General Circulation Model Errors Are Variable across Exoclimate Parameter Spaces

Pushkar Kopparla1,6, Russell Deitrick1, Kö, Kevin Heng1,2,3, João M. Mendonça4, and Mark Hammond5

1 Center for Space and Habitability, University of Bern, Bern, Switzerland; pushkarkopparla@gmail.com
2 Department of Physics, University of Warwick, Coventry, UK
3 Ludwig Maximilian University, University Observatory Munich, Scheinerstrasse 1, D-81679 Munich, Germany
4 National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark
5 Department of Geosciences, University of Chicago, Chicago, IL, USA

Received 2021 May 10; revised 2021 October 1; accepted 2021 October 4; published 2021 December 9

Abstract

General circulation models (GCMs) are often used to explore exoclimate parameter spaces and classify atmospheric circulation regimes. Models are tuned to give reasonable climate states for standard test cases, such as the Held–Suarez test, and then used to simulate diverse exoclimates by varying input parameters such as rotation rates, instellation, atmospheric optical properties, frictional timescales, and so on. In such studies, there is an implicit assumption that the model works reasonably well for the standard test case will be credible at all points in an arbitrarily wide parameter space. Here, we test this assumption using the open-source GCM THOR to simulate atmospheric circulation on tidally locked Earth-like planets with rotation periods of 0.1–100 days. We find that the model error, as quantified by the ratio between physical and spurious numerical contributions to the angular momentum balance, is extremely variable across this range of rotation periods with some cases where numerical errors are the dominant component. Increasing model grid resolution does improve errors, but using a higher-order numerical diffusion scheme can sometimes magnify errors for finite-volume dynamical solvers. We further show that to minimize error and make the angular momentum balance more physical within our model, the surface friction timescale must be smaller than the rotational timescale.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (487); Exoplanet atmospheric variability (2020); Computational methods (1965)

1. Introduction

General circulation models (GCMs) are a powerful tool to explore the space of possible climates on exoplanets, especially to capture three-dimensional, multicomponent, and nonlinear interactions within the climate system that are hard to study with simpler, low-dimensional models (Kaspi & Showman 2015). One of the main takeaways from exoclimate parameter space studies has been that transitions between different flow regimes can be understood by examining trends in circulation with parameters such as planetary rotation rate, strength of irradiation, or drag (for e.g., Mitchell & Vallis 2010; Pascale et al. 2013; Pierrehubert & Hammond 2019).

Modern GCMs can produce atmospheric circulations that are qualitatively similar to observations for solar system planets (Read et al. 2016), but there remain many unsolved challenges particularly in predicting time variability, the most famous instance being the decades-long effort to accurately model climate change on Earth (Eyring et al. 2016). GCMs are imperfect models limited by grid resolution and finite time steps, employing parameterizations for complex processes that are either too computationally expensive to represent accurately or not well understood physically. Therefore, there are significant omissions, biases, and uncertainties involved (Heng et al. 2011; Amundsen et al. 2014). In the presence of high volumes of observational data, such as in the case of the Earth, GCMs can be improved iteratively to yield the right spatial and temporal variability (Koutoulis et al. 2016). For exoplanets, time variability is now beginning to be observed, such as the hot Jupiters Kepler-76b (Jackson et al. 2019), WASP-12b (Hooton et al. 2019), and the super-Earth Cancri 55e (Tamburo et al. 2018). The spatial and temporal patterns associated with this variability could help us calibrate our exoclimate GCMs, but sufficiently high-resolution observations may never be possible for many exoplanets, given the large uncertainties in physical properties (Heng & Showman 2015). Thus, it is not clear how well our GCMs can capture the processes that drive exoclimate variability.

When direct model validation is not possible due to data scarcity, we rely on intermodel comparisons to understand if the model results are credible, whether it be between a GCM and a hierarchy of simpler models (Liu & Showman 2013; Zhang & Showman 2017) or between multiple GCMs using standard test setups (Heng et al. 2011; Polichtchouk et al. 2014), such as the well-known Held–Suarez test (Held & Suarez 1994) used commonly for Earth system models. In case of disagreements between two GCMs, Held–Suarez-type standard tests cannot judge which GCM is more correct because they only check for qualitative characteristics of the flow, so other metrics become necessary to determine model credibility.

There is a precedent for such a metric in planetary science: a study using identical forcings in different GCMs gave differing results for simulations of Venus’s superrotating atmosphere, indicating dependencies on the nature of the dynamical core used and in particular the parameterization of the numerical diffusion (Lee & Richardson 2010). Further investigations showed that atmospheric angular momentum conservation varied significantly between models (Lebonnois et al. 2012, hereafter Leb12), which explained why there were differences in circulation and the models with the best angular momentum conservation properties produced the most credible atmospheric states (Lee & Richardson 2012). Since tidally locked terrestrial exoplanets are expected to have atmospheres dominated by a superrotating jet (Hammond et al. 2020), arguments relating angular momentum conservation...
to realism of the model circulation for Venus can plausibly be extended to these exoplanets as well.

In this paper, we explore model errors in the simulation of Earth-like tidally locked planets with different rotation rates. The paper is organized as follows: Section 2 describes the model setup and angular momentum budget calculations. Section 2 after that outlines the trends in angular momentum variability and their origins. Section 3 summarizes the findings and suggests directions for future research.

2. Methods and Model Setup

In this work we use THOR, an open-source GCM, for our simulations (Mendonça et al. 2016; Deitrick et al. 2020). It solves the nonhydrostatic Euler equations on a sphere with a finite-volume solver (Tomita et al. 2001) using an icosahedral grid. The simulation setup is the tidally locked Earth benchmark case described in Section 4.1 of Deitrick et al. (2020), which was first proposed in Heng et al. (2011) and is based on the Held–Suarez test commonly used for Earth GCM studies (Held & Suarez 1994) with a tidally locked forcing function. The planet has a radius equal to the Earth, with a 1 bar atmosphere and a mean molecular weight of 0.0048. It is one of the main tuning parameters responsible for model stability.

The paper is organized as follows: Section 2 describes the model setup and angular momentum budget calculations. Section 3 after that outlines the trends in angular momentum variability and their origins. Section 3 summarizes the findings and suggests directions for future research.

The Astrophysical Journal, 923:39 (8pp), 2021 December 10

Kopparla et al.
calculated from Equation (6). Thus, we have

\[
\begin{align*}
\left( \frac{du}{dt} \right)_F &= \frac{u_{\text{mid}} - u_{\text{start}}}{\delta t} \\
\left( \frac{du}{dt} \right)_c &= \frac{u_{\text{end}} - u_{\text{mid}} - \delta u_D}{\delta t},
\end{align*}
\]

(7)

where \( \delta t \) is the time step, \( u_{\text{start}} \) is the zonal wind at the start of the time step, \( u_{\text{mid}} \) is after the physics module (friction) is calculated, and \( u_{\text{end}} \) is after dynamics calculations and is at the end of the time step.

In a perfect model, eventually the atmospheric angular momentum equilibrates and the net torque should become vanishingly small, but this rarely happens in practice due to finite simulation times and long-period atmospheric and/or numerical oscillations. The ratio between the absolute value of the physical \((F)\) and spurious numerical \((\epsilon+D)\) torques is calculated at each output time step over the runtime. The time median of this ratio over the entire simulation time gives the Global Credibility Score (GCS):

\[
\text{GCS} = \text{median}\left( \frac{\sum F_{\text{loc}}}{\sum (\epsilon_{\text{loc}} + D_{\text{loc}})} \right).
\]

(8)

Simulations where \( \epsilon + D \) is equal to or larger than \( F \) on average or, equivalently, GCS \( \leq 1 \) are totally not trustworthy, and GCS > 10 is preferable so that physical torques are at least an order of magnitude larger than numerical error. See Appendix A for more details on this score. Taking ratios at individual grid points can therefore blow up to arbitrarily large values, which is why we first sum over all grid points before calculating the ratio. However, this global average metric does not give the full picture. If, for instance, there are large surface frictional torques of very similar magnitude but eastward at the tropics and westward in the midlatitudes, the global surface torque sum would be very small even though these torques may significantly affect local flows. Thus, the above GCS metric might indicate that the model is unphysical, even if the local physical torques were quite large compared to numerical errors. We recognize that this is a shortcoming of this metric.

3. Results and Discussion

Figure 1 shows the results of the above-described angular momentum analysis with the GCS metric for four planets with different rotation rates. After a spin-up period of a few hundred days, the angular momentum reaches a steady state with some variability about a mean value. For slow rotators, angular momentum changes correspond to changes in physical torques like surface friction contributions, and dynamical error is negligible. But as rotation rate increases, conservation errors from the dynamical core become larger, and eventually for the fastest rotators, the equilibrium is set by a balance between the physics and the dynamical errors of nearly equal magnitude. So the model goes from being quite credible to very unreliable over this range.

In the absence of large physical torques such as mountain torque, numerical torques being prominent is somewhat expected (Lebonnois et al. 2012; Lauritzen et al. 2014). Since exoplanet simulations rarely have topography because we have never observed an exoplanet surface, this problem is likely to persist. The global averages seem to reduce numerical torques more than physical ones, leading to a larger range of values for the GCS.

Since there is no sponge layer in this model run, the physical sources and sinks for atmospheric angular momentum are at the surface. For the slow rotators, the frictional timescale acts to dissipate energy on short timescales (i.e., shorter than the rotational period). This energy is then lost, and it is artificially replenished by the forcing. For the fast rotators, when surface friction acts too slowly to adjust the atmosphere to the surface leading to low variability, the dynamical errors take over the torque balance. Since the frictional timescale is an important control on this process, we varied it between 10 and 100 days from the baseline value of 1 day to test the model error sensitivity. As the frictional timescale becomes slower, the credibility scores consistently worsen (see Figure 2).

The modeling convention is to keep the Rayleigh frictional timescale fixed at a standard value (very often a 1 day timescale inherited from the Held–Suarez test) regardless of the rotation rate (Penn & Vallis 2018; Wang et al. 2018) or vary it as a free parameter over arbitrarily large ranges (Pascale et al. 2013; Carone et al. 2016). But these results show that when the frictional timescale is equal to or larger than the rotational timescale, the model becomes unreliable.

We also experimented with changing grid resolution and numerical dissipation order within in the model. A finer grid allows for the resolution of smaller-scale flows, and a higher-order numerical diffusion is more scale selective in damping at the grid scale, both of which are desirable (Mendonça et al. 2016; Skinner & Cho 2021). For the high-resolution run the spatial resolution was doubled and the model time step was proportionally halved to achieve an approximately equal Courant number, thereby keeping the simulations amenable to intercomparison. The Courant number is calculated as

\[
C = \frac{U\Delta t}{d},
\]

(9)

where \( U \) is the magnitude of horizontal velocity at a given grid point, \( d \) is the grid spacing (see Equation (2) of Mendonça et al. 2016), and \( t \) is the model time step. Please see Appendix B for further Courant number distribution statistics.

The numerical diffusion timescale, defined as (Mendonça et al. 2016)

\[
\tau_d = \frac{\Delta t}{2^n+1}D_n,
\]

(10)

was held constant (with a value of 3900 s) when changing the numerical diffusion order \( n \) by adjusting the tuning parameter \( D_n \) so that the simulations can be intercompared. As seen in Figure 3 increasing the resolution decreases error, but increasing the order of the numerical diffusion scheme from 4 to 6 and further to 8 has mixed results. Furthermore, the eighth-order diffusion simulations take much longer to spin up, though why this happens is as of yet unknown.

The fact remains that all GCMs must be tuned to be run stably, and this tuning often takes the form of numerical diffusion, sponge layers, and/or spectral filters. It is a very subjective choice in terms of how it is implemented, what its magnitude is, and what the goal of the tuning is (e.g., to get a smooth kinetic energy spectrum or to have the smallest
Figure 1. Evolution of the total atmospheric angular momentum (left panel) and total torque (dM_r/dt) and its contributions from surface friction, numerical dissipation, and conservation errors in the dynamical core for planets with rotation rates from 0.1 to 100 days. The quantities plotted in the right column are global net torques. Simulations with higher scores are more credible.
possible values of numerical diffusion that will not crash the model; see Jablonowski & Williamson (2011) for a review of various tuning philosophies. But model tuning and model error are frequently underdiscussed in exoplanet climate literature, and often the tuning setup for a test case is used unchanged for simulations across a wide range of physical parameters without checking for changes in model errors.

4. Conclusions

In this work, we took a model tuned to work at a rotation period of 1 day and examined its behavior at over four orders of magnitude in rotation rate, with a focus on angular momentum changes. We find that angular momentum changes are closely related to surface friction torques for the slow-rotating planets, while model errors dominated the balance for the fast rotators. The sensitivity of model errors to numerical diffusion order, grid spacing, and frictional timescale were also examined. The main takeaway is that model credibility can vary widely over a given parameter space. In short, for a given GCM if different settings show different variabilities, then this framework argues that you must trust the setting that yields the most physical angular momentum balance. The big challenge is to find tunings that allow for reasonably physical GCM runs across wide parameter spaces, but it is fairly straightforward to tune the model if you are interested in a single planet. Therefore, exoclimate parameter studies are much harder to get right than individual exoplanet modeling. Overall, it might be good to do an ensemble of different GCM runs to overcome individual GCM quirks when trying to
Figure 3. Same as Figure 1, but with higher model grid resolution (second row) and higher numerical diffusion order (third and fourth rows) for a rotation rate of 1 day.
estimate exoclimate variability, as has become standard practice in the Earth climate modeling community. The THAI project is one such intercomparison effort focusing on terrestrial exoplanets (Fauchez et al. 2020).

It is well understood within the modeling community that changing the model tuning parameters changes the resulting atmospheric flow. But a less recognized corollary is that each tuning is only appropriate for a given set of physical model parameters. By holding the tuning constant and varying the physical parameters, there is a risk of over- or underdamping the numerical effects within the model, exactly as will happen by holding the physical parameters constant and changing the tuning.

Since we have only tested one model within a relatively narrow parameter space, this work is intended to be indicative and not exhaustive. Atmospheric angular momentum conservation is considered an important metric in hot Jupiter simulations (Rauscher & Menou 2010; Polichtchouk et al. 2014; Mayne et al. 2017), so our framework can be readily applied in those cases as well. The magnitude of the model error is sensitive to the type of dynamical core (with finite-volume cores having much higher errors than spectral ones; Lauritzen et al. 2014) and both physical and model tuning parameters, as shown in this work. Other model error metrics based on conservation laws could be devised. We hope that exoclimate modeling efforts begin to include model error metrics when describing results, particularly for parameter space exploration type studies.

P.K. acknowledges funding support from the CSH Fellowship of the University of Bern. R.D. and K.H. acknowledge funding support from the European Research Council via a Consolidator Grant awarded to K.H. (project EXOKLEIN; number 771620). K.H. acknowledges a honorary professorship from the University of Warwick.

Appendix A
Mean versus Median

Here we show why the choice of median was necessary in the calculation of the credibility scores. In the right panels of Figure A1, the distribution of the ratio of frictional torques to numerical torques has a long tail with a few extremely high values that occur during the model spin-up (indicated by the shaded region in the corresponding left panels). The mean is biased high because of these extreme values and not a good central measure of the distribution. Therefore, we use a median measure that is more resistant to such outliers.

Figure A1. The first two rows of Figure 3 are shown here with the histogram of the quantity $F/(D + \epsilon)$ calculated at each output time step, which is used to calculate the GCS through Equation (8). The right panels show the mean and median of the distribution of these ratios, which were used to calculate two possible GCS values.
Appendix B

Courant Number Statistics

Here we show further statistics for the Courant number with changing grid resolution and time step. As shown in Figure B1, when the resolution is increased without adjusting the time step (second panel), the Courant number spread is roughly twice as wide and is not a fair counterpart for intercomparison with the lower-resolution run. When the time step is halved (third panel) the Courant number range is similar to the lower-resolution run, but the extreme values are more dispersive as higher-resolution simulations allow for greater variability in velocities.

ORCID iDs

Pushkar Kopparla https://orcid.org/0000-0002-8951-3907
Russell Deitrick https://orcid.org/0000-0001-9423-8121
Kevin Heng https://orcid.org/0000-0003-1907-5910
João M. Mendonça https://orcid.org/0000-0002-6907-4476
Mark Hammond https://orcid.org/0000-0002-6893-522X

References

Amundsen, D. S., Baraffe, I., Tremblin, P., et al. 2014, A&A, 564, A59
Carone, L., Keppens, R., & Decin, L. 2016, MNRAS, 461, 1981
Deitrick, R., Mendonça, J. M., Sadowski, T., & Schrofenegger, U., et al. 2020, ApJS, 248, 30
Eyring, V., Bony, S., Meehl, G. A., et al. 2016, GMD, 9, 1937
Fauchez, T. J., Turbet, M., Wolf, E. T., et al. 2020, GMD, 13, 707
Hammond, M., Tsai, S.-M., & Pierrehumbert, R. T. 2020, ApJ, 901, 78
Held, I. M., & Suarez, M. J. 1994, BAMS, 75, 1825
Heng, K., & Showman, A. P. 2015, AREPS, 43, 509
Hooton, M. J., de Mooij, E. J. W., Watson, C. A., et al. 2019, MNRAS, 486, 2397
Jablonowski, C., & Williamson, D. L. 2011, in Numerical Techniques for Global Atmospheric Models, ed. P. Lauritzen et al. (Berlin: Springer), 381
Jackson, B., Adams, E., Sandidge, W., Kreyche, S., & Briggs, J. 2019, AJ, 157, 239
Kaspi, Y., & Showman, A. P. 2015, ApJ, 804, 60
Koutoulas, A., Grillakis, M., Tsanis, I., & Papadimitriou, L. 2016, CIDy, 47, 1881
Lauritzen, P. H., Bacmeister, J. T., Dubos, T., Lebonnois, S., & Taylor, M. A. 2014, JAMES, 6, 129
Lebonnois, S., Covey, C., Grossman, A., et al. 2012, JGRE, 117, E12004
Lee, C., & Richardson, M. 2010, JGRE, 115, E104002
Lee, C., & Richardson, M. 2012, Icar, 221, 1173
Liu, B., & Showman, A. P. 2013, ApJ, 770, 42
Mayne, N. J., Debras, F., Baraffe, I., et al. 2017, A&A, 604, A79
Mendonça, J. M., Grimm, S. L., Grosheintz, L., & Heng, K. 2016, ApJ, 829, 115
Mitchell, J. L., & Vallis, G. K. 2010, JGRE, 115, E12008
Pascale, S., Ragone, F., Lucarini, V., Wang, Y., & Boschi, R. 2013, P&SS, 84, 48
Penn, J., & Vallis, G. K. 2018, ApJ, 868, 147
Pierrehumbert, R. T., & Hammond, M. 2019, AnRFM, 51, 275
Polichtchouk, I., Cho, J.-K., Watkins, C., et al. 2014, Icar, 229, 355
Rauscher, E., & Menou, K. 2010, ApJ, 714, 1334
Read, P., Barstow, J., Charnay, B., et al. 2016, QJRMS, 142, 703
Skinner, J., & Cho, J. 2021, MNRAS, 504, 5172
Tamburo, P., Mandell, A., Deming, D., & Garhart, E. 2018, AJ, 155, 221
Tomita, H., Tsugawa, M., Satoh, M., & Goto, K. 2001, JCoPh, 174, 579
Wang, Y., Read, P. L., Tabataba-Vakili, F., & Young, R. M. 2018, QJRMS, 144, 2537
Zhang, X., & Showman, A. P. 2017, ApJ, 836, 73

Figure B1. Distribution of Courant numbers for three runs with a 1 day rotation period and variable spatial and time resolution. The maximum and minimum values are also plotted at each time step to make sure that there are no anomalous grid cells with very large values.