new additional Figures.

Figure 10 Here you are showing results from integrations which exhibit multiple regimes. Based on past experience (e.g., Wang et al. 2012), regime transitions can introduce very long time scales, as the model switches between states. You can see this of this Figure 5 of your text, which corresponds to pTw=400, gamma 2 integration shown in the right panel (I think.) Therefore, you have to be very careful in establishing convergence. Earlier in the text you suggested that runs were done for 1825 days; it seems that you have longer runs (3000 days are shown in Fig. 5), but I'm not sure that would be sufficient. It would be good to check/comment on the sampling uncertainty in these climatologies.

Agreed, and indeed we find very long time-scales in the simulations with regime transitions (the simulation with gamma = 2 and pTw = 400 presented in old Fig.5 has a decorrelation time scale of ~100 days). We agree that an integration length of 1825 days is thus far too short to establish convergence (indeed, for a simulation with regime transitions, convergence might be never reached). We commented on this issue at the beginning of Section 3, and further in Section 4.2. We have extended the newly simulations (with the standard set-up, see top) to ~10000 days to test for the robustness of the results, but are not able to extend all simulations to this length. However, the new extended simulations do show a generally similar behavior than the old (short) simulations, letting us believe that the sampling uncertainty does not influence our conclusions majorly.

Figure 12 and discussion. I suspect that the strength of the overturning (difference between T and T_eq) near the model top will be dominated by the drag layer. Hence, it's likely to be determined by \gamma: if you force a stronger vortex, you need a stronger drag. At lower layers, the strength of overturning is dominated by "wave pumping", and so the resolved circulation. I worried about this a lot in preparing my 2012 paper, but convinced myself that in the mid-to-upper stratosphere, the differences in the residual circulation in response to changing gamma were still being dominated by the waves, and so not an artifact of the sponge layer. I'm not exactly sure how far down you need to go to be free of the sponge layer, but perhaps 10 hPa would be a better choice than 1 hPa? This would be supported by Figure 7, where you find that the spong layer has a negligible impact below 10 hPa. I'd also be curious to see if the nonlinearity in the vortex shown in Figure 11, bottom left, shows up in the overturning at 10 hPa in the model with heating.

Thanks for pointing this out, and we decided to change the figure to show the 10 hPa results, which also provide more interesting insights – as you suspected, the non-linearity in the vortex is in agreement with the deviation from T_eq at 10 hPa.

22:18 consider a paragraph break after "high."

Done.

22:19 consider "high latitudes (north of 60N), driven by the strong wave dissipation that effected the SSW; see the red line in the top panel of Fig. 13. This transports …"

Done.

22:22 consider "latidues, evident in Fig. 13 ... 15 ms-." (no paratheses). I'd also consider breaking the paragraph after this sentence.

Done.

22:29 Isolated from what? Consider cutting "in an isolated manner," or to be more specific, e.g., "independent of the annual cycle" or "isolated from all other chemical processes".

Moved sentence to the end of the section, and added more specific statement.

Figure 13 Consider reworking the caption, as you first refer to the middle panel. It might also be nice to include a second axis on the top panel, or to make "w[10ˆ-5 hPa]" in red*
Reworked figure, and done.
23:1 Consider a paragraph break after "is steep."
We chose to keep the text in one paragraph as still the same Issue is discussed.
23:3 "downwelling is maxized at the"
Done.
23:4 same as above
Done.
24:10 consider a pargarph break after "state."
Done.
24:17 10ˆ20 J sounds like a lot, but could you provide some context? Say, what is the effective heating rate per square meter (W/mˆ2), which could be more easily compared to solar or precipitation forcing. With hope this number is in the ball park for what you'd expect from monsoon precipitation.
Thank you for this comment. We added the following sentence to the text: „This heating is of the same order of magnitude as the idealized heat source of 6×10^{19}\unit{J} prescribed in Siu and Bowman (2019) to model the North American monsoon anticyclone (see their experiments 5a-5e).” Keeping in mind e.g. that the Asian monsoon anticyclone is clearly more pronounced than the North American monsoon anticyclone, the higher energy input in our study seems reasonable.
24:19 "produced in response to the additional heating"
Done.
24:23 consider a paragraph break after "respectively.)"
Done.
24:26 Perhaps the anticyclonic centers could be marked/labeled in the figure.
Done.
24:30 You could break the paragraph after "2016).
Done.
24:30 Consider. "An example of eastward eddy shedding was observed during the second period, as displayed on the right of Fig. 16. This phenomenon has been previously investigated…"
Done.
25:7 Your summary opens with a hard sentence to parse. Consider from line 8"...model system is documented. The set-up, denoted EMIL (explain the acronym), is shown to perform consistently with established dry dynamical core benchmarks, both earlier configurations of the ECHAM core, and those developed by other modeling centers."
Thanks, and Done.
25:26 "used setups. The polar"
Done.
26:1 This is an interesting result, as we see this coupling in observations (i.e., with the ozone hole, or following an SSW). It is my understanding that the tropospheric state of the Lingren et al. (2018) model is substantially different, and might explain why does not couple to the stratosphere. As you have shown in Figure 10 (right panel), for instance, easterlies are generated in the UTLS region of the winter hemisphere.
We added substantial discussion of the tropospheric jet response in the different set-ups of the model, mentioning also that from observational evidence we expect a vertically coherent response of the tropospheric jet (which is not seen in the „Lindgren“ set-up).
26:3 consider "we present, as a proof-of-concept, a"
Done.

Reviewer Name: Penelope Maher

Summary of the Review

This manuscript describes the implementation of the Held–Suarez configuration, with the Polvani–Kuchner amendment for the stratosphere, within the ECHAM/MESSy modelling framework. From the model description, it seems the model has been implemented in a modular nature which is a credit to the modelling effort (this can be a development nightmare otherwise). The manuscript has a well described parameterisation equation set, and has tested the relevant parameter spaces for the tunable variables and compares their results with the literature. The new model set-up is then used, as a proof of concept, for looking at how CFCs impact the polar vortex and monsoon circulation.

Thank you for the positive description of our modeling efforts.

Unfortunately, we realize that a general misunderstanding arose: in the model version we describe in this paper, it is possible to analyze the impact of dynamical variability and forced changes ON tracer distributions, including diagnostic chemical tracers as shown for the CFC example (Section 5.1), but NOT the impact of e.g. CFCs on dynamics. No feedback of the chemical tracers on dynamics exists currently in the model.

Thanks to your following comments and to avoid this misunderstanding, we reworked the model description and motivation (in Abstract, Introduction and Summary), as detailed below.

In these regards the manuscript is both novel and interesting. There were, however, a number of things that I was confused about and that need further clarification or description. I also feel there are a number of figures that could benefit from further work. This manuscript is well suited for publication in GDM, is written in a way consistent with the journal style and with further revision I believe it will be suitable for publication in GMD. In this review I have used the notation "PxLy" and this should be interpreted as page x and line y.

2 Major Comments

1. The introduction is well motivated in terms of the using idealised models in general (the philosophy of idealised models), however, I think more introductory material is needed for describing the need for adding chemistry into the hierarchy and what these styles of models are used for. For example, it may not be clear to readers if/why chemistry models are needed to investigate the polar vortex or monsoons.

The motivation to implement the idealized model set-up in the framework of a chemistry-climate model system is, for the current set-up, the ability to study the impact of idealized dynamical variability and forced changes on to the distribution of (chemically active) tracers. This model set-up is motivated by a large number of research questions on the distribution of chemical substances in the atmosphere, e.g. the question how changes in the circulation in a changing climate will affect stratospheric ozone, and how important the monsoon systems are in transporting tracers from the troposphere to the stratosphere. As detailed in the paper, it will be the next step to couple the chemistry and dynamics, a task that will be possible to perform within the chemistry-climate model system framework we use. This second step is motivated by the research question on how circulation-induced anomalies in radiative trace gases (e.g. ozone and water vapor) feed back on the dynamics, a question that is relevant both on climate time-scales as well as in intra-seasonal timescales (e.g. during sudden warmings).

We added the motivation for the inclusion of diagnostic (chemical) tracers in the Introduction (see p3, line 16ff), and motivation for the next step in the hierarchy (including the coupling), that we moved to last paragraph of the paper (see also Ed Gerbers major comment 4).

2. I felt the abstract, introduction and conclusions did not sufficiently describe what is currently possible within the ECHAM vs MESSy models. I initially assumed this paper was the first to implement the Held-Suarez test case within ECHAM but realised my mistake on page 10 when the authors described the study of Wan et al 2008. I think what options are (or not) previously

available needs to be said much earlier or more clearly. I understand the RELAX submodel is new (ie implementing the parameterisations of newtonian cooling and drag), but were changes to the dynamical core needed or where they already available (if it was available, is it the same/similar as Wan et al 2008?)? I am confused by what is new and what was existing in ECHAM.
Thank you for pointing out the fact that we need to state much clearer the difference of our model set-up, i.e. using ECHAM as base model within the MESSy framework, to the original ECHAM model. MESSy is a framework that allows to link a base model (i.e. a dynamical core, here ECHAM) to submodels (e.g., physical parameterizations, diagnostics, and among others the chemistry scheme). While Wan et al. (2008) used a dynamical core version of ECHAM in their paper, this version was to our knowledge only used for testing purposes of the model core and is not part of the general model distribution of ECHAM. The implementation we performed here is new in that it was developed within the MESSy framework. One advantage of MESSy is that the implementation of the model was possible simply through the implementation of a new submodel for the relaxation, and the other physical parameterizations could be simply switched off. Thus, no changes to the dynamical core had to be made. Within MESSy, it is now possible to run the dynamical core model with the same executable as more complex versions of the model, and the idealized model set-up is available for all model users. Moreover, the full infrastructure on tracer set-up, transport and chemical reactions, available in the MESSy framework, can be exploited also with the dynamical core model.
We added text on the distinction and motivation of our implementation of the dry dynamical core model within MESSy to the Introduction (p4, l20 ff) and Abstract (p1, line 6-7).

3 Minor Comments

1. The introduction would benefit from a plane language description of ECHAM vs ECHAM/Messy (ie what is the standard GCM, atmosphere only or ECM).
We expanded the description of the MESSy framework in the Introduction, to state that ECHAM is (one possible) dynamical core used within the MESSy framework:
"The MESSy framework couples a base model (dynamical core) to submodels, that contain the physical parametrizations as well as diagnostics. Among other base models, the ECHAM dynamical core is available in MESSy." (p3, line 33)
2. The manuscript would be easier to read to non-ECHAM specialists if there was a table of acronyms with a short description of each model and where if fits in with the other options.
While we do see that the number of Acronyms are confusing (in particular for non-MESSy users), we do not feel like we can omit any of them, as the paper also serves as documentation. However, we did try to reduce the usage of the acronyms in the text as much as possible to increase readability, and explain the model framework in more detail (which we hope serves the purpose more than adding a table).
3. From my perspective, sections 1 through to 4 are describing the implementation and the validation of the code. While in section 5, the model infrastructure is now well justified to use with the chemistry models. I think at the beginning of section 5 this should be more clearly communicated to let the reader that we have reached to point of advertising why a model like this is useful.
Thanks for the suggestion, and we added a sentence at beginning of section 5.
Figures often reflect personal styles and different perspectives. I have listed quite a few changes to the figures. These are separated into changes I would like to see made (below) and suggestions which I feel would help (these can actioned at your discretion, see clarifications section).
Requesting the following changes be made to the figures/captions:
1. Fig 1: Are there four options on the y-axis or more? I found it hard to interpret this figure and I am not sure which set-up has which chemistry option. What does '/...' in the 3D dynamical core mean?
The figure caption was expanded to clarify the meaning of the axis, and the Figure was slightly reworked to emphasize on the model set-ups with /without coupled chemistry.

2. Consistent colour bars are needed. Fig 2, 3, 10 use a yellow-to-red colour bar to describe T , uv and vT. Suggest Fig 3 has different colour bar for the fluxes. The blue-to-red colour bar is used for diffs in Fig7 but for T in Fig 4 and Φ in Fig 15-16. Suggest diffs for blue-to-red, T use yellow-to-red

and another colour option for Φ etc.
We reworked these Figures according to your suggestion to use consistent colour bars (Figs. 3, 4, 15, 16 (now 17, 18))
3. Many subplots all have the colour bar repeated. Suggest only having one colour bar or legend per plot.
Done.
4. Fig 5 caption should explain PT W and y are from the legend and point to relevant equations.
We added the description of p_Tw and gamma in the caption.
5. Fig 13: The jet colour map is generally considered bad practice and I suggest a different colour. I found it hard to interpret the zonal mean zonal wind in white and it look me a while to identify what the breaks were a SSW (also is this surface wind or aloft?). On first reading I thought the top panel was divided by w so the title was confusing for me. Are both u and w essential (could it be described instead as the inverse in general)?. I suggest exploring some other formats for this plot to help draw out the features.
The plot was reworked by changing the color map, and adding an additional y-axis in the upper panel to avoid confusion on the time-series of u and w*. The caption now clarifies that the wind contours in the middle panel are displaying winds at 50 hPa.

.
4 Clarifying Comments

4.1 Figures
Suggestion the following changes be made (optional):
1. Fig 1: The y -axis title 'Chemistry' is floating in a way that it feel out of place (either remove or move). I am not sure what the purpose of the two dotted vertical lines are.
The label „Chemistry" was moved, and the caption expanded to make the purpose of the vertical lines more obvious.
2. Fig 2: The title on the fig is not helpful (suggest removing it).
Done.
3. Fig 3: I find the left plots very hard to see. Suggest moving the left panel to a new plot and then keep the flux plots together. Alternatively you could consider only plotting 0-90 in one hemisphere given they are symmetric in this case. What does the 'MA' in the caption (and text) mean? The fonts are too small (also in other plots).
The plot is redone and resized to make it more visible, The „MA" is omitted from the title to avoid confusion, and explained in the text (MA="Middle Atmosphere", i.e. high-top version of model levels).
4. Fig 4 Suggest subplot titles are larger and also included in bottom panel.
Added titles to subplots, and omitted upper panels to keep balance of number of figures (two new additions).
5. Fig 5: Suggest you use the same seaborn colours as in Fig 8 (matching the values of y in fig 8) – this assumes these are in python though I notice a mix a languages used to generate the plots which is fine.
Well spotted, there is indeed a mixture of languages due to the different habits of the authors. Most figures are reworked now (in python), as is Fig, 5. Colors of Fig. 8 (new Fig. 9) and the time-series in Fig. 5 are consistent now.
6. Fig 6 (but in general): might want to consider skipping either red or green in your line plots so everyone can easy see it. I would prefer you use the same colour choices as in Fig 7.
Usage of Red and Green lines in Fig. 6 (new 7) is omitted now.
7. Fig 7 (and text): add the equilibrium temp to the legend. What do you mean by 'plane' in titles (and text)?
The title „plane surface" is omitted to avoid confusion (without topography was meant) and the equ. temp. line is explained in the caption.
8. Fig 11-12 bottom panels: suggest using colours not already used in the top panels and they are different so as not to confuse latitude, wind speed and temp differences.
Due to the addition of the new experiments, the coloring changed and is unique in each set of panels (top versus bottom).

9. Fig 14, if this image is a a pdf/png/jpg or similar, then I would suggest replacing the error bars with filled upper and lower intervals with lower alpha values (ie shading). If your using ps/eps this won't work.
Decided to leave figure as is (it is an eps file).
10. Fig 14: the line width is not consistent (thinner is nicer) and I suggest removing titles. The choice of black gives this authority (as is commonly done for obs), was this intentional?
Fixed (thinner lines in all panels, title removed). Black was intentional as this is the base case presented in Fig. 13.

4.2 Abstract
1. P1L1: I think it would help to explain why you mean by 'a need emerges'. I know what you mean but it might help to explicitly say it.
Thank you for the suggestion. As the reason for the „emerged need" is detailed in the first sentences in the introduction, we felt that adding it to the Abstract as well would blow up the Abstract too much. Also, we feel that the addition „to improve process understanding" does explain already why we need the model hierarchies.
2. P1L2: I would suggest a more general description instead of 'process understanding', perhaps ' simulations of the climate'
As detailed above, we think that the reason for the simplified models is exactly the seek for process understanding, while the complex models are the ones that are used for the best possible „simulations of the climate". Therefore, we would rather keep the formulation as is. See first paragraph of introduction.
4.3 Introduction
1. P2L8-13: In terms of the Held-Suarez description, a reader could easily get confused about a models dynamical core vs the parameterisation set-up of HS. Suggest rewriting L8-13 to make it clearer that HS was designed as a test for the dynamical core.
True, and reformulated to make it more clear that we mean a „Held-Suarez" type model here, not the particular functions for the equilibrium temperature they propose.
2. P3L2: The sentence starting 'The motivation of the MESSy framework was' is an excellent sentence that helped ground me in the context of the configurations. Could I requestion you add this (word for word is fine) to the abstract (something similar is already in the abstract but not as clear) and something just as cleanly described for the EMIL.
Thanks for the suggestion, we revised the paragraph in the Abstract to describe the MESSy hierarchy and the motivation for the implementation of the dry dynamical core model within this framework more clearly:
"The Modular Earth Submodel System (MESSy) was developed to incorporate chemical processes into an Earth System model. It provides an environment to allow for model configurations and set-ups of varying complexity, and as of now the hierarchy reaches from a chemical box model to a fully coupled Chemistry-Climate model. Here, we present a newly implemented dry dynamical core model set-up within the MESSy framework, denoted as ECHAM/MESSy IdeaLized (EMIL) model set-up. EMIL is developed with the aim to provide an easily accessible idealized model set-up that is consistently integrated in the MESSy model hierarchy. The implementation in MESSy further enables the utilization of diagnostic chemical tracers."
3. L9: Could you include what MECCA stands for? Is MECCA the chemistry model of EMAC or it more subtle?
Yes, MECCA is the chemistry module used in EMAC. The full name is added to the text.
4.4 Model Description
1. What surface conditions are you using? Is it generally an aquaplanet with 'water' mountains or does have a land like surface heat capacity?
As in the dry dynamical core model, the only interaction with the ground is via the prescribed friction, the ground does not actually have any heat capacity (it is implicitly included in the prescribed equilibrium temperature).
2. Equ 2: Worth mentioning the extra term ($\varrho \sin \varphi$) that is not in the original HS formulation is from PK.
We added the remark that the asymmetry term was added is from the PK study (see equ. A1, moved to appendix in response to Reviewer Ed Gerber).

3. P6L12: suggest replacing (40κa)−1 with 0.025κ−1 s and (4κs) −1 with 0.25κ−1 s
The default values have been omitted from the text, and moved to the new Table 1 in the supplement.
4. Equ 8: Why use σ0 here and σb in equ 3?
As those are two distinct parameters, we chose to keep the distinct labeling to be able to keep them apart.
4.5 Model test cases, sensitivity simulations and application examples
1. Fig 4: titles and caption have inconsistent window for the averages, were they 10 or 11 years?
True, indeed 11 years as on the Figure titles were used. However, as we have chosen to remove the panels showing ERA-Interim climatologies to compensate for the addition of new Figures, this is obsolete.
2. P15L1: might help to define UTLS and what it's acronym is
Thanks and done.
3. P24L5: a key task of what?
Rephrased.
4. P24: I think you should state early in section 5.2 that the monsoon simulations are run with the chemistry scheme. I think you should also say if this is usual, and if not, why is the chemistry scheme is helpful.
In the monsoon simulations presented in the paper, no chemical tracers were included (see clarifications at top of review). However, for future applications, we plan to use diagnostic tracers to study tracer transport in the idealized model framework, one of the major research questions about monsoon circulation systems (see addition to the Introduction, new p3, line 20 ff).
5. Sect 5.2: Is there temporal variability in these simulations? If it was said then I missed it.
The forcing term is set constant, so there is no variability in the forcing. This is described in the second paragraph of Section 5.2.
4.6 Summary and Outlook
1. P25L7: Suggest removing 'In the paper presented here,' (it works better without it).
Done.
2. L2511: suggest replace 'based on the suggestions by HS94 and PK02' with 'as described in HS94 and PK02'. I would then start a new sentence that described what is new in your implementation.
Done.
3. P25L25-27: I found the description of 'climate states' confusing. I also suggest rewriting this sentence (or even multiple sentences) as the grammar has gotten complicated.
Done.
4. Please add in the acknowledgements where the SPARC and Era-I data can be downloaded from.
Done (for SPARC, ERA-I is not used directly anymore).

5 Editorial comments
There were a number of times where latex has compiled with " instead of " (please review).
Done.
5.1 Title and abstract
I think the title is too technical. By their very nature GMD papers are technical but I think you could make your title easier to read/understand (and remove some of the acronyms where possible).
We tried to reduce the usage of acronyms in the text of the paper, including the Abstract, which hopefully is now much more readable also to non-MESSy -users. However, we think the title needs to include the model system/ set-up names (indeed, it is requited by GMD to include version numbers), so we decided to keep it as is.
1. I found the number of acronyms hard to digest. Given the subject matter, I think ECHAM, MESSY and EMIL are probably fine to use but I would suggest removing RELAX and EMAC as they are not essential. I also think referring to the model as only EMIL or ECHAM/Messy idealied model would help readability. There is a lot of switching between model names which makes it hard to read at times for non-ECHAM experts.
We agree, and as said above, we tried to reduce the usage of the acronyms in the text, and hopefully explain the nature of the model framework more clearly now. However, we don't think we can remove any of the acronyms completely, as they are important informations to the model

users.
2. I don't think you need to citations in the abstract. I think it is fine to say Held-Suarez model.
References are omitted.
5.2 Introduction
1. P1L1 'more and more processes and compartments' is awkward, suggest changing to 'increasing complexity' or similar
Rephrased (see Ed Gerbers review).
2. P2L13 (or there abouts) suggest stating that the HS model will be described in detail in section 2.
Done.
3. P2L32: The description 'currently underway' reads as though these models are yet to be released. Suggest rewording as both Isca and CESM are broadly used for idealised studies.
Rephrased to „are aiming to...".
4. P3L3: suggest adding '0d' before box model.
Done.
5. P3L22-27: I very much liked this paragraph. I would suggest moving it earlier in the introduction. Maybe even as the first paragraph.
Thanks, and we moved to paragraph to an earlier place (before the whole MESSY framework is explained, to make it clear why we use it).
6. P3L28: What do you mean by 'consistent' ?
With „consistent", we here refer to a model hierarchy which adds processes in a logical order to study chemistry-dynamics interactions. The paragraph has been moved to the end of the paper, where it is put more in context.
7. P3L28: I like the description of model hierarchy of chemistry-dynamical coupling and I suggest you use this more often (esp in abstract).
See above, the paragraph has been moved to the very end of the paper, to make it more clear that this is an outlook.
8. P3L28-P4L2: Suggest moving this paragraph to outlook section.
Done.
9. P4L5: replace 'to' with 'two' and suggest separating into two sentences, one that talks about section 3 and one for section 4.
Done.
10. P3L6: Suggest joining the two paragraphs that describe what is coming up in the paper.
Done.

5.3 Model description
1. P5L12: Suggest adding convection to the list of parameterisations.
Done.
2. P5L13: suggest you mention Held-Suarez in this sentence.
Done.
3. P5L17-20: Suggest you add these as dot points rather than a list in a sentence.
Done.
4. P5L25: should cite HS in here.
As this is the generic equation for temperature relaxation, we don't think the reference in necessary here – it is cited again two sentences later.
5. P5L27 (and else ware): suggest replace 'local' with 'environmental'.
Changed to „actual temperature"
6. P5L28: suggest adding this paragraph to the one before and list dot points for the ways of implementing κ and Tequ
Obsolete due to restructuring of section.
7. P6L12: The discussion on hf rac should start with a description of ϱ.
Done.
8. P6L14: suggest replacing 'sign' with ±
We keep the „sign" function here, as this function is returning the sign of the given parameter.
9. P8L15: suggest replacing 'employed' with 'added'.
Done.

10. P8L18: suggest replacing 'reads' with 'is given by'
Kept as is to avoid duplication of „given by"

5.4 Sensitivity simulations
1. P16: Fig 11-12 are referred to before 9-10. You might want to consider moving figs or mentioning 9-10 earlier.
Consistent order of Figures is ensured now.
2. P17L6 and Fig 9: Could you explain why there are multiple Lindgren lines on these plots (which variables are changed)?
The differences in the set-up (i.e. in the Equilibrium temperature) are explained in the second paragraph of the section, and in the figure legends.
5.5 Supplementary
The tables would benefit from latex hlines and vlines so they look more like tables. Suggest removing the quotations from all variables. I don't think Fig1-2 are needed but I do not feel strongly about this. Fig 3 is a nice aid to see the call sequence (well done).
The supplement, including tables, has been revised.

Anonymous Referee #3

The authors introduce a new idealized and modular modeling setup and demonstrate its use in a couple of ways. I believe the paper would benefit from some restructuring – the paper goes back and forth between model setup issues (choices of values for various parameters) and scientific results which could potentially be a bit confusing to a reader. Perhaps the authors might wish to consider splitting the manuscript in two?
We thank the reviewer for the suggestion. Following Ed Gerber's review, we restructured the paper to enhance readability (moved considerable part of the technical descriptions to Appendix), and added more scientific discussion on the results (namely, stratosphere-troposphere coupling and dynamical regimes). We agree with the reviewer in that the scientific results might be expandable, and a second paper on the dynamical regimes of the polar vortex, that we find in the parameter sensitivity experiments, is in preparation.

Would it be possible to set up a github with a downloadable version of the model? I have doubts about reproducibility which the availability of the model would help to dispel.
The MESSy model is available upon obtaining a licence (see code availability section), so a freely downloadable version at github is not possible. As detailed in response to Ed Gerber's review, the next model release will contain sample namelists for the experiments conducted in the current study, so that the reproducibility will be ensured. Further, we added a new table with the specifications of the simulations, and the supplement contains now more detailed instructions for setting the parameters. Also, we decided to provide the data of the presented experiments via zenodo to enable future comparisons to other models.

On the science front, I think the authors are up against some regime issues in dynamical core models, which it would be good to clarify. The original PK02 model shows a very large response to stratospheric perturbations in comparison with observations, and in the absence of a quantitative theory of how stratospheric perturbations affect the troposphere, responses of the model when planetary scale waves are forced by topography are not necessarily the "correct" response.
We agree, and added discussion on regimes and stratosphere-troposphere coupling in our simulations (for details, see response to Ed Gerbers general comment 5).

On readability, the manuscript would benefit from some proofreading and fixing of minor typos (in particular, the quotation marks all appear reversed?).
We implemented the suggested changes by the other reviewers, and revised the manuscript with fixing all typos etc. we were able to identify.