University of Victoria
PO Box 3065
Victoria BC V8W 3V6 Canada

Tel (250) 721-6120
Fax (250) 721-6200
Email seos@uvic.ca
Web www.seos.uvic.ca

School of Earth and Ocean Sciences

Dr Colin Goldblatt
Associate Professor
Tel (250) 472 4060
Email czg@uvic.ca

4th July 2017

Dear Dr Hargreaves,

I enclose the revised version of our manuscript, which has been revised in response to reviewers comments. Both the paper, and the online supplementary material, are revised. In addition to the scientific comments, there was a lot of “i”s to dot, and “t”s to cross (particularly in the supplement) – which took a long time because of personnel changes in my group.

There is an extra author on the paper now. All authors have given their explicit consent to the revised author list.

I have appended my responses to review, and additionally a graphical diff made with LaTeXdiff. One of the changes in the manuscript was the addition of a giant table which, unfortunately, looks a bit messy after LaTexdiffing. For clarity: that table is all new, so one can simply consult the revised manuscript.

Thank you again for considering this manuscriopt.

Yours sincerely,

Colin Goldblatt

Colin Goldblatt

Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-24-AC1, 2017

Geoscientific
Model Development
Discussions

Interactive comment on "The Palaeoclimate and Terrestrial Exoplanet Radiative Transfer Model Intercomparison Project (PALAEOTRIP): experimental design and protocols" *by* Colin Goldblatt and Lucas Kavenagh

Colin Goldblatt and Lucas Kavenagh

czg@uvic.ca

Received and published: 22 June 2017

We thank Reviewer 1 for a helpful review of our experimental protocol. We provide a full response to the reviewers comments (reproduced in italics) below, together with revisions to the manuscript.

This is a well written paper describing the protocol for an intercomparison between radiation codes which will be useful for the Palaeoclimate and Terrestrial Exoplanet modelling communities. The paper addresses a relevant scientific modelling question within the scope of GMD ; namely how to assess the skill of radiation codes used for a wide range of conditions which may be outside of those for which they were originally developed. It presents a modelling protocol that is suitable for addressing this question, involving the submission of outputs from radiative transfer codes run with standardised inputs covering a range of conditions. These results will be compared with reference calculations from 'line-by-line' codes and published in a subsequent paper. The concepts involved are not particularly novel, being an extension of the approach used in previous studies; nonetheless they are useful because they will be applied to a wider range of radiation codes and conditions than previously. The results from this intercomparison will help modellers in the community to select the radiation codes best suited to their purposes, and to improve others, and so will be likely to result in substantial advances in modelling science. The methods and assumptions are valid and clearly outlined, but unfortunately the description is not sufficiently complete and precise to allow the protocol to be executed by a modelling group. I think the paper would be suitable for publication once the major issues below are addressed. I also list some minor issues which would be good to address.

We address specific issues discussed raised here below.

Major issues:

1/ Page 9 I think that your timetable is very unrealistic. The amount of time you're giving people to send in their results is so small that there is a real risk of not getting sufficient participation to maximise value of this activity to the community. I would recommend seeking advice from other similar projects and coming up with something more realistic. I would have thought the modellers would need at least 6 months to send their data, and it would be a good idea to allow additional time for you to spot any errors in the data or its formatting and to allow them to resubmit.

Sober reflection and plain passage of time has us agree on this point. We have relaxed the timescale very substantially, with the goal of a summer 2018 completion, not summer 2017.

2/ The supplementary information needs to be improved. The text of the paper reads "We have provided MATLAB and Python codes which will write them automatically from your output. These scripts, and sample output files, are available at www.palaeotrip.org and included in the supplementary information for this paper." In spite of what the text says, the SI does not contain any sample output files. These are needed because the text doesn't explain the full naming convention. These aren't available on the website either at the time of writing, but even if they were, the protocol is supposed to be fully described by the paper. Also, the list of input files in incomplete; for example there is no specification for the Stellar Spectra in the SI or on the website. Please also list the files individually in the readme file and say what each of them are. For example palaeotrip_profiles.mat file seems to be a binary and I have no idea what it is.

The SI is revised, completed, and fully described in a readme file.

Minor issues:

1/ The abstract provides a concise and complete summary. A minor point however is that I found the aim to "constrain the ranges of far-from-modern atmospheric compositions in which the codes perform well" a bit unclear. Why would you want to constrain the ranges over which the codes perform well? I can understand why you would want to identify those ranges, and subsequently to allow to community to expand those ranges. Constraining them makes little sense to me.

We have changed "constrain" to "identify".

2/ The overall presentation well generally well structured and clear. I did however find the list of experiments on page 5 somewhat redundant given the more informative and detailed list which appears on Page 6. I would recommend merging the list on Page 6 into the list on Page 5. If you want an ?at a glance? summary of the experiments then I would add a table.

The descriptions have been merged into a new table, Table 1.

3/ Page 1 Line 1 time on the order of
Fixed.

4/ Page 3 Line 25. It might be nice to explain why a single global mean profile is considered sufficient. Previous intercomparisons have used profiles from different regimes/seasons, e.g. McClatchey Mid-Latitude Summer/Winter etc.

We now say “For simplicity, all experiments use a Global Annual Mean (GAM) profile”. Others would be nice, but this simple approach is probably sufficient. There isn't much more to say!

5/ Page 3 Line 31 number/name, a brief
6/ Page 5 Line 9 is expected
7/ Page 6. Please be consistent - i.e. Experiment 1 / Experiments 11.
9/ Page 7 Line 2 which is best done
10/ Page 7 Line 8 with a
11/ Page 7 Line 9 consisting of the
12/ Table 1 diffuse

All fixed.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-24, 2017.

Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-24-AC2, 2017

Geoscientific
Model Development
Discussions

Interactive comment on "The Palaeoclimate and Terrestrial Exoplanet Radiative Transfer Model Intercomparison Project (PALAEOTRIP): experimental design and protocols" *by* Colin Goldblatt and Lucas Kavenagh

Colin Goldblatt and Lucas Kavenagh

czg@uvic.ca

Received and published: 22 June 2017

We thank Reviewer 2 for an exceptionally detailed review of our experimental protocol. We provide a full response to the reviewers comments (reproduced in italics) below, together with revisions to the manuscript.

The manuscript introduces a new intercomparison project for radiation codes, PALAEOTRIP, aimed at evaluating the accuracy of and differences between radiation codes applied to study paleo- and terrestrial exoplanet climates. As the field of modelling climates of planets significantly different from the present-day Earth matures, it will be increasingly important to evaluate the accuracy of the radiation schemes used, as radiative transfer is one of the most important components of climate models. I commend the authors for taking the initiative to begin such an intercomparison.

Thank you! We hope this project will be useful to the community.

I found the manuscript to be well written, and the different experiments to be explained in sufficient detail. My main concerns are with the large number of different runs proposed (> 200), that important parts of the parameter space are not included, and that the experiments including clouds will lead to differences between codes that may not be errors and have no distinction from conditions found on present day Earth. I discuss my concerns in more detail below, and recommend publication once my major concerns have been addressed.

We address these points in more detail below. In summary...

To facilitate broad participation, the authors could consider adopting an experiment design similar to that proposed for CMIP6, with a core set of experiments that all participating groups are expected to do, and with remaining experiments organised in terms of increasing optionality: http://www.mpimet.mpg.de/en/communication/news/single-news/?no_cache=1&tx_ttnews%5Btt_news%5D=606.

In section 2.1, we now say:
"Participating groups should run the experiments that their models are configured for, and omit any which are not possible (or onerous) to run. We do not expect groups do perform model development in order to participate in this project. For example, a model which had the solar spectrum hard coded and did not include N2O absorption would run experiments 1, 2a–b, 3–6 and 13–16. A model without clouds would omit experiments 13–16. If, for any reason, there is a limit to the number of experiments that a group can run then experiments 1–6 should be considered "core" and prioritized. A minimal set of experiments would be 1 and 2. "

Main comments:

1) Experiment 2: In this experiment, well-mixed greenhouse gas (WMGHG) amounts are varied. Only one gas is varied at a time, with the rest kept at standard conditions, amounting to a total of 113 different runs. I have several suggestions on how this experiment could be improved:
1.1) I think the number of experiments here is unnecessarily large, which may put off some potential participants. I think the number of gas concentrations per log unit can be reduced to one or two without losing a significant amount of information.

We have reduced the number to two per log-unit. This makes the lead author a bit jittery to have such course spacing, but I accept it is probably for the best.

1.2) The maximum N2O amount seems quite large to me, I am curious about how the authors arrived at this number (1e-2 volume mixing ratio). Also, some radiation schemes may not include N2O, so it might be worth having some experiments without N2O to facilitate broader participation.

We made this up. No-one has any idea of what Archean or Proterozoic N2O levels were (I say this as someone who works on the early nitrogen cycle!). Fluxes may well have been quite high given incomplete denitrification in suboxic environments. Even a first order estimate would be a good paper, but is beyond the scope here. We have set the upper bound high to be inclusive.

As we emphasize further now in section 2.1 (see above), groups should run experiments which they can, a model without N2O should run 2a and 2b, but not 2c. As N2O absorbtion would be omitted in the standard run as well as experiments 2 and higher, then comparison of differences between test and standard conditions would still be meaningful.

1.3) As only one gas is varied at any given time, a significant part of the parameter space is not being considered, including the authors example of a typical late Archean

atmospheric composition in the introduction where both CO2 and CH4 amounts are elevated compared to present day Earth. I think it would be beneficial to add some runs with compositions that have previously been used in climate models of the Archean Earth.

We have included a new experiment with overlap as experiment 3 for solar, and 9 for M-star. To keep the setup simple, this simply replaces standard background with a nominal set of high background levels.

1.4) Should these experiments include oxygen and ozone? The Archean atmosphere is thought to have had very little oxygen and ozone, including some experiments without these absorbers may be useful.

Indeed, the Archean atmosphere was around 1ppmv O2 and no O3, whereas Phanerozoic O2 and O3 levels are essentially modern. Thus there is a dilemma about what levels to use. For the purpose of an intercomparison, however, our focus in on simple experiments on gas addition, relative to standard conditions. Therefore, we have kept O2 and O3 in when changing most WHGHGs. There is the issue of overlap of course, but this should be minor: O3 absorption overlaps with CO2 only in the thermal region.

In summary, I would encourage the authors to significantly reduce the number of WMGHG amounts in this experiment and also to include other, very common compositions such as atmospheres where both CO2 and CH4 amounts are elevated compared to present day Earth.

2) Experiment 3: The water vapour mixing ratio could be reduced even further in this experiment, perhaps by a factor of 0.01 or 0.001. Five mixing ratios per log unit may also be unnecessarily many, this could potentially be reduced to two or three. Also, planets receiving large near-IR fluxes may have very large stratospheric water vapour mixing ratios (up to 1e-3), the authors could consider adding an experiment with a modified water vapour profile where the stratospheric water vapour amount is elevated compared to present day Earth.

We have reduced the minimum factor to 0.01.

A specific moist stratosphere experiment would be interesting, but this would be more complicated to set up and we are motivated (and advised by this reviewer!) to keep the number of experiments simple. The experiments with high water vapour with the M-star spectrum will offer some guidance here. Thus we do not wish to add an additional experiment.

3) Experiment 5: Why have the authors decided to turn off oxygen absorption while still having ozone absorption turned on? This test also moves the upper boundary to a higher or lower pressure. In particular, the largest surface pressure will move the upper boundary to 1 mbar, which is a rather large pressure. I would recommend defining a few separate P-T profiles with varying surface pressures (but constant upper boundary pressures) to use for this test instead of simply multiplying the values in the GAM profile with a constant factor.

Re GHG concentrations: the motivation in this experiment was to keep amounts of each absorber constant. For minor species, this is easy to achieve, as described. However, it is obviously impossible for oxygen when surface pressure is reduced strongly. Therefore, oxygen absorption is turned of in all for self-consistency. There is a minor loss of physical realism (though oxygen absorption is minor), which is justified because the motivation is self-consistent inter-comparison of codes, not accurate climate prediction.

Re upper boundary pressure, in the standard case the difference in flux between 0.1 and 1mbar is $< 0.1\,\mathrm{W\,m^{-2}}$ in all radiation streams (sample output now in SI), therefore we do not expect this to cause a problem.

4) Experiments 6-8: WMGHG amounts are not varied for the experiments using an M-star spectrum. CO2, and particularly CH4, are significant near-IR absorbers. It would be very interesting to see how well the different radiation schemes deal with the overlapping absorption in the near-IR between H2O, CO2 and CH4 for cases with large amounts of CO2 and CH4. Any errors in this region can become significantly larger with an M-star spectrum compared to that obtained with a Sun-like star spectrum due to the large near-IR flux.

Fair point. So, we have added additional experiments to make the suite of experiments for an M-star spectrum identical to those for the solar spectrum. We considered further picking-and-choosing, but the simpler approach of duplicating all seemed easier (all participatiing groups should then have to do is change the spectrum for each, so all GHG experiments can be done for both spectra).

5) The temperature-pressure profile is kept the same in all experiments (except in experiment 5 where it is scaled to achieve a smaller/larger surface pressure). For the stellar (short-wave) component of the radiation I would not expect errors in most radiation codes to depend strongly on temperature (except if temperatures become high enough to warrant the use of high temperature line lists). Errors in the thermal (long-wave) radiation, however, can depend on the temperature due to the shift in the peak of the Planck function with temperature, which will emphasise different wavelengths. It may be worth adding another experiment where the temperature is varied within a reasonable range to see how well codes deal with somewhat lower and higher temperatures than those found on present day Earth.

This would be an interesting experiment. However, we have added other experiments in response to this review, and also have the mandate to keep the total number of experiments low.

6) Experiments 9-12 involve adding a low or high altitude cloud to the setup of experiment 1 and vary the water path or cloud particle size. Currently these experiments feel somewhat out-of-place:
6.1) The motivation for including these experiments is not clear from the current manuscript. The other experiments are designed to test how well radiation codes perform for conditions potentially significantly different from present day Earth. In these

experiments conditions are similar to those found on present day Earth, and most approximations used have been tested for these conditions by present day Earth climate modellers (see e.g. Oreopoulos et al. 2012, http://dx.doi.org/10.1029/2011JD016821, Barker et al. 2015, http://dx.doi.org/10.1175/JAS-D-15-0033.1). I think a stronger motivation for these experiments is required.

We now add to the manuscript:
"There a range of good choices of representation of cloud microphysics in models (i.e. which are different but entirely reasonable), so variation in the radiative effects of clouds may arise from these rather than error per se. Nonetheless, it is of primary interest to us how the radiative effects of clouds do vary when every attempt has been made to specify cloud physical properties equivalently."

6.2) Experiments 11-12 include ice clouds with a prescribed effective size Deff with optical properties from Baum et al. (2014). For several participating groups this may involve implementing new ice cloud scattering properties in their radiation codes, solely for the purpose of participating in this intercomparison. I think it may be too much to ask groups to do this, and results would not directly reflect those obtained in the respective climate models.

This is a misunderstanding of our intention, so we have improved the clarity of the manuscript. We add:
"We emphasize that the normal implementations of clouds in participant models should be used; single scattering properties are provided only for cases where this necessarily needs to be input. "

6.3) It is not clear how the benchmark results will be defined and obtained in these tests. Different and entirely reasonable choices with regards to e.g. the size distribution of cloud particles may result in differences between radiation codes that cannot be considered to be errors as in the other experiments. This should be discussed in more detail.

We now state in the manuscript:
" There a range of good choices of representation of cloud microphysics in models (i.e. which are different but entirely reasonable), so variation in the radiative effects of clouds may arise from these rather than error per se. Nonetheless, it is of primary interest to us how the radiative effects of clouds do vary when every attempt has been made to specify cloud physical properties equivalently."

6.4) The number of different runs may also here be unnecessarily large, 54 in total. I would suggest reducing it to about three runs per experiment (e.g. with a low, medium and high value) to ease participation.

We have reduced it to 6 or 7 runs per experiments, total 25. In our opinion, three would simply not be enough.

I my opinion these points will need to be addressed in order to justify including Experiments 9-12 in this intercomparison.

All the points are addressed above.

Minor comments:

7) The abstract and introduction paints a rather negative view of the current state of radiation codes used to study paleo- and terrestrial exoplanet climates. While it is true that the accuracy of several radiation codes remains unevaluated, at least in the literature, there have been some work to address this. Examples are Wolf & Toon (2013) (dx.doi.org/10.1089/ast.2012.0936), who evaluated the accuracy of their new radiation scheme by comparing it to the LBLRTM, and Yang et al. (2016) (dx.doi.org/10.3847/0004-637X/826/2/222), who evaluated differences between several radiation schemes when applied to the inner edge of the habitable zone. These works should be mentioned and referenced.

Now cited.

8) Introduction, first paragraph, first sentence ("A typical model of ..."): One or more *references are needed. Also, giving gas amounts in units of pressure is ambiguous (a gas? contribution to the surface pressure and the gas? partial pressure at the surface are generally different). Please consider using ppmv for all gas amount units, or clarify which pressure is used.*

We've changed ppm to ppmv for clarity, and used percent for CO2, and changed "A typical model of" to "An example model of". There isn't one or a few good references for the nominal composition chosen, and it won't help maters to put in a few paragraphs of justification here - it is really just an example to set the tone.

9) Introduction, second paragraph: The statement that deriving the surface temperature for a given atmospheric composition and incident flux is conceptually a simple physics problem is somewhat oversimplifying the problem. Uncertainties in e.g. ground albedos, cloud physics and ocean heat transport (with a potentially unknown land/ocean distribution) can potentially impact surface temperatures significantly. In my opinion this discussion should be modified to argue for why performing accurate radiative transfer is both important and difficult, while at the same time acknowledging that other uncertainties remain.

We have added in "surface properties specified"; in respect to other points we beg to be allowed some artistic licence in motivating the experiment.

10) Introduction, second paragraph: In my experience line-by-line calculations can, with a reasonable number of layers (? 40), take as little as a few minutes for a single column. Still several orders of magnitude too slow for use in a GCM, but not as bad as indicated.

We now say "minutes to hours".

1) Introduction, third paragraph: A statement is made that atmospheric composition is equivalent to column abundances. This is strictly speaking not correct as gas mixing ratios are 3D fields, while column densities are vertically integrated fields.

We have revised this to remove the erroneous statement of equivalence: "To optimize efficiency, these parameterizations may made for limited ranges of atmospheric composition or column abundances of absorbing molecules."

12) Section 2.1, second paragraph: To argue for why H2-dominated atmospheres are not included, it is stated that the altered mean molar weight would lead to different pressure-broadened line shapes. While it is indeed true that H2 pressure-broadened widths are different from air-broadened widths, this is not only due to H2 molecules being lighter than air molecules; calculating pressure-broadened line widths is a rather complicated quantum-mechanical problem. Please reformulate.

We now say: "One class of model atmospheres that we exclude is H2 dominated atmospheres (Wordsworth, 2013), as air-broadened line shapes will likely not be appropriate and thus a majority of codes may not perform well (that is, these atmospheres require rather specialist treatment, beyond the scope of this intercomparison). "

13) Section 2.2.1: I assume the mixing ratios provided online with the GAM profile are volume mixing ratios, but I could not find this specified anywhere. Also, it would be nice if the GAM profile could be specified on both levels and layers to avoid potential slight inconsistencies between codes.

These are indeed volume mixing ratios. This is now specified at 2.2.1

We have additionally provided layers in the SI.

14) Section 2.2.3: Will the supplied stellar spectra be normalised such that, integrated over wavelength, they give the TOA flux to be used in the experiments? Otherwise the TOA flux will need to be specified.

Yes, and a solar constant is now additionally specified.

15) Section 2.2.5: The effective temperature of the surface is missing.

Now specified.

16) Section 2.3: Currently, the list of experiments is provided twice, one on the form of an overview and one as a list with details on each experiment. I understand why, but to me this seems a bit awkward. I would consider making a large table with details on the different experiments to provide a better overview, and refer to this in the main text when discussing them.

Fair point. We have moved all of this into “One Table to bring them all and in the lightness bind them”.

17) Section 2.4, second paragraph: Consider moving the definition of layers and levels to section 2.2.4 as they are used there.

Done.

18) Please consider adding more references to recent radiation intercomparisons, e.g.: Oreopoulos et al. (2012): http://dx.doi.org/10.1029/2011JD016821 Pincus et al. (2015): http://dx.doi.org/10.1002/2015GL064291

These are now cited.

9) From statements in section 2.2.2, and 2.4, I deduce that benchmark results from line-by-line codes are meant to be submitted along with results from other radiation codes. Please make this more clear.

We state clearly:“For line-by-line models, spectrally resolved output should be subsampled to 1 cm?1 resolution. Contact the PALAEOTRIP project team directly to discus how to submit this (info@palaeotrip.org).” The point is that LBL output may be too large for the online submission system that we have.

Typos
-Page3,line31: "aa" –"a"
- Page 4, line 4: "and and" – "and an"
- Page 6, line 22: Runcode for experiment 8 should be PT8_x.
- Page 7, lines 8-9: "an ten line" – "a ten line"

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-24, 2017.

The Palaeoclimate and Terrestrial Exoplanet Radiative Transfer Model Intercomparison Project (PALAEOTRIP): experimental design and protocols

Colin Goldblatt, Lucas Kavenagh, and Maura Dewey
School of Earth and Ocean Sceinces, University of Victoria, Victoria, BC, Canada.
Correspondence to: Colin Goldblatt (czg@uvic.ca and info@palaeotrip.org)

Abstract. Accurate radiative transfer calculation is fundamental to all climate modelling. For deep palaeoclimate, and increasingly terrestrial exoplanet climate science, this brings both the joy and the challenge of exotic atmospheric compositions. The challenge here is that most standard radiation codes for climate modelling have been developed for modern atmospheric conditions, and may perform poorly away from these. The palaeoclimate or exoclimate modeller must either rely on these or use bespoke radiation codes, and in both cases rely on either blind faith or *ad hoc* testing of the code. In this paper, we describe the protocols for the Palaeoclimate and Terrestrial Exoplanet Radiative Transfer Model Intercomparison Project (PALAEOTRIP) to systematically address this. This will compare as many radiation codes used for palaeoclimate or exoplanets as possible, with the aim to ~~constrain~~ identify the ranges of far-from-modern atmospheric compositions in which the codes perform well. This paper describes the experimental protocol and invites community participation in the project through ~~2017.~~ 2017–2018.

1 Introduction

Earth's atmospheric composition has varied dramatically through time, and yet-to-be-discovered terrestrial exoplanets will add untold diversity. A ~~typical~~ example model of late Archean atmospheric composition ~~, for example,~~ would be of 30,000 ~~ppm~~ ppmv CO_2, 1000 ~~ppm~~ ppmv CH_4, no oxygen or ozone and an unknown nitrogen inventory, whereas escape from 'snowball Earth' glaciation may take ~~0.1 bar~~ 10% CO_2. A fundamental part of the palaeoclimate problem, and equivalently the exo-climate problem, may be stated as: given some atmospheric composition, what was the energy balance of the planet? Or~~equivalently:~~, for given atmospheric composition and incident solar flux, what was the surface temperature?

This is a conceptually simple physics problem. An atmospheric composition and structure needs to be given, surface properties specified, then the radiation field must be simulated, the equations for which are well known (e.g. **?**). Regrettably, implementation is far from simple. Millions of gas absorption lines from numerous gases are relevant to the climate problem. Herculean work has assembled most of these into large and oft-revised databases (e.g. **?**). From these databases, absorption cross sections may be calculated as a function of temperature and pressure. Even these cross sections, calculated with standard assumptions regarding the shape of absorption lines, have some notable disagreement with observations and smoothly varying "continuum" absorption must be added to produce realistic cross sections. Armed with cross sections, the radiative transfer equations may then be solved at the natural resolution of the lines—a so-called line-by-line calculation. Alas, these can take time on the order of ~~the order one to ten~~ minutes to hours for a single column, hence are too slow by many orders of magnitude to be used in a climate model.

In a general circulation model (GCM), the radiative transfer for a single column must be evaluated in a fraction of a second. Consequently, simplifications must be made in the treatment of the radiative transfer and the spectral dependence must be heavily parameterized. To ~~optimise~~ optimize efficiency, these parameterizations ~~are~~ may made for limited ranges of atmospheric composition or ~~, equivalently, for certain~~ column abundances of absorbing molecules. Often, these parameterizations were made a decade or more ago, with poor documentation. Where an older (and likely faster) GCM is used for palaeoclimate research, one is automatically in the situation of using a legacy radiation code.

At the other end of the modelling spectrum, there still exists a cottage industry of bespoke development of fast-enough radiative transfer codes for deep palaeoclimate, planetary atmospheres, or other obscure radiative transfer problems, where all the required steps are made ad hoc. However, ~~this can mean that~~ in some cases, the resources required to sufficiently test the code are unavailable locally.

Three broad classes of problem arise. First, whilst excellent parametrization is possible within design ranges, some parameterizations do not perform as well as a third-party user may hope. For example, intercomparison of radiation codes used for the IPCC Forth Assessment Report (**?**) showed that many codes simulated the changes due to a doubling of carbon dioxide poorly. Second, performance of codes will decrease outside design ranges, which often includes the regions that we are interested in for palaeoclimate (e.g. **?**). Third, errors are made in parameterizations (especially in bespoke codes) which can remain undetected through review and for some years afterwards.

The palaeoclimate or exoplanet modeller is thus in a bind. The science interest is in novel atmospheric compositions, whose radiation properties are outside the intuition of most non-specialists. It would be prudent to test *any* fast radiation code that one planned to use against a well-trusted line-by-line code across the parameter space of interest (e.g. **???**); however, doing this requires both the specialist knowledge in radiative transfer, the local availability of such a ~~mode~~ model and a lot of time and energy. All of these can be hard to come by.

With the Palaeoclimate and Terrestrial Exoplanet Radiative Transfer Model Intercomparison Project (PALAEOTRIP), we hope to alleviate this problem. Our aim is to test a large number of fast radiation codes, both GCM and bespoke, against line-by-line models for a wide range of conditions applicable to palaeoclimate and terrestrial exoplanet research. Such intercomparison studies have a long history in application to modern conditions and anthropogenic global change ~~(e.g. ????)~~ (e.g. **??????**) and have contributed markedly to improvements in the fidelity of radiation codes and thus the robustness of climate models. Our hope is that exporting such systematic intercomparison to deep palaeoclimate and exoplanets will yield similar improvements. In this paper, we describe the experimental design and protocol[1]. Up-to-date project information will be available at www.palaeotrip.org throughout the project.

---

[1] Community input on the experimental design and protocols ~~can, at this stage, be provided via~~ was gathered during the open peer review process “discussion” phase of *Geoscientific Model Development*.~~These will be considered final in post review manuscript.~~

2 Experimental Design

2.1 Philosophy

Our hope it that by assembling and analyzing results from many radiative transfer codes outside of modern conditions, we will both help future investigators to make an educated choice of which radiative transfer code is applicable for a particular experiment, and inform model developers of opportunities for improvement of models.

The standard method of radiative transfer intercomparison is to compare model output—especially changes in fluxes in response to changes atmospheric composition—calculated on *fixed* atmospheric profiles. The use of fixed profiles is essential to isolate the fidelity of the radiative transfer codes (to be evaluated) from the myriad of other processes that determine the atmospheric profile. This methodology has a long history (e.g. **??**); see **?** for an in-depth discussion of this methodology. We use instantaneous (unadjusted) radiative forcings~~; the~~. The most modern radiative transfer intercomparison project for IPCC class models (**?**) additionally use *effective radiative forcings* that account for a variety of rapid adjustments in GCMs; these are not included here. Our method here corresponds to the (**?**) assessment of "parametrization error".

~~Twelve sets of experiments ,~~Three groups of experiments are included, addressing changes to clear-sky properties under both a solar and a M-star spectrum, and adding clouds under the solar spectrum. These give fourteen experiments in total, each of which varies a parameter of key importance for palaeoclimate and Earth-like exoplanets~~, should be performed~~. The choice of parameter space represents a range of mainstream assumptions about atmospheric composition through Earth history. We have explored all of this parameter space previously: see **?** and **??** for well mixed greenhouse gases, **??** for clouds and **?** for varying atmospheric pressure. One class of model atmospheres that we exclude is H_2 dominated atmospheres (**?**), ~~because variation in mean molecular weight will mean that~~ as air-broadened line shapes ~~may~~ will likely not be appropriate ~~and thus~~ for these, consequently a majority of codes may not perform well (that is, these atmospheres require rather specialist treatment, beyond the scope of this intercomparison). ~~Our hope it that by assembling and analyzing results from many radiative transfer codes outside of modern conditions, we will both help future investigators to make an educated choice of which radiative transfer code is applicable for a particular experiment,~~

Participating groups should run the experiments that their models are configured for, and omit any which are not possible (or onerous) to run. We do not expect groups do perform model development in order to participate in this project. For example, a model which had the solar spectrum hard coded and did not include N_2O absorption would run experiments 1, 2a–b, 3–6 and 13–16. A model without clouds would omit experiments 13–16. If, for any reason, there is a limit to the number of experiments that a group can run then experiments 1–6 should be considered "core" and prioritized. A minimal set of experiments would be 1 and ~~inform model developers of opportunities for improvement of models.~~ 2.

All of the required input files for the project are available at www.palaeotrip.org, and as an online supplement to this paper.

2.2 Model atmosphere

2.2.1 Atmospheric profile

~~All~~ For simplicity, all experiments use a Global Annual Mean (GAM) profile. This based on a profile derived from averaging of reanalysis data by **?**. This specific profile should be used, and none substituted for it. We refer to model levels are the boundary between model layers. Experiments 1–4, 6–8 and 10 use the GAM profile unmodified, whereas experiments 5 and 9 modify it as described for experiment 5.

Radiatively active species in the atmosphere are CO_2, CH_4, N_2O, H_2O, O_3 and O_2. All mixing ratios are in parts per volume. Standard mixing ratios are 0.21 for O_2, and vertically resolved profiles supplied with in the GAM profile for H_2O and O_3. For the remaining gases, referred to as well-mixed greenhouse gases (WMGHG), mixing rations are supplied in Table 1.

2.2.2 Line data

Line-by-line codes should use line data from HITRAN2012 (e.g. **?**).

Bespoke, GCM and legacy radiation codes will use a variety of line data. It is acceptable to submit either the most current/standard version, or a variety of versions corresponding to different applications. The model version number/name, a ~~a~~ brief description and/or link to the full description should be included as metadata with the model output, especially the version number/name of the code.

2.2.3 Stellar fluxes

Stellar fluxes are supplied for both the Sun and ~~and~~ an example M-star (ADLeo) ~~. Where a solar flux is specified ad-hoc, these should be used.~~ for models in which these are input directly.

As with line data, ~~these may be pre-set in some bespoke, legacy and GCM codes : use the standard~~ for codes which use a standard stellar flux, use this standard configuration and include whatever description possible. ~~Where~~ For such codes, where it is impractical to modify the stellar flux to an M-star, perform experiments ~~1–5 and 9–12~~ 1–6 and 12–16 only.

All experiments should use an integrated stellar flux (solar constant) of $1360\,\mathrm{W\,m^{-2}}$.

2.2.4 Clouds

Experiments with both low and high clouds are included. Calculations should be done with a single profile, with a cloud fraction of unity. Clouds may be specified in different ways in different radiation codes; the nominal descriptions here should be matched as well as possible given how clouds are specified in the particular radiation code, and appropriate description provided as metadata. We emphasize that the normal implementations of clouds in participant models should be used; single scattering properties are provided only for cases where this necessarily needs to be input. There a range of good choices of representation of cloud microphysics in models (i.e. which are different but entirely reasonable), so variation in the radiative effects of clouds may arise from these rather than error *per se*. Nonetheless, it is of primary interest to us how the radiative effects of clouds do vary when every attempt has been made to specify cloud physical properties equivalently.

Vertical position: if clouds are specified in a layer, low clouds should be in the ~~800–825~~900–925 hPa layer, high clouds in the 250–300 hPa layer. If they are specified on levels, they should be at ~~812.5~~912.5 hPa and 275 hPa, and can be specified with minimal vertical extent (or extent not exceeding the boundaries of the layer).

Low clouds are taken to be made of liquid water droplets. Thus cloud particles are well described as Mie spheres, so consistent specification across models should be straightforward. A standard low cloud should have a water path of ~~40~~$W = 40$ g m^{-2} and effective radius of 10 μm (?). ~~Single scattering properties~~For codes which require single scattering properties, output from a Mie code ~~can be provided by the PALAEOTRIP team if requested~~is provided in the supplementary information (for simplicity, a single particle radius of $r = r_{\mathrm{eff}}$ is used) and the Henyey-Greenstein phase function should be used.

High clouds are taken to be made of ice crystals, and are thus more complicated to describe, as there are a variety of ice habits which are all non-spherical. The normal parameter to describe the size of particles is the effective diameter, D_{eff}. A standard high cloud should have a water path of ~~20~~$W = 20$ g m^{-2} and effective diameter of ~~70~~80 μm (?). ~~Where~~ For codes which require single scattering properties~~need to be specified, these should use~~, these are taken from the “general habit mixture” ~~from ? which are available for all sizes required from .~~ of ?, see also http://www.ssec.wisc.edu/ice_models/polarization.html are provided in the supplementary information and the Henyey-Greenstein phase function should be used.

For codes which specify cloud thickness via an optical depth τ, this can be calculated directly from the extinction efficiency, Q: $\tau = \pi r^2 n Q$ where n is the number of cloud particles in the column. n is found directly as $n = W/m$, where for liquid droplet mass is found $m = \rho(4/3)\pi r^3$, given density ρ and for ice droplets m is supplied.

2.2.5 Miscellaneous details

A solar zenith angle of 60° should be used for all experiments.

The surface should be black for thermal calculations and have a grey albedo of 0.12 in solar calculations. If a combined solar-thermal calculation is performed, the separation between solar and thermal albedos should be at 3 μm.

2.3 ~~Experiments~~

~~In addition to standard conditions, four sets of clear-sky experimentsapplicable to palaeoclimate should be performed:~~

1. ~~Standard conditions~~
2. ~~Varying the concentrations of well-mixed greenhouse gases (WMGHG).~~
3. ~~Varying the amount of water vapour.~~
4. ~~Switching atmospheric ozone and oxygen on/off.~~
5. ~~Varying the surface pressure.~~

The surface temperature is 288.24 K in all experiments.

~~A further three clear-sky experimentsapplicable to exoplanets should be performed, with the solar spectrum is replaced with an M-star spectrum. This is relevant because it will be relatively easy to find and observe planets receiving the same insolation as Earth around M-stars, but this stellar spectrum is significantly to the red of the solar spectrum (so less Rayleigh scattering but more water absorption are expected):~~

1. ~~Standard conditions, with M-star spectrum.~~
2. ~~Varying the amount of water vapour, with M-star spectrum.~~
3. ~~Varying the surface pressure, with M-star spectrum.~~

~~Four experiments with clouds should be performed, to test the representation of cloud radiative properties. All should use standard conditions with a solar spectrum:~~

1. ~~Inclusion of a low level (e.g. stratus) cloud, with fixed droplet size distribution and varying cloud water path.~~
2. ~~Inclusion of a low level (e.g. stratus) cloud, with fixed cloud water path and varying droplet size distribution.~~
3. ~~Inclusion of a high level (e.g. cirrus) cloud, with fixed droplet size distribution and varying cloud water path.~~
4. ~~Inclusion of a high level (e.g. cirrus) cloud, with fixed cloud water path and varying droplet size distribution.~~

~~**General Notes:** Experiments 1–4~~Note that, for most experiments, ~~6–8 and 10–12 all use the Global Annual Mean temperature-pressure (~~a literal interpretation of the changes to atmospheric conditions will imply some physical inconsistencies: there is no change in atmospheric pressure when CO_2 mixing ratio increases to 10^{-1}, water vapour may become super-saturated, there is no change to the T-p ~~) profile. Climatological profiles of water vapour and ozone are supplied. Experiments 5 and 9 use modifications of the GAM profile. Experiments 6–8 use the M-star spectrum, all others use the solar spectrum.~~ profile when gas concentrations change. These inconsistencies are tolerated, with the philosophy of designing simple and easy to compare experiments which test the fidelities of the radiation codes, which is best done on fixed profiles.

2.3 Experiments

The experiments are described in Table 1.

The runcode is a unique identifier for each run, which should be used as the name of the output file for each run (e.g. `runcode.dat`). These all begin `PT` (for palaeotrip, and to avoid starting a filename with a number), followed by the number of the experiment and the run number (x) within each experiment, counting from the lowest value of any quantity varied.

Table 1: Description of experiments

Expt #	Paremeter	Value/Desription
1 ~~(Standard Conditions):GAM profile with climatological water and ozone abundances. Volume mixing ratios of WMGHGs:~~	Name	Standard Conditions
	Description	-
	Spectrum	Solar
	Profile	GAM
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, ~~and~~ 1×10^{-6} N_2O ~~. Oxygen mixing ratio is 0.21, and the remainder of the atmosphere is nitrogen. Runcode :~~
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	`PT1` ~~.~~
~~**Experiment 2:**~~	Num. of runs	1
2	Name	WMGHG variation
	Description	The concentration of each WMGHG ~~should be~~ is varied in series (ranges below), with the other two held at standard conditions~~). The ranges used should be: CO_2 from 10^{-9} to 10^{-1}, CH_4 from 10^{-9} to 10^{-2} and N_2O from 10^{-9} to 10^{-2}. The~~ . The lower end of each range is selected for minimal radiative significance of that gas (see **?**). The upper limit is an arbitrary guess at an upper bound for an Earth-like planet. Models should be run with concentrations evenly spaced in log units, with ~~five~~ two runs per one log unit (e.g. $\{1\times10^{-9.0}$, ~~$1\times10^{-8.8}$, $1\times10^{-8.6}$, $1\times10^{-8.4}$, $1\times10^{-8.2}$,~~ $1\times10^{-8.5}$, $1\times10^{-8.0}, \ldots\}$).~~Runcodes:~~
	Spectrum	Solar
	Profile	GAM

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
	WHGHG	(a) CO_2 from 10^{-9} to 10^{-1}, 1×10^{-6} CH_4, 1×10^{-6} N_2O (b) 400×10^{-6} CO_2, CH_4 from 10^{-9} to 10^{-2}, 1×10^{-6} N_2O (c) 400×10^{-6} CO_2, 1×10^{-6} CH_4,N_2O from 10^{-9} to 10^{-2}
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	`PT2a`_x, `PT2b`_x, `PT2c`_x for CO_2, CH_4 and N_2O respectively.
~~Experiment 3~~ ~~The~~	Num. of runs	$17+15+15=47$
3	Name	WMGHG variation, high background
	Description	The concentration of each WMGHG is varied in series, with the other two held at high conditions potentially representative of the Archean: $30,000\times10^{-6}$ CO_2, 300×10^{-6} CH_4, 30×10^{-6} N_2O. Otherwise, as experiment 2.
	Spectrum	Solar
	Profile	GAM
	WHGHG	(a) CO_2 from 10^{-9} to 10^{-1}, 300×10^{-6} CH_4, 30×10^{-6} N_2O (b) $30,000\times10^{-6}$ CO_2, CH_4 from 10^{-9} to 10^{-2}, 30×10^{-6} N_2O (c) $30,000\times10^{-6}$ CO_2, 300×10^{-6} CH_4,N_2O from 10^{-9} to 10^{-2}
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	`PT3a`_x, `PT3b`_x, `PT3c`_x for CO_2, CH_4 and N_2O respectively.
	Num. of runs	$17+15+15=47$
4	Name	Water vapour variation
	Description	The water vapour mixing ratio ~~should be~~ is changed by a constant factor, with all other gases as standard conditions. The ~~factors to use should be 0.1 to 10~~range of factors is $0.01 < x < 10$, which correspond to the differences ~~between~~ a range of saturation vapour pressures from ~~255~~230 K to 330 K. Models should be run with concentrations evenly spaced in log units, with ~~five~~ four runs per one log unit. ~~Runcodes: `PT3_`~~
	Spectrum	Solar
	Profile	GAM, altered water vapour profiles
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
	Runcode	PT4_x
~~**Experiment 4**~~ ~~Runcodes: PT4~~	Num. of runs	13
~~**Experiment 5**~~ 5	Name	Surface pressure variation
	Description	The surface pressure is varied between 0.1 and 10 bars. This is done by multiplying the pressure vector in the GAM profile by a factor $0.1 \leq y \leq 10$, and dividing mixing ratio vectors of minor absorbing species (CO_2, CH_4, N_2O and O_3) by y so that the mass of each absorber is conserved. Absorption by atmospheric oxygen should be turned off, because the mass of this absorber cannot be conserved at low pressure. Models should be run with y evenly spaced in log units, with ~~five~~ four runs per one log unit.~~Runcodes:~~
	Spectrum	Solar
	Profile	GAM with modified pressure.
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3.
	Clouds	None
	Runcode	PT5_x
~~**Experiments 6**~~	Num. of runs	9
6	Name	No oxygen or ozone absorption
	Description	Absorption by atmospheric oxygen and ozone should be turned off, with all other conditions as standard. Note there is no change to the T-p profile.
	Spectrum	Solar
	Profile	GAM
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O
	Clouds	None
	Runcode	PT6
	Num. of runs	1
7	Name	Standard Conditions, M-star spectrum

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
	Description	As experiment 1, ~~with Mstar spectrum . Runcode:~~ M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	~~`PT6`~~`PT7`
~~**Experiments 7 As experiment**~~	Num. of runs	1
8	Name	WMGHG variation, M-star spectrum
	Description	As experiment 2, M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM
	WHGHG	(a) CO_2 from 10^{-9} to 10^{-1}, 1×10^{-6} CH_4, 1×10^{-6} N_2O
		(b) 400×10^{-6} CO_2, CH_4 from 10^{-9} to 10^{-2}, 1×10^{-6} N_2O
		(c) 400×10^{-6} CO_2, 1×10^{-6} CH_4,N_2O from 10^{-9} to 10^{-2}
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	`PT8a_`x, `PT8b_`x, `PT8c_`x for CO_2, CH_4 and N_2O respectively.
	Num. of runs	$17 + 15 + 15 = 47$
9	Name	WMGHG variation, high background, M-star spectrum
	Description	As experiment 3, ~~with Mstar spectrum . Runcodes:~~ M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM
	WHGHG	(a) CO_2 from 10^{-9} to 10^{-1}, 300×10^{-6} CH_4, 30×10^{-6} N_2O
		(b) $30,000 \times 10^{-6}$ CO_2, CH_4 from 10^{-9} to 10^{-2}, 30×10^{-6} N_2O
		(c) $30,000 \times 10^{-6}$ CO_2, 300×10^{-6} CH_4,N_2O from 10^{-9} to 10^{-2}
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
~~Experiments 8 As experiment~~	Runcode	~~PT7~~PT9a_x, PT9b_x, PT9c_x for CO_2, CH_4 and N_2O respectively.
	Num. of runs	$17+15+15=47$
10	Name	Water vapour variation, M-star spectrum
	Description	As experiment 3, M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM, altered water vapour profiles
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3, O_2
	Clouds	None
	Runcode	PT10_x
	Num. of runs	13
11	Name	Surface pressure variation, M-star spectrum
	Description	As experiment 5, ~~with Mstar spectrum . Runcodes:~~ M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM with modified pressure.
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O, O_3.
	Clouds	None
	Runcode	~~PT7~~PT11_x ~~.~~
~~Experiments 9~~	Num. of runs	9
12	Name	No oxygen or ozone absorption, M-star spectrum
	Description	As experiment 4, M-star spectrum substituted for solar spectrum.
	Spectrum	M-star
	Profile	GAM
	WHGHG	400×10^{-6} CO_2, 1×10^{-6} CH_4, 1×10^{-6} N_2O
	Absorbers	CO_2, CH_4, N_2O, H_2O
	Clouds	None
	Runcode	PT12

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
	Num. of runs	1
13	Name	Low cloud, thickness variation
	Description	A low altitude water cloud is added to the standard profile (experiment 1)~~. Water path should vary~~, and the liquid water path varied between 10 and 100 g m^{-2}.
	Spectrum	Solar
	Profile	GAM
	WHGHG	400×10^{-6} CO$_2$, 1×10^{-6} CH$_4$, 1×10^{-6} N$_2$O
	Absorbers	CO$_2$, CH$_4$, N$_2$O, H$_2$O, O$_3$, O$_2$
	Clouds	Water cloud, effective radius 10 μm, water path {10, ~~with 11 values evenly spaced in log space (10.00, 12.59, 15.85, 19.95, 25.12, 31.62, 39.81, 50.12~~15, ~~63.10~~25, ~~79.43~~40, ~~100.00). Standard effective radius of 10~~63, 100} ~~μ~~g m~~. Runcodes:~~ $^{-2}$
	Runcode	~~`PT9`~~`PT13`_x ~~.~~
~~**Experiments 10**~~	Num. of runs	6
14	Name	Low cloud, effective radius variation
	Description	A low altitude water cloud is added to the standard profile (experiment 1)~~. Fixed water path of 40 g m^{-2}. Effective radius should vary~~, and the effective radius varied between 5 and 25 μm~~in 1~~.
	Spectrum	Solar
	Profile	GAM
	WHGHG	400×10^{-6} CO$_2$, 1×10^{-6} CH$_4$, 1×10^{-6} N$_2$O
	Absorbers	CO$_2$, CH$_4$, N$_2$O, H$_2$O, O$_3$, O$_2$
	Clouds	Water cloud, effective radius {5, 7.5, 10, 12.5, 15, 20, 25} μm~~steps. Runcodes:~~, water path 40 g m^{-2}
	Runcode	~~`PT10`~~`PT14`_x ~~.~~
~~**Experiments 11**~~	Num. of runs	7
15	Name	High cloud, thickness variation
	Description	A high altitude water cloud is added to the standard profile (experiment 1)~~. Water path should vary~~, and the ice water path varied between 10 and 100 g m^{-2}~~, with 11 values evenly spaced in log space. Standard effective diameter of 70~~.
	Spectrum	Solar
	Profile	GAM

Continued on next page

Table 1 – *Continued from previous page*

Expt #	**Paremeter**	**~~Experiment~~ Value/Desription**
	WHGHG	400×10^{-6} CO$_2$, 1×10^{-6} CH$_4$, 1×10^{-6} N$_2$O
	Absorbers	CO$_2$, CH$_4$, N$_2$O, H$_2$O, O$_3$, O$_2$
	Clouds	Ice cloud, effective diameter 80 μm~~. Runcodes:~~, water path {10, 15, 25, 40, 63, 100} g m^{-2}
	Runcode	~~`PT11`~~`PT15`_x~~.~~
~~**Experiments 12**~~	Num. of runs	6
16	Name	High cloud, effective radius variation
	Description	A high altitude water cloud is added to the standard profile (experiment 1)~~. Fixed water path of 20 g m^{-2}. Effective diameter should vary~~, and the effective diameter varied between 20 and 120 μm~~in 10~~.
	Spectrum	Solar
	Profile	GAM
	WHGHG	400×10^{-6} CO$_2$, 1×10^{-6} CH$_4$, 1×10^{-6} N$_2$O
	Absorbers	CO$_2$, CH$_4$, N$_2$O, H$_2$O, O$_3$, O$_2$
	Clouds	Water cloud, effective diameter {20, 40, 60, 80, 100, 120} μ~~m steps. Runcodes: `PT12`_~~m, water path 25 g m^{-2}
	Runcode	`PT16`_x~~.~~
	Num. of runs	6

~~Note that, for most experiments, a literal interpretation of the changes to atmospheric conditions will imply some physical inconsistencies: there is no change in atmospheric pressure when CO$_2$mixing ratio increases to 10^{-1}, water vapour may become super-saturated, there is no change to the T-p profile when gas concentrations change. These inconsistencies are tolerated, with the philosophy of designing simple and easy to compare experiments which test the fidelities of the radiation codes, with is best done on fixed profiles.~~

2.4 Submission of results

To facilitate comparison of many codes, each of which undoubtedly has its own output format, we ask that contributing scientists reformat output into the standard plain text format described below. ~~Not only are these~~ These formats simple, ~~but~~ and we have provided MATLAB ~~and Python~~ codes which will write them automatically~~from your output~~. These scripts, and sample output files, are available at www.palaeotrip.org and included in the supplementary information for this paper.

For spectrally integrated output (dimensions W m^{-2}) the PALAEOTRIP data format consists of a plain text file with ~~an ten~~ a twelve line header that includes the metadata in Table 3 followed by the data header describing each column, consisting the ~~the~~ variables in Table 2. Each data column is twelve characters long~~and therefore all quantities will be rounded to twelve~~

Table 2. Model output that will be accepted by PALAEOTRIP.

Variable	**Description**	**Unit**
Quantities on levels (bold variables are required):		
plevel	pressure on levels (layer bountaries)	Pa
Fswdndir	direct solar flux down	W m^{-2}
Fswdndif	~~difuse~~ diffuse solar flux down	W m^{-2}
Fswdn	total solar flux down (Fswdndir+Fswdndif)	W m^{-2}
Fswup	solar flux up	W m^{-2}
Fswnet	net solar flux	W m^{-2}
Flwdn	thermal flux down	W m^{-2}
Flwup	thermal flux up	W m^{-2}
Flwnet	net thermal flux (Flwdn-Flwup)	W m^{-2}
Quantities on layers (all should be included if any are)		
player	pressure at layer centre	Pa
Qsolar	solar heating rate	K day^{-1}
Qtherm	thermal heating rate	K day^{-1}

~~figures~~. The formatting codes accept model output that corresponds to either pressure levels or layers and will automatically distinguish between these (levels are the boundary between model layers). Quantities on layers and levels will be exported to separate data files but in both cases the first column will correspond to the pressure at the level or centre of the layer in pascals. The filename convention is `runcode_levels.txt` and `runcode_layers.txt` (e.g. `PT2a_1_layers.txt`, `PT2a_1_layers.txt`).

For spectrally resolved output ~~(dimensions W m^{-2} μm^{-1})~~ other than from line-by-line models, where available, a separate file should be provided for each flux, with pressure levels as rows and each spectral bin as a column. ~~Two rows of column headers should give the minimum and maximum wavelength~~ There should be a twelve line header that includes the metadata in Table 3 and a field with the flux name, then column headers of the spectral bin ~~. Other aspects of the output files should be as the spectrally integrated fluxes .~~ edges in microns and the dimension of the column. The bin edges should be those native to the model. The fluxes in each bin should be provided in W m^{-2} (that is the integrated flux within that spectral bin). If layer properties are provided, they likewise should be integrated within each bin such that heating rates are in K/day for each bin. The filename convention is `runcode_variable.txt` (e.g. ~~`PT2a_1_layers.txt,`~~`PT2a_1_Fswdn.txt`).

All model output should be put into a single `.zip` file called `yourname_model.zip` and can be uploaded via the palaeotrip website. Include a `readme.txt` file as necessary.

For line-by-line models, spectrally resolved output should be subsampled to 1 cm^{-1} resolution. Contact the PALAEOTRIP project team directly (info@palaeotrip.org) to discus how to submit this~~()~~, as it will likely have too large a file size for our online submission system.

Table 3. Model metadata to be included with PALAEOTRIP submissions.

Variable	Metadata Description
runcode	String with the code of run (see experiment descriptions)
modelname	String with the name (and version number) of model
username	String with your name (e.g. 'Colin Goldblatt')
useremail	String with your email (e.g. 'czg@uvic.ca')
usernotes	String with any notes about this run

Table 4. Proposed PALAEOTRIP timeline

Timeframe	Activity
January 2017	Submit description/protocol paper
January – ~~March~~ June 2017	Review of description/protocol paper. Community feedback on experimental design.
~~February~~ May – ~~March~~ June 2017	Respond to review of protocol paper and finalize protocol.
~~April – July~~ July – August 2017	
	Final protocol published
August – December 2017	Contribution of radiative transfer model runs.
January – April 2018	
	Nag participants for contributions.
May – ~~August 2017~~ July 2018	Analysis of model output by PALAEOTRIP team.
~~August~~ July – ~~September 2017~~ August 2018	Write results paper, circulate to co-authors.
~~October 2017~~ September 2018	Co-author comments.
~~November 2017~~ October 2018	Revise and submit results paper.

3 Protocol and information for contributors

The final experimental design and protocols for the PALAEOTRIP are described in this paper[2]~~.~~. These were revised following formal review and informal discussion during the *Discussion* phase of the manuscript. If you intend to submit model output to the PALAEOTRIP project, we ask that you register your intention at www.palaeotrip.org or contact us directly. This will ensure that models are not run in duplicate by different groups, and that your model output is expected.

The anticipated timeline of the project is in Table ~~3.~~ 4. Sadly, few deadlines survive contact with academics, but we hope that this schedule is realistic and it is our intention to keep ~~a tight schedule~~to it. We will post any updates to www.palaeotrip.org and communicate schedule changes directly to all participating scientists.

We intend that everyone submitting unique model results will be offered authorship on the final paper. Lead authorship will be by one of the project team, who will additionally determine the order of authorship (likely project team followed by contributing scientists, listed alphabetically). This paper will be circulated amongst all co-authors prior to submission.

---

[2] ~~Community input on the experimental design and protocols can, at this stage, be provided via the open peer review process "discussion" phase of *Geoscientific Model Development*. These will be considered final as described in the post review manuscript.~~

A motivation of this project is to find out how a variety of radiation codes perform across a range of conditions applicable to palaeoclimate and exoplanets, so that future model users may know the range of conditions across which each model is likely to be accurate. Therefore, it is essential that models are able to be identified in the final paper. The analysis will be restricted to the range of conditions specified here, as an indicator of performance in palaeo- and exoclimate studies. We have no interest in, or intention of, commenting on the fitness of any model for any other purpose. It is the responsibility of scientist submitting model results to assert that the model can be identified in the final paper.

4 Summary and Discussion

PALAEOTRIP will run ~~twelve~~ fourteen controlled experiments addressing the radiative transfer through a subset of conditions expected through Earth's past climate, and applicable to Earth-like exoplanets. We invite community participation in the experiment. Over the course of the next year, the model runs will be performed and compared. The anticipated outcome is that the community will be better informed about the performance of available radiative transfer codes for palaeo- and exoclimate research.

The range of conditions which we have specified experiments for is somewhat "vanilla". It likely does not represent the full range of conditions seen in Earth's past, and will be a tiny fraction of the parameter space for Earth-like exoplanets. This is motivated to get wide participation; that is to specify conditions which most models which derive from Earth atmospheric sciences should be capable of being run for. We anticipate that, if this intercomparison is successful, we may be able to lead a more wide-ranging intercomparison in the future.

5 Code and data availability

A zip file containing the GAM profile~~and~~, scripts to be used to write model output into the specified format and sample output is available in an online supplement to this article. Version ~~0.9 (i.e. beta version)~~ corresponds to this ~~version, in *Geoscientific Model Development Discussions*. Version 1.0 will accompany the final peer reviewed manuscriptin *Geoscientific Model Development*~~manuscript. Updated versions ~~of these~~ will be made available through the project website, www.palaeotrip.org, as necessary.

Final model output will be available from www.palaeotrip.org and as an online supplement to the paper which will describe the results of the intercomparison.

Author contributions. CG has designed the experiment and is responsible for the scientific content herein. LK and MD have helped prepare materials and provided technical assistance.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Thanks to Tony del Genio, Christos Matsoukas, Robert Pincus, Yun Yang and two anonymous reviewers for discussion and comments on the manuscript and draft protocol. Financial support has been provided by ~~and~~ a NSERC discovery grant and UVic startup funds to CG.