Geosci. Model Dev. Discuss., 8, 4375–4400, 2015
www.geosci-model-dev-discuss.net/8/4375/2015/
doi:10.5194/gmdd-8-4375-2015

Geoscientific
Model Development
Discussions
Open Access

Importance of bitwise identical reproducibility in earth system modeling and status report

L. Liu[1,2], S. Peng[1], C. Zhang[3], R. Li[3], B. Wang[1,2,4], C. Sun[1], Q. Liu[1], L. Dong[4], L. Li[4], Y. Shi[1], Y. He[1], W. Zhao[1], and G. Yang[1,2,3]

[1]Ministry of Education Key Laboratory for Earth system modeling, Center for Earth System Science (CESS), Tsinghua University, Beijing, China
[2]Joint Center for Global Change Studies (JCGCS), Beijing, China
[3]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[4]State Key Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Received: 30 April 2015 – Accepted: 25 May 2015 – Published: 12 June 2015

Correspondence to: L. Liu (liuli-cess@tsinghua.edu.cn), B. Wang (wab@tsinghua.edu.cn), and G. Yang (ygw@tsinghua.edu.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Reproducibility is a fundamental principle of scientific research. Bitwise identical reproducibility, i.e., bitwise computational results can be reproduced, guarantees the reproduction of exactly the same results. Here we show the importance of bitwise identical reproducibility to Earth system modeling but the importance has not yet been widely recognized. Modeled mean climate states, variability and trends at different scales may be significantly changed or even lead to opposing results due to a slight change in the original simulation setting during a reproduction. Out of the large body of Earth system modeling publications, few thoroughly describe the whole original simulation setting. As a result, the reproduction of a particular simulation experiment by fellow scientists heavily depends on the interaction with the original authors, which is often inconvenient or even impossible. We anticipate bitwise identical reproducibility to be promoted as a worldwide standard, to guarantee the independent reproduction of simulation results and to further improve model development and scientific research.

1 Importance of bitwise identical reproducibility

Earth system modeling, which simulates behavior and variation of the climate system, plays a critical role in understanding the past and predicting the future climate. An increasing number of numerical models have been developed for Earth system modeling, including stand-alone component models (e.g., atmospheric models, ocean models, land surface models, and sea ice models) and coupled models consisting of multiple component models, such as climate system models, Earth system models, etc. A large number of papers have been published with simulation results using these models.

Bitwise identical reproducibility, which guarantees the reproduction of exactly the same results, has already been used within some modeling groups to improve model development (Easterbrook and Johns, 2009; Ford et al., 2012). However, it is rarely used worldwide in sharing model codes and results. One possible reason is that it was extremely difficult to achieve bitwise identical reproducibility. As bitwise results are determined by the whole simulation setting (including the model code, input data, parameter setting, computing environment, etc. (Ford et al., 2012)) and are very sensitive to round-off errors determined by the finite precision of floating-point computations on modern computer systems (Monniaux, 2008), it requires scientists to preserve the whole simulation setting and recreate exactly the same simulation setting during a reproduction. It is highly unlikely that a simulation setting can be recreated exactly after a number of years, because some parts of the original simulation setting are no longer preserved or the original computing environment (including parallel setting, compiler version, compiling option, processor version, etc. (Ford et al., 2012)) is no longer available with the rapid upgrade of computer software and hardware. Moreover, as the whole simulation setting includes a lot of seemingly uninteresting information and is generally of a large size, it is not feasible to be detailed in a published paper or included as a Supplement. A recent study (Liu et al., 2015) shows that the information for recreating the same simulation setting can be easily recorded into a package of small size (called simulation setting package) with the help from an upgraded model software platform (a runtime software environment for configuring, building and running models), and bitwise identical results can be reproduced with upgraded computer software and hardware.

Another possible reason is that some scientists may feel it is unnecessary to reproduce the results of climate simulations at a bitwise identical level, because these results are generally statistical characteristics of output data from Earth system modeling on time scales longer than a few months. However, it has been shown that climate simulation results can be sensitive to round-off errors (Song et al., 2012; Hong et al., 2013). It is widely known that changes of the computing environment can introduce new round-off errors. Slight changes of the model code, input data or parameter setting can also introduce new round-off errors, because the floating point computations in the simulation as well as their inputs will be changed. Round-off errors therefore can be viewed

as the tiniest error in climate simulations. The sensitivity of climate simulation results to round-off errors indicates that slight changes of the model code, input data, parameter setting or computing environment may lead to failed reproduction. In such cases, it is required that the reproduction be conducted at the bitwise identical level.

To further illustrate the importance of bitwise identical reproducibility to Earth system modeling, we re-ran the historical experiment of Coupled Model Intercomparison Project Phase 5 (CMIP5) for the beginning 60 years (from 1 January 1850 to 31 December 1909) using two CMIP5 models: CESM1 (Gent et al., 2011) and FGOALS-g2 (Li et al., 2013). For each model, we designed ten simulations with slight differences in the computing environment (Table 1), while keeping the rest of the simulation setting unchanged. Any simulation of a model can be considered as "correct", while the other simulations can be viewed as an attempt of reproduction. Thus, different results among the simulations of each model can be used to evaluate the importance of the bitwise identical reproducibility.

Figure 1 uses standard deviation to quantify differences between the climatological mean surface air temperatures (SAT) by the ten simulations of each model. Although the globally averaged standard deviations (area weighted) are small (less than 0.15 °C), significant standard deviations (greater than 1 °C) exist in the high latitudes. Moreover, the standard deviations of the seasonal mean (using June–July–August and December–January–February as examples) are much greater than the annual mean. The domains with significant differences of the climatological mean SAT also show significant differences or even contraries in their decadal variations of the 10-year-mean SAT (Fig. 2). As a result, significant differences or even contraries are observed in the linear trend of time series of spatially averaged SAT (Fig. 3). When reducing the domain from the global to the Northern Hemisphere and then to a high-latitude region (60–90° N), the differences or contraries become more serious. Although low-latitude regions only show slight differences in the climatological mean and decadal variation of SAT, obvious differences in the interannual variability also exist, such as the El Niño–Southern Oscillation (ENSO). There are significant differences in terms of phase, amplitude, power, and periods of Niño-3 index (Fig. 4). Similarly, significant differences are observed in Niño-3.4 index. Like differences in SAT, significant differences in wind and precipitation also exist due to a slight change in the simulation setting. For example, significant differences are present in the correlation between the monsoon index and precipitation in the Asia Monsoon region (Fig. 5).

The above results show that modeled mean climate states, variability and trends at different scales may be significantly changed or even lead to opposing results due to the new round-off errors resulting from a slight change of the original computing environment during a reproduction. Similar results can be observed when changing the other parts of the original simulation setting (i.e., the model code, input data and parameter setting) during a reproduction, because new round-off errors are also introduced due to the changes.

1.1 Current status of bitwise identical reproducibility

The previous section reiterates the importance of bitwise identical reproducibility to Earth system modeling. So, what is the current status of bitwise identical reproducibility of published results?

In this study, we conducted a survey in two major steps. The first step is selecting papers. Only recent papers published between 2006 and 2014 were considered. In order to highlight high-impact papers, a simple criterion was designed using the number of citations (Table 2). To make the selected papers distribute evenly among journals as well as publication years, for each year, at most three papers with new simulation results of Earth system modeling were picked from each journal. As a result, in each year, a number of papers were selected and the average citation number is much higher than the corresponding threshold in the criterion (Table 2). Finally, 351 high-impact papers from 17 journals were selected (Supplementary Table S1).

The second step is bitwise identical reproduction of the simulation results from the selected papers. Since none of the papers includes the information of the whole simulation setting, we started to email all corresponding authors of each paper in July 2014,

in order to interactively reproduce the published results. After the authors of a paper provided us all required information, we tried to recreate exactly the same simulation setting. It was a challenge for us to prepare various computing environments for the bitwise identical reproduction. When lacking the same computing environment, we tried to re-run the simulation in our available computing environments. At the end of this step, a survey result was concluded for each paper (Supplementary Table S2).

Finally, we did not have responses for 283 papers (80.6 %), due to no corresponding author (five papers, 1.4 %), automatic email rejection (66 papers, 18.8 %) or no active reply (212 papers, 60.4 %). For the remaining papers, we did not obtain the required information on the simulation settings for 54 papers (15.4 %), among which the authors of 47 papers (13.4 %) confirmed their inconvenience for the bitwise identical reproduction. For the rest 14 papers (4.0 %), most of which were published after 2010, we received the required information from the authors and then tried to reproduce the bitwise identical results. Because we did not have the same computing environments, the simulations in four papers (1.1 %) were successfully re-run but without producing the bitwise identical results, and the simulations in another five papers (1.4 %) were not successfully re-run. Only the simulation results in five papers (1.4 %) were bitwise identically reproduced at the end.

The survey results demonstrate that the importance of bitwise identical reproducibility to Earth system modeling has not yet been widely recognized. Fellow scientists can easily download a paper with research findings independently of the authors, but it heavily depends on the authors' help to reproduce the simulation results. As the whole simulation setting is rarely kept for a long time (say more than 10 years), it is always inconvenient even impossible to recreate the same simulation setting. Even when the whole simulation setting can be recalled, the authors still have to spend a lot of efforts to help the fellow scientists who want to reproduce these results, while the bitwise identical reproduction may fail at the end due to the lack of an appropriate computing environment. Although ensemble with enough members of simulations can make some simulation results insensitive to changes of the computing environment (Song et al., 2012), only 71 selected papers (20.2 %) used ensemble approach and the numbers of ensemble members are generally small, for example no more than 20 for most of these papers (Supplementary Table S1). Moreover, it needs to be investigated that whether or not ensemble or other approaches can make various simulation results of a model insensitive to changes of computing environments.

2 Discussion

2.1 Uncertainty due to round-off errors

More and more evidences, including this study, have shown that round-off errors can introduce significant uncertainty to climate simulation results. Some authors involved in the survey of this study stated that they had realized similar phenomenon for a number of years and believed that their published results would not be sensitive to round-off errors and the reproduction was unnecessarily at the bitwise identical level. We also intuitively believed that mean climate states would not be sensitive to round-off errors before this study, but Fig. 1 indicates a very different conclusion. Scientists may rarely examine the sensitivity of simulation results to round-off errors in the past. Moreover, given the same kind of climate simulation results, different models may have different scales of sensitivity. As round-off errors are random, unpredictable and unavoidable, the impact of the uncertainty due to round-off errors to simulation results is hard to be controlled and understood. We therefore propose scientists to quantify the sensitivity of various kinds of simulation results of various models to round-off errors in the future.

2.2 Worldwide standard of bitwise identical reproducibility

Although the bitwise identical reproducibility of Earth system modeling is currently at a very low level, we propose to promote it as a worldwide standard: original scientists of published results should ensure the whole simulation setting publicly available for bitwise identical reproduction, so that any fellow scientists can independently obtain

the whole simulation setting and then independently repeat the original simulation or reproduce the original results.

Such a standard will guarantee that the published simulation results can be reproduced exactly and independently, so as to improve the trust of published results. It will not introduce any new burden to fellow scientists, because it does not enforce every reproduction by fellow scientists at the bitwise identical level. However, it will enable fellow scientists to easily and independently obtain all detailed information of the original simulation setting for further researches. It therefore will promote sharing and spreading model code, data, results, knowledge and experiences in a worldwide region.

The worldwide bitwise identical reproducibility can also lead to a rapid improvement in code quality (Easterbrook, 2014) with more and more test cases. In the field of computer science, there is a valuable concept of “record and replay” for bug tracking. A program should be tested with an increasing number of test cases, while each test case should be sufficiently recorded and then can be replayed (exactly reproduced) when required (for example, if a bug is detected). The worldwide bitwise identical reproducibility can help bug tracking for model development. A model simulation by anyone can be viewed as a test case for the model codes. If a model simulation detects a bug but cannot be reproduced by the modeling group who is responsible for the model development, an important chance for improving the model codes is wasted. Bitwise identical reproducibility can guarantee the exact reproduction of the bug.

Figure 6 shows our proposed framework for achieving worldwide bitwise identical reproducibility. It requires scientists and journals to cooperatively take actions and also requires some technical supports.

2.3 Scientists’ actions

Scientists of the Earth system modeling community should pay attention to bitwise identical reproducibility when developing models or conducting simulations. The model code, input data, computing environments and simulations should be managed by the model software platforms that have been upgraded with the enhancement of bitwise identical reproducibility (Liu et al., 2015), so that information of the whole simulation setting of any simulation can be recorded into a small package automatically. Moreover, the model code and input data of a simulation should be preserved and open for the future reproduction by anyone (Ince et al., 2012; Easterbrook, 2014). Although this requirement will introduce a new burden to scientists, we believe that further advances in model software platforms will minimize this burden greatly.

2.4 Journals' actions

To enhance the reproducibility of published research results, journals such as the *Nature* family, *Science* and *Geoscientific Model Development* now encourage authors to publicly share their code and input data and ask them to state the availability of the code and input data in their papers (Hanson et al., 2011; GMD Executive Editors, 2013; Nature, 2014a). However, this study shows that the availability of the code and input data alone is not enough for reproducing the results in the field of earth system modeling. We therefore expect journals to unite (Nature, 2014b) to play a critical role in promoting bitwise identical reproducibility to become a worldwide standard. They can encourage authors to provide the information package of the whole simulation setting as a supplementary material in their submission, to enable their simulation results independently reproducible. For the simulation results whose reproduction does not depend on bitwise identical reproducibility, authors should be asked to clearly state such independence in the paper. Moreover, journals should allow fellow scientists to leave feedbacks online on the reproducibility of each paper.

The reproducibility corresponding to a submitted manuscript should be tested by journals before their publication. Some journals have already made such kind of effort. For example, *Geoscientific Model Development* encourages referees to compile the code and run test cases supplied by the authors (GMD Executive Editors, 2013). However, such a way of testing will introduce a new burden to referees and will be inconvenient to check the reproducibility when the simulation requires a large amount of computing resource or a long time to be finished. The testing for bitwise identical

reproducibility will be more practical because it can be conducted automatically with a short run of the simulation (say for several model days) (Easterbrook and Johns, 2009).

2.5 Technical supports

Model software platforms should be continuously upgraded for the worldwide bitwise identical reproducibility of simulation results from various models. As it cannot be guaranteed that scientists are able to individually preserve the whole simulation setting of published results for a long time (say for more than ten years), we call for third-party open repositories for archiving and sharing the whole simulation setting and testing platforms with various computing environments for automatically checking the bitwise identical reproducibility of published results. Before publishing a paper, journals can ask authors to upload the whole simulation setting to third-party open repositories and to testing platforms for automatically checking the bitwise identical reproducibility. Third-party open repositories and testing platforms can be constructed in different countries or different cities and work cooperatively for worldwide bitwise identical reproducibility, so that scientists at different places of the world can conveniently upload and download the whole simulation settings of published results. Model software platforms can serve the whole process of uploading or downloading with a simple user command. As the open repositories will include the whole simulation settings of more and more simulations, they can enable fellow scientists to search a number of interesting simulations according to a set of detailed information. The ongoing development of metadata (data describing data) for earth system modeling (Lawrence et al., 2012; Guilyardi et al., 2013; Moine et al., 2014) will provide substantial supports to such kind of search.

2.6 Model intercomparison projects' actions

Similar to journals, model intercomparison projects, which mainly aim to improve and to develop Earth system models and their components, as well as to share the outputs, should also unite to play a critical role in promoting the bitwise identical reproducibility to be a worldwide standard; for example, encourage modeling groups to provide the information package of the whole simulation setting when they submit the outputs. Model intercomparison projects can also take consideration of the framework in Fig. 6 for achieving worldwide bitwise identical reproducibility.

3 Conclusions

This work focuses on the reproducibility of simulation results of Earth system modeling. As the results from an individual model simulation are potentially sensitive to a slight change of the original simulation setting during a reproduction, bitwise identical reproducibility that guarantees exact reproduction therefore is important to Earth system modeling. The survey with hundreds of published papers reveals that the importance of bitwise identical reproducibility has not been satisfactorily recognized. Considering reproducibility is a fundamental principle of scientific research, we propose to promote bitwise identical reproducibility as a worldwide standard. We believe the worldwide bitwise identical reproducibility is practical and will improve the model development and scientific research of Earth system modeling, e.g., improve the trust of published results, promote sharing and spreading model code, data, results, knowledge and experiences in a worldwide region, and rapidly improve the code quality.

Code availability

The historical simulations of CESM1 were created according to the corresponding historical experiment named “b40.20th.track1.2deg.001” (http://www.cesm.ucar.edu/experiments/cesm1.0/): code version CESM1.0.5 was used, the component configuration (“CCSM_COMPSET”) was set to “B_1850-2000_CN”, the resolution (“GRID”) was set to “1.9x2.5_gx1v6”, and the machine name (“MACH”) was set to “generic_linux_intel”. All historical simulations were restarted from 01-01-0501 of a Pre-Industrial Control experiment (“b40.1850.track1.2deg.003”). The code version can

be downloaded from website http://www.cesm.ucar.edu/models/cesm1.0/. When building a simulation, the input data could be downloaded automatically.

The historical simulations of FGOALS-g2 were based on its CMIP5 historical experiment. All simulations were restarted from 01-01-0440 of a Pre-Industrial Control experiment. Please contact us for more detailed information of the whole simulation setting, such as the model code, parameter setting and input data of FGOALS-g2.

The Supplement related to this article is available online at doi:10.5194/gmdd-8-4375-2015-supplement.

Acknowledgements. This work is supported in part by the Natural Science Foundation of China (no. 41275098), the National Grand Fundamental Research 973 Program of China (no. 2014CB441302) and the Tsinghua University Initiative Scientific Research Program (no. 20131089356).

References

Easterbrook, S. M. and Johns, T: Engineering the Software for Understanding Climate Change, IEEE Comput. Sci. Eng., 11, 65–74, 2009.

Easterbrook, S. M.: Open code for open science?, Nat. Geosci., 7, 779–781, 2014.

Ford, R., Riley, G., Budich, R., and Redler, R. (Eds.): Earth System Modelling – Volume 5 Tools for Configuring, Building and Running Models, Series: Springer Briefs in Earth System Sciences, 97 pp., 2012.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, J. Climate, 24, 4973–4991, 2011.

GMD Executive Editors: Editorial: The publication of geoscientific model developments v1.0, Geosci. Model Dev., 6, 1233–1242, doi:10.5194/gmd-6-1233-2013, 2013.

Guilyardi, E., Balaji, V., Lawrence, B., Callaghan, S., Deluca, C., Denvil, S., Lautenschlager, M., Morgan, M., Murphy, S., and Taylor, K. E.: Documenting Climate Models and Their Simulations, Bull. Am. Meteor. Soc., 94, 623–627, 2013.

Hanson, B., Sugden, A., and Alberts, B.: Making data maximally available, Science, 331, p. 649, 2011.

Hong, S.-Y., Koo, M.-S., Jang, J., Kim, J.-E. E., Park, H., Joh, M.-S., Kang, J.-H., and Oh, T.-J: An Evaluation of the Software System Dependency of a Global Atmospheric model, Mon. Weather Rev., 141, 4165–4172, 2013.

Ince, D. C., Hatton, L., and Graham-Cumming, J.: The case for open computer programs, Nature 482, 485–488, 2012.

Lawrence, B. N., Balaji, V., Bentley, P., Callaghan, S., DeLuca, C., Denvil, S., Devine, G., Elkington, M., Ford, R. W., Guilyardi, E., Lautenschlager, M., Morgan, M., Moine, M.-P., Murphy, S., Pascoe, C., Ramthun, H., Slavin, P., Steenman-Clark, L., Toussaint, F., Treshansky, A., and Valcke, S.: Describing Earth system simulations with the Metafor CIM, Geosci. Model Dev., 5, 1493–1500, doi:10.5194/gmd-5-1493-2012, 2012.

Li, L., Lin, P., Yu, Y., Wang, B., Zhou, T., Liu, L., Liu, J., Bao, Q., Xu, S., Huang, W., Xia, K., Pu, Y., Dong, L., Shen, S., Liu, Y., Hu, N., Liu, M., Sun, W., Shi, X., Zheng, W., Wu, B., Song, M., Liu, H., Zhang, X., Wu, G., Xue, W., Huang, X., Yang, G., Song, Z., and Qiao, F.: The Flexible Global Ocean-Atmosphere-Land System Model: Grid-point Version 2: FGOALS-g2, Adv. Atmos. Sci., 30, 543–560, 2013.

Liu, L., Li, R., Zhang, C., Yang, G., Wang, B., and Dong, L.: Enhancement for bitwise identical reproducibility of Earth system modeling on the C-Coupler platform, Geosci. Model Dev. Discuss., 8, 2403–2435, doi:10.5194/gmdd-8-2403-2015, 2015.

Moine, M.-P., Valcke, S., Lawrence, B. N., Pascoe, C., Ford, R. W., Alias, A., Balaji, V., Bentley, P., Devine, G., Callaghan, S. A., and Guilyardi, E.: Development and exploitation of a controlled vocabulary in support of climate modelling, Geosci. Model Dev., 7, 479–493, doi:10.5194/gmd-7-479-2014, 2014.

Monniaux, D: The pitfalls of verifying floating-point computations, ACM Trans. Program. Lang. Syst., 30, 1–41, 2008.

Nature: Code share, Nature, 514, p. 536, 2014a.

Nature: Journals unite for reproducibility, Nature, 515, p. 7, 2014b.

Song, Z., Qiao, F., Lei, X., and Wang, C.: Influence of parallel computational uncertainty on simulations of the Coupled General Climate Model, Geosci. Model Dev., 5, 313–319, doi:10.5194/gmd-5-313-2012, 2012.

Webster, P. J. and Yang, S.: Monsoon and ENSO: Selectively interactive systems, Q. J. Roy. Meteor. Soc., 118, 877–926, 1992.

Table 1. Simulations of CMIP5 historical experiment of CESM1 and FGOALS-g2. The corresponding simulation settings are only slightly different in terms of computing environments, including compiler versions and compiling options as well as parallel settings. Table 1a lists the name of each simulation, Table 1b provides information of each parallel setting and Table 1c shows detailed compiling options. All simulations were run on a homogeneous supercomputer consisting of a number of Intel Xeon X5670 CPU. Intel compiler with different versions was used to compile the model code. **(a)** Name of each simulation. The name is formatted as *VVV_PPP_CCC*, where "*VVV*" is the version of the Intel compiler, "*PPP*" labels the parallel setting and "*CCC*" labels compiling options. **(b)** Process numbers of component models in each parallel setting. "ATM" means the atmospheric model, "OCN" means the ocean model, "LND" means the land surface model, "ICE" means the sea ice model, "CPL" means the coupler, and "GLC" means the glacier model. FGOALS-g2 does not include a glacier model as its component. **(c)** Detailed information of compiling options.

(a)

Model	Simulation				
CESM1	11.1_120_C1	11.1_128_C1	11.1_128_C2	12.1.3_128_C1	11.1_96_C1
	11.1_96_C2	11.1_104_C1	11.1_104_C2	11.1_112_C1	11.1_112_C2
FGOALS-g2	11.1_104_C1	11.1_108_C1	11.1_108_C2	12.1.3_108_C1	11.1_106_C1
	11.1_106_C2	11.1_110_C1	11.1_110_C2	11.1_112_C1	11.1_112_C2

(b)

Model	Label	Number of processes					
		ATM	OCN	LND	ICE	CPL	GLC
CESM1	96	96	96	96	96	96	96
	104	104	96	104	96	104	104
	112	112	96	112	96	112	112
	120	120	120	120	120	120	120
	128	128	128	128	128	128	128
FGOALS-g2	104	30	20	18	24	12	–
	106	30	20	18	24	14	
	108	30	20	18	24	16	–
	110	30	20	18	24	18	–
	112	30	20	18	24	20	

(c)

Model	Label	Compiling option
CESM1	C1	-O2 -convert big_endian -assume byterecl -ftz -FR -fp-model precise
	C2	-O2 -convert big_endian -assume byterecl -ftz –FR
FGOALS-g2	C1	-c -r8 -i4 -O2 -zero -132 -convert big_endian -assume byterecl -no-vec -mp1 -fp-model precise -fp-speculation=safe
	C2	-c -r8 -i4 -O2 -zero -132 -convert big_endian -assume byterecl

Table 2. Statistical characteristics of the paper selection. We selected all the papers from 24 February 2014 to 27 April 2014. Citation numbers are obtained from Web of Science (https://apps.webofknowledge.com/).

Year of publishing	2006	2007	2008	2009	2010	2011	2012	2013	2014
Threshold of citation number	≥ 10	≥ 9	≥ 8	≥ 7	≥ 5	≥ 4	≥ 3	≥ 1	≥ 0
Number of selected papers	35	35	42	41	42	45	48	46	17
Average citation number per selected paper	92.1	74.5	65.9	52.6	26.0	31.8	20.4	5.1	0.4

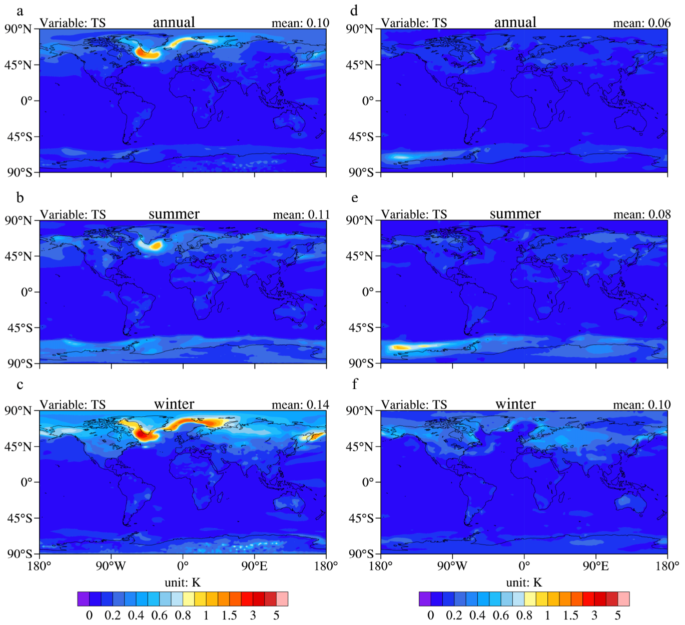

Figure 1. Standard deviations of climatological mean surface air temperature (SAT) from ten simulations by two models. **(a–c)**: Corresponding to annual mean, June–July–August (JJA) and December–January–February (DJF) SAT of CESM1; **(d–f)**: the same as **(a–c)**, except for FGOALS-g2. The ten simulations of each model are conducted following the CMIP5 historical experiment from 1 January 1850 to 31 December 1909, under different computing environments, e.g., parallel settings, compiler versions and compiling options (Table 1).

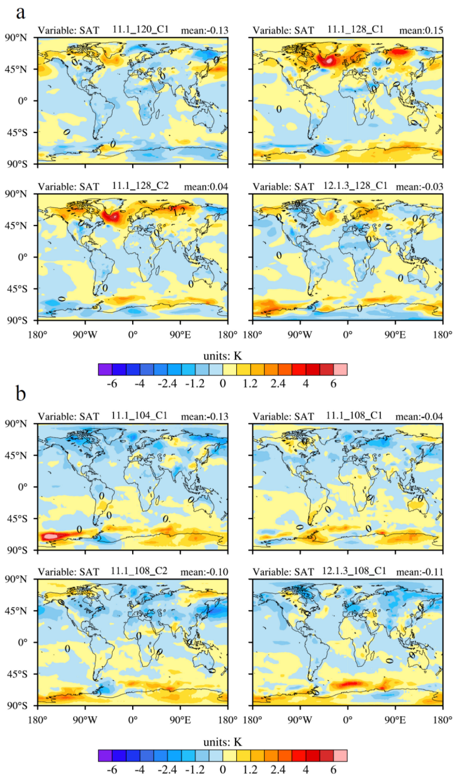

Figure 2. Decadal variation of mean surface air temperature (SAT) in 1900–1909 with respect to 1850–1859. **(a)** Corresponding to the first four simulations of CESM1; **(b)** corresponding to the first four simulations of FGOALS-g2. Significant differences are observed at high latitudes. For example, average (area weighted) decadal variations of the four simulations are −0.01, 1.41, 1.07, and 0.66 in the domain 60–90° N, and are −0.16, 0.17, −0.07, and 0.22 in the domain 60–90° S. Average decadal variations of the four simulations are −0.87, −0.27, −0.43, and −0.78 in the domain 60–90° N, and are 0.44, 0.28, 0.24, and 0.26 at the domain 60–90° S.

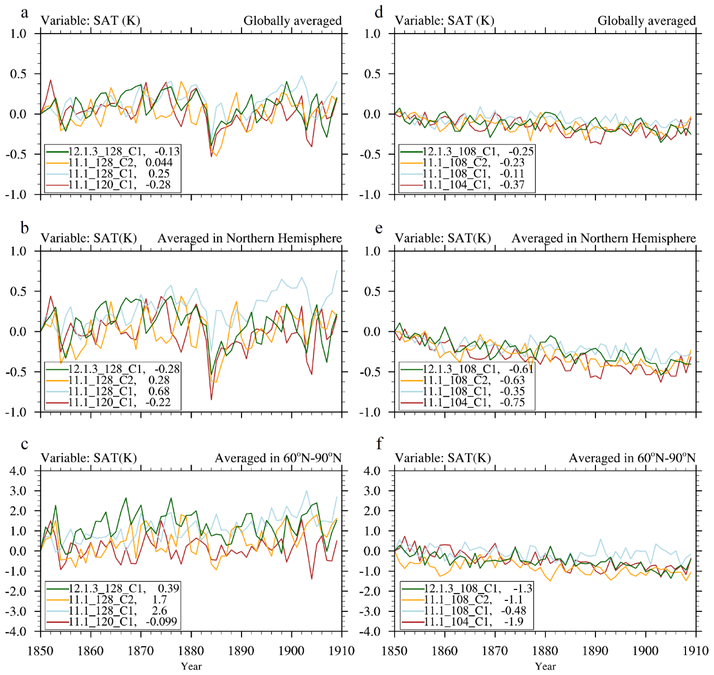

Figure 3. Time series of area-averaged surface air temperature (SAT). **(a–c)**: From the first four simulations of CESM1 at the global scale, Northen Hemisphere and a high-latitude region (50–90° N), respectively; **(d–f)**: from the first four simulations of FGOALS-g2. In each panel, the linear trend (K per 100 years) of the time series of each simulation is listed following the simulation name.

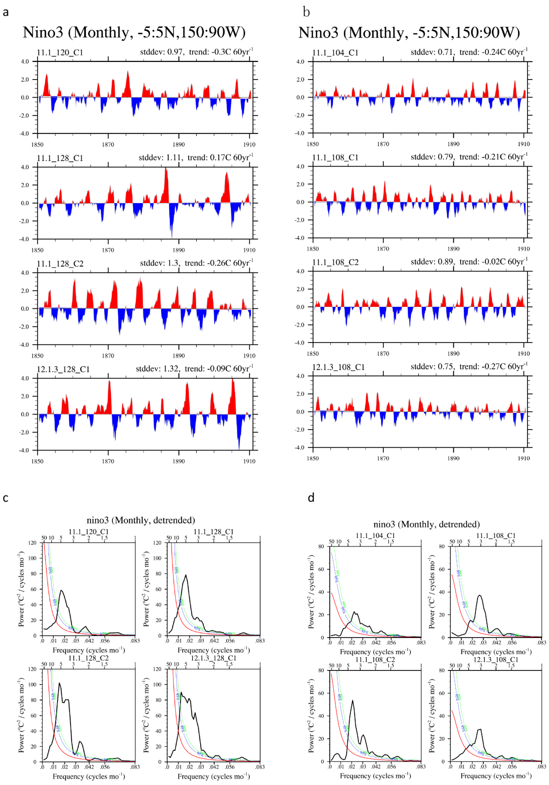

Figure 4. ENSO characteristics. **(a–b)** Time series of Niño-3 index (5° S–5° N, 150–90° W) from the first four simulations of CESM1 and FGOALS-g2, respectively. **(c–d)**: Power spectrum of Niño-3 index, corresponding to **(a–b)**.

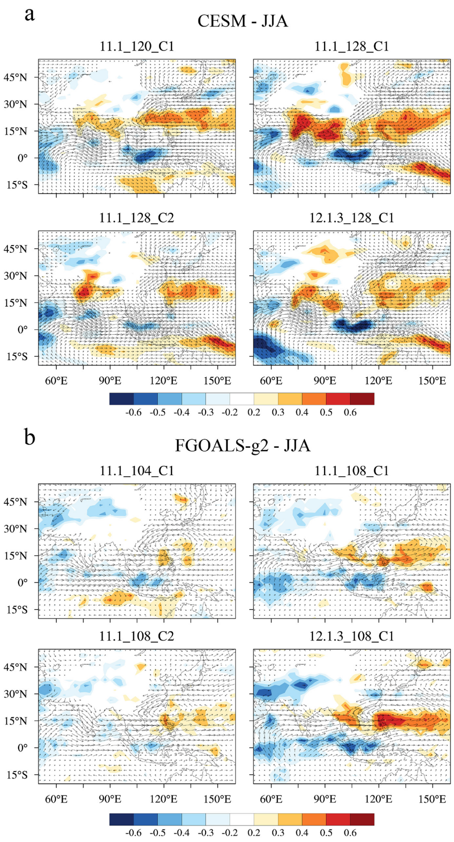

Figure 5. Correlation between monsoon index and total precipitation of June–July–August (JJA) in the Asia Monsoon region. **(a)** Corresponding to CESM1; **(b)** corresponding to FGOALS-g2. The monsoon index used here is the Webster–Yang index (Webster and Yang, 1992).

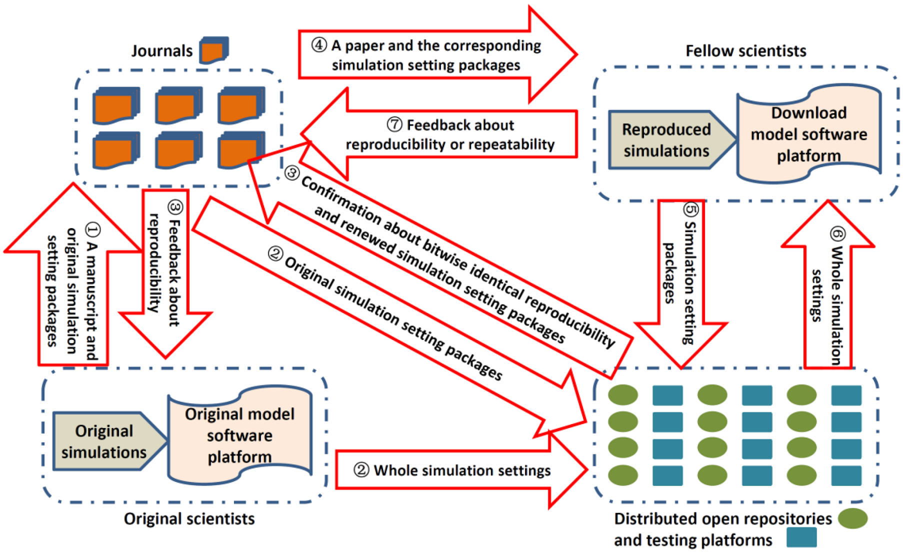
Journals
④ A paper and the corresponding simulation setting packages
Fellow scientists
⑦ Feedback about reproducibility or repeatability
Reproduced simulations
Download model software platform
③ Confirmation about bitwise identical reproducibility and renewed simulation setting packages
① A manuscript and original simulation setting packages
③ Feedback about reproducibility
② Original simulation setting packages
⑤ Simulation setting packages
⑥ Whole simulation settings
Original simulations
Original model software platform
② Whole simulation settings
Original scientists
Distributed open repositories
and testing platforms

Figure 6. A framework for achieving worldwide bitwise identical reproducibility. When original scientists submit a manuscript to a journal, they will be asked to submit the corresponding original simulation setting packages that were automatically produced by the original model software platform. The journal will automatically send these packages to the union of open repositories and testing platforms, to make the whole simulation settings corresponding to the manuscript be automatically uploaded from the original scientists. Next the journal will obtain the testing results about bitwise identical reproducibility and obtain renewed simulation setting packages (if the simulation results can be bitwise identically reproduced) that will be supplementary materials of the manuscript, and then notifies original scientists the feedback about bitwise-identical reproducibility. If the simulation results cannot be bitwise identically reproduced, original scientists can call for help from the modeling groups who are responsible for the development of the corresponding models. Thus modeling groups can get more test cases for the improvement of models. After fixing the problems in the simulations, original scientists can resubmit the revised simulation setting packages to the journal (simulation results referred in the manuscript may be changed). Fellow scientists can obtain the corresponding simulation setting packages when downloading a paper. Using the simulation setting packages, fellow scientists can independently download the corresponding whole simulation settings (including the original model software platform) from open repositories and then independently repeat the original simulations or independently reproduce bitwise identical results for conducting new simulations. Journals will welcome fellow scientists to post feedbacks about the bitwise identical reproducibility.