American Journal of Theoretical and Applied Statistics
2022; 11(3): 83-88
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20221103.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Tuning a Model in Climatology and Calibrating One in Hydrogeology: An Informative Comparison

R. A. Weitzman

Naval Postgraduate School, Monterey, USA

Email address:
ronweitzman@redshift.com

To cite this article:
R. A. Weitzman. Tuning a Model in Climatology and Calibrating One in Hydrogeology: An Informative Comparison. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 3, 2022, pp. 83-88. doi: 10.11648/j.ajtas.20221103.11

Received: April 17, 2022; **Accepted**: May 5, 2022; **Published**: May 12, 2022

Abstract: A November 2021 article in the journal *Chance* on a misuse of statistics by hydrogeologists in their modeling of water levels below ground raises the question of whether climatologists might be committing the same statistical errors in their modeling of global warming above ground. In seeking to answer that question, the research reported in this article finds the answer to be, yes, both research communities corrupt data by altering values of independent variables to reduce error variation or to achieve particular model results. That data alteration not only creates an impermissible negative correlation between estimates and errors but also creates model estimates that exaggerate trends in the observations. The exaggerated trends occur regardless of the nature or the intent of the data alteration. For that reason, use of trends in model estimates resulting from data alteration as a guide to future research or as a basis for conclusions may lead researchers astray. This article suggests an alternative research strategy consisting of random sampling of observation zones which, by limiting a study to thousands rather than millions of zones, could enable researchers to obtain sufficiently accurate input data to make the alteration of data unnecessary. Use of this procedure could also help avoid exaggerated and misleading predictions from models.

Keywords: Tuning, Calibration, Linear Model, Estimation, Prediction, Error, Global Warming

1. Introduction

Does tuning a model in climatology have the same meaning as calibrating a model in hydrogeology? That is an interesting question because in a November 2021 *Chance* article that took a forensic look at the misuse of statistics in hydrogeology the villain turned out to be model calibration [14]. To find the answer, a good place for someone who is not a climatologist to begin is the recent book *Unsettled: What Climate Science Tells Us, What It Doesn't, and Why It Matters* by physicist Steven E. Koonin [7], with particular attention to Chapter 4 on modeling. The short answer there is, yes, the two have the same meaning. This article is about the rest of the story.

Except for books like Koonin's (e.g., [8, 13]), much of the literature cited here dates prior to 2020. The reason is that, unlike weather (which changes from day to day), climate varies over decades, and the world's climatology community has organized its research on climate change accordingly, beginning with the First Assessment Report of the Intergovernmental Panel on Climate Change (AR1) in 1990. The most recent complete report, AR5, was issued in 2014, with most of the literature cited here centering on that date. To say that the study of global temperature is a hot topic would be a gross understatement. Like Koonin's, a number of the books published recently on climate change are critiques of AR5 and its predecessors. Differing from these books, in which tuning is considered (if at all) as only one of a number of concerns, this article focuses on tuning, identifies specific and consequential statistical problems with the practice, and suggests a statistical alternative that could avoid those problems. The presentation will being with a description of the modeling used in both fields,

In each zone of a layered checkerboard of zones, which in climatology covers the whole earth upward in the atmosphere and downward in the oceans, the models involved break down an observed measurement—of water level in hydrogeology and temperature in climatology—into estimate and error components. In that breakdown, the estimate is a weighted sum of values of independent variables, like well-pumping rate in hydrogeology and number of parts per million of airborne carbon dioxide (CO_2) molecules in climatology. Like the observed measurements, values of the independent variables can vary over time and from zone to zone while the weights remain constant over time and all zones. In model development, based on the observed measurement and the independent-variable values in every zone, the weights are determined to minimize the error variation while keeping the average error and the correlation between estimates and errors equal to zero.

Often, to reduce the error variation even further, hydrogeologists calibrate and climatologists tune their models by adjusting unreliably-determined values of independent variables.

2. Calibration in Hydrogeology

As shown in the *Chance* article [14], this process can create a negative correlation between estimates and errors because movement of the error component of a measurement toward zero moves its estimate component equally in the opposite direction to avoid changing the observed measurement, which is the sum of its error and estimate components. As Figure 1 shows, that is in fact what happened in a project involving the modeling of water levels by hydrogeologists. The question now is whether the same thing has occurred in the modeling of temperatures by climatologists.

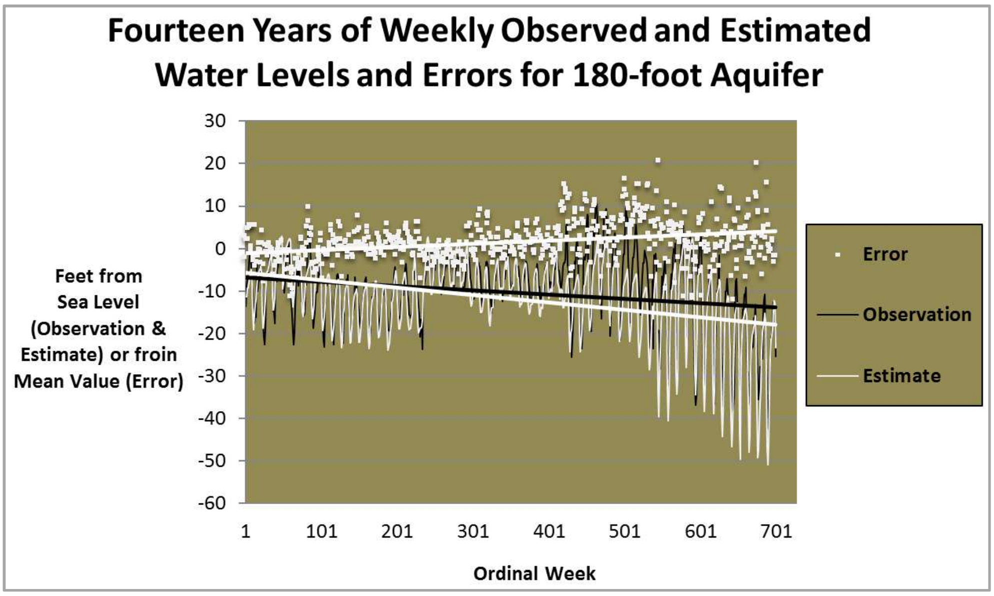

Figure 1. *Calibrated model estimates (white line) of observed water levels (black line) with errors (filled circle) and corresponding trendlines over time.*

Exploration of the answer to that question can benefit from a corresponding exploration of Figure 1, which shows that errors go up (from negative to positive) as water levels go down over time. Because errors, by definition, should not be predictable, the non-zero correlation between errors and water levels was so troubling that the project abandoned the use of the model to estimate water levels. The hydrogeologist making that decision based it on the belief that the observed change in water level over time was too fast for the model to catch up. If that were true, the model would be overestimating water levels when, as shown in Figure 1, it is underestimating them: The bottom trendline (estimates) is lower than the middle trend line (observations). So, what is the real problem?

Though based on virtually the same data, Figure 1 is not the figure shown in the report of the project. The figure in the report shows the errors trending downward rather than upward. That is because the hydrogeologist who created the figure in the report determined errors by subtracting observations from estimates rather than vice versa, which is the correct way to do it and which is the way the errors shown in Figure 1 were determined. That mistake was not trivial. It prevented the hydrogeologist from discerning the actual cause of the non-zero correlation between errors and declining water levels over time.

A sufficient cause for the rise of errors with the decline of water levels is that when water levels go down, estimates follow them down, as shown by the bottom two trendlines in Figure 1. Meanwhile, as shown by the top trendline there, errors—being negatively correlated with estimates, as a result of calibration—go up. The cause, at least the demonstrated cause, of the worrisome non-zero correlation is not the model; the cause, at least the sufficient cause, is the calibration of the model. The hydrogeologist, not the model, is the real problem...

3. Tuning in Climatology

Can the same be said about at least some climatologists in their modeling of world temperature over time? Figure 2, which is a copy of the figure on page 91 in *Unsettled* [7], shows rising observed and estimated mean-global-surface-temperature "anomalies" over time for 26 different models, where the anomalies are departures from the mean global surface temperature between 1880 and 1910. (Koonin in *Unsettled* cites the original source of Figure 2, [11], which is the source of the copy shown here.) Each plotted point is an eleven-year average. The four dark black lines represent separate sets of observations, and the 26 light black lines represent the 26 model estimates, which more or less follow the observations. Among other things, the models generally vary in their tuning practices.

The apparent grey swath weaving around the dark lines and encompassing the bulk of the model estimates represents them as a group, whereas the substantial variation of the model estimates provides some credence for Koonin's choice of the title for his book, *Unsettled*.

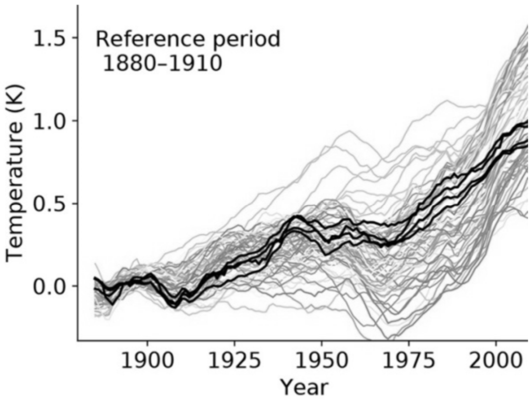

Figure 2. *Tuned model estimates (26 grey lines) of observed temperatures (4 black lines) with the apparent grey swath showing the trend of the bulk of the estimates over time.*

4. Global Warming

From about 1970, both the observations and the estimates in Figure 2 show a rather steep rise in mean global surface temperature, now popularly identified as "global warming." As the use of tuning might predict, the rise is steeper for the estimates than for the observations: Just as calibration resulted in underestimation of falling water levels, so here tuning results in overestimation of rising temperatures. How much is that overestimation? Koonin toward the end of Chapter 4 in [7] provides information that may suggest an answer to that question. In a so-called "budget analysis," he compared the mean global temperature rise over the past 140 years with total human and natural forcings (measured in Watts per square meter) that occurred during the same period and, after some correction of the data, showed that model tuning may have led to overestimating the effect of human influences on global warming by a factor as high as two.

5. Sensitivities

Terminology varies among statisticians and users of statistics in different fields. That variation can obscure the occurrence of mistakes in the use of statistics by non-statisticians. Terms used for the weights in estimates consisting of weighted sums provide an apt example. Some simply refer to the weights as constants. Statisticians call them parameters, and that could lead a statistician to misinterpret the term "parameter adjustment" when used by hydrogeologists and climatologists. The statistician might think that term meant the development of a new model, with new weights, to reduce error variation. The reason for that misinterpretation is that hydrogeologists and climatologists understand the word "parameter" to mean not a weight but the value of the independent variable to which the weight applies. By "parameter adjustment," they mean adjustment of variables, not constants—in other words, the adjustment of data. So, what terms do hydrogeologists and climatologists use to identify a model's weights? Interestingly, they both use the same term: "sensitivities."

The sensitivity that is of particular interest in the study of global warming is the weight that applies to the concentration of CO_2 in the atmosphere. The modelers producing the estimates shown in Figure 2 generally agree that the "equilibrium" value (Equilibrium Model Sensitivity, or EMS) of that sensitivity should be equal to about 3.0 degrees Centigrade (C). That means that doubling the concentration of CO_2 in the atmosphere from its value prior to the use of fossil fuels would increase the mean global temperature by about 3.0°C, provided no other influences, which climatologists call "forcings," were affecting it. As noted earlier, reflecting the different slopes of the curves for the estimated and observed surface temperatures in Figure 2, that number might be too high, perhaps by a factor of two.

In support of this possibility, the actual sensitivity for CO_2 concentration in climatology models (Transient Climate Response, or TCR) over the years has tended to hover around 1.5°C, half the 3.0°C EMS (e.g., Nijsse et al. [11] and Koonin [7]). Climatologists generally believe that the EMS is the correct long-term value for that sensitivity because transient conditions in zones might tend to lower the steepness of observation curves. Typical among those conditions are changing cloud formations, decreasing aerosol emissions, and melting icebergs. Perhaps even more to improve the fit of their models to data, climatologists use tuning to help guide their development of models having TSR values which are increasingly close to the EMS. Nijsse et al. [11] provides examples of that practice.

6. Uneasiness of Climatologists with Tuning

Although almost all the models cited in the 113 pages of the Fiato & Marotzke et al. [3] chapter in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (AR5) employ tuning, the word "tuning" appears there only 14 times. That number includes once in the table of contents and once in the reference list. Of the chapter's 1,423 references, only one [9] contains the word "tuning" in its title! Contrast that with the number of appearances in that chapter of the word "cloud" or "clouds" (165), "aerosol" (120), "ice" (333), and "ocean" or "oceans" (638). Although AR6 is not yet complete, neither the word "model" or "models" nor the word "tuning" appears in the title of any of the report's 12 listed chapters.

Interpreting such information to be indicative of a deliberate lack of transparency, the 15 authors of Hourdin et al. [5] ascribe it to an uneasiness within the climatology community over its use of tuning, especially since 22 of the 23 modeling centers contacted in a survey reported in the article said their models had used tuning, and all responded that they believed tuning to be important in model development. Climatologists evidently have two minds about tuning.

According to Hourdin et al. [5], its title being "The Art and Science of Climate Model Tuning," the practice of tuning is partly subjective and partly objective. Uneasy about the subjective part, the article's authors cite as authoritative support for the use of tuning an article by the highly-respected statistician R. A. Fisher [4] identifying "parameter estimation" as one of three steps comprising the process of model development. As noted earlier, to those authors, but not to Fisher, parameter estimation meant the partially subjective process of estimating an independent variable. To Fisher, it meant an entirely objective process of estimating the weights in a weighted sum of independent variables, commonly to minimize error variation. So, the citation of Fisher was hardly authoritative support for the subjective part of tuning. To the extent that subjectivity plays a part in it, climatologists have every reason to be uneasy about tuning.

Subjectivity-objectivity, however, is not the correct scale to use in evaluating the practice of tuning. As the next section will show, the correct scale to use is the right-wrong one, and on this scale, regardless of the extent of subjectivity or objectivity involved in the practice, tuning is simply wrong. "Parameters" as the term is used by both hydrogeologists and climatologists are data, and Fisher in [4] was not endorsing the adjustment of data.

Whether tuning or calibration, the adjustment of the value of any independent variable during model development, regardless of whether it is an increase or a decrease, will produce trends in estimates that are steeper than the trends in their corresponding observations. Interpretation of those exaggerated trends as forecasts of the future or as corrections of data have no more validity than the reading of tea leaves.

Simply the alteration of an independent variable's value itself, however, does not constitute tuning or calibration. An alteration made to use a model to estimate a future event when the value of the independent variable may differ from its current value is prediction, not tuning or calibration. The difference is that in tuning and calibration the observations being estimated remain unchanged throughout the process whereas in prediction those observations are free to vary.

Hydrogeologists and climatologists have a notable difference in their evaluation of the results of "parameter" adjustment. Whereas the hydrogeologists cited in Weitzman [14] blamed their calibrated models for being too slow to catch up with the data, climatologists have tended to blame the data for being too slow to catch up with their tuned models. Neither the models nor the data are to blame, however. The blame belongs entirely to the practice of "parameter" adjustment itself.

7. What Is Wrong with Tuning

By exaggerating upward or downward tendencies of observations over time, tuning corrupts data. When it is done to improve the appearance of a model or to help produce desired model predictions, it descends to the level of cheating. It is like using a cheat sheet to answer questions you would otherwise get wrong on a test. On the binary scale of right and wrong, regardless if the motivation, it is simply wrong.

Why? Every useful independent variable in a model uniquely increases the predictable portion of each observed measurement the model has been developed to estimate while simultaneously reducing the measurement's unpredictable portion, in other words, its error. So, error, by definition, is unpredictable. By creating a negative correlation between model estimates and errors, however, tuning, like calibration, makes errors predictable. That, as noted by Weitzman [14], is an oxymoron, which should be anathema to every member of any research community.

A mindset that allows tuning to help a model achieve a desired purpose can allow not only its extreme use but also the use of other forms of such motivated data manipulation. In Chapter 4 of *Unsettled* [7], Koonin cites a glaring example. Global warming being a United Nations concern, the models described in Figure 2 come from countries all over the world. To correct for a prediction of over twice as much global warming as was actually observed, some highly-regarded German climatologists [10] tuned one of their independent variables by a factor of ten from its initial value in their model-improvement process. Not to be outdone, a hydrogeologist cited in Weitzman [14] adjusted an independent-variable value without supporting data to be seven orders of magnitude lower than its initial value in zones crossed by a river to show that no aquifer beneath the river could possibly get any water from it. For the purpose of achieving a sensitivity of 3.0°C for CO_2 concentration, a modeler could avoid tuning altogether simply by fixing the sensitivity for CO_2 concentration at 3.0°C while allowing data to determine the sensitivities for the other independent variables in model development. Such blatant fudging of results should sound an alarm in every research community, not only climatology, to avoid the practice of data manipulation by any means to help achieve a desired purpose.

8. Yet More to the Story

In addition to the steep rise in mean global surface temperature from about 1970 onward, Figure 2 also shows an equally steep rise earlier, between about 1900 and 1940, prior to the steep rise in the use of fossil fuels. As shown by the grey swath in the figure, however, the model estimates rise about twice as steeply for the later than for the earlier period, a difference noted with concern by Koonin (in [7]). Koonin feared that the models were not sensitive enough to natural conditions, like a burst of unrecorded volcanic activity beneath the sea (author's, not Koonin's, example), causing the earlier rise that might also, possibly together with CO_2, be the cause of the later rise. Tuning might also help explain the difference.

Prior to tuning, the model estimates tracked the observations with mostly randomly-occurring under- and overestimations. That tendency applied to both the earlier and the later periods of steep observation rise, but only during the later period did the measured concentration of CO_2 rise precipitously. So, prior to tuning, the models could better, likely much better, account for the rise in that later period than for the rise in the earlier period. By exaggerating both rises, tuning improved the performance appearance of the models during the earlier period while having the opposite effect during the later period of steep observation rise.

9. What to Do Now

The United Nations' studies of global warming have employed enormous resources involving the use of many computers working in concert for months to analyze data from all the defined zones above the surface of the land and below the surface of the oceans throughout the world. Because of the enormity of the undertaking, much of the data collected is either unreliable or just an expert guess, a condition that invites and may, in some minds, even justify tuning. Some hydrogeologists who are aware of the problems resulting from their calibration of a model have in each instance resolved them by developing a new model based on the adjusted data in the calibrated one. Climatologists have done likewise, now in their sixth iteration (AR6), with increasingly unsatisfactory results motivating more rather than less tuning and showing increased divergence among modelers from iteration to iteration, duly noted by Koonin in *Unsettled* [7].

Perhaps in an excess of hutzpah, a statistician who is not a climatologist might offer the following possible solution to the tuning problem in climatology: Analyze the data obtained from a random sample of zones that is large enough to yield results having an acceptable margin of error and small enough for researchers to collect reliable data from all the zones in the sample. As observed by Hourdin et al. [5], a number of climatologists have already led the way, some via classical statistical methods (Bellprat et al. [1], Yang et al. [17], Zou et al. [19], and Zhang et al. [18]) and some via Bayesian ones (Rougier [12], Jackson et al. [6], Edwards et al. [2], and Williamson et al. [16]). Rather than being an exception, random sampling should become the norm. The number of zones needed to do that would be in the low thousands rather than the millions now under study.

Of the two statistical methods, the Bayesian one should require smaller samples. The difference could be considerable. In the field of survey methodology, Weitzman [15] provides this example: To achieve a .03 margin of error in a two-choice case, a survey which requires a sample of 1.067 using classical methods would require a sample of only 522 using Bayesian methods. What a .03 margin of error might mean in a sampling study of climate change is that the sample estimate of mean global surface temperature should differ from the population (all zones sampled from) mean global surface temperature by no more than .03 with odds of 20 to 1. Use of a Bayesian method could achieve error margins of .02 or even less with tractable sample sizes.

10. Conclusion

The planet is going through an interglacial period of global warming. Carbon dioxide fills little holes in the blanket of water vapor in the sky that helps keep the earth warm. Human activity that varies the production of CO_2 can somewhat affect the rate but cannot stop the occurrence of global warming. Despite what the cock might believe, the sun will continue to rise even if he stops crowing before daybreak. Guided by research in climatology, with due respect for Mother Nature, human beings in this century should plan for steadily rising seas resulting from melting glaciers and icebergs—along with other daunting challenges—created by increasingly warm nights. The name of the effort could be Project Noah. It has happened before.

References

[1] Bellprat, O., Kotlarski, S., Lüthi, D. & Schär, C. (2012). Objective calibration of regional climate models. *Journal of Geophysical Research*, 117, 1-13.

[2] Edwards, N. R., Cameron, D. & Rougier, J. (2011). Precalibrating an intermediate complexity climate model. *Climate Dynamics*, 37, 1469–1482.

[3] Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason and M. Rummukainen, 2013: Evaluation of Climate Models. In: *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[4] Fisher, R. A.: 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal. Society*, A222, 309–368.

[5] Hourdin, F., Mauritsen, T, Gentleman, A., Golaz, J., Venkatramani, B., Duan, Q., Folini, D., Ji, D., Klocke, D., Rauser, F., Rio, C., Tomassini, L., Watanabe, M. & Williamson, D. (2017). The Art and Science of Climate Model Tuning, *Bulletin of the American Meteorological Society*, 98 (3), 589-602.

[6] Jackson, C. S., Sen, M. K., Huerta, G. Deng, Y. & Bowman, K. P. (2008): Error reduction and convergence in climate prediction. *Journal of Climate*, 21, 6698–6709.

[7] Koonin, S. E. (2021). *Unsettled: What Climate Science Tells Us, What It Doesn't, and Why It Matters*. BenBella Books, Inc., Dallas, TX.

[8] Lomborg, B. (2020). *False Alarm: How Climate Change Costs Us Trillions, Hurts the Poor, and Fails to Fix the Planet*. Basic Books, New York.

[9] Mauritsen, T., Roeckner, E., Crueger, T, Stevens, B. et al. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, 4 (3), 1-18.

[10] Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V. et al., (2019). Developments in the MPI-M Earth System Model Version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *Journal of Advances in Modeling Earth Systems,* 11: 998–1038.

[11] Nijsse, F. J. M. M., Cox, P. M. & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models." *Earth System Dynamics* 11, 737–750.

[12] Rougier, J. C. (2007): Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, 81, 247–264.

[13] Shellenberger, M. (2020). *Apocalypse Never: Why Environmental Alarmism Hurts Us All*. Harper Collins, New York.

[14] Weitzman, R. A. (2021): From divining rods to statistics: a forensic analysis of the misuse of statistics in the estimation of environmental impact. *Chance*, 34 (4), 18-20.

[15] Wettzman, R. A. (2022). The determination of sample size in a Bayesian estimation of population proportions: how and why to do it in a regression framework *American Journal of Theoretical and Applied Statistics*, 11 (1), 13-18.

[16] Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. & Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, 41, 1703–1729.

[17] Yang, B. & Coauthors (2013): Uncertainty quantification and parameter tuning in the CAM5 Zhang-McFarlane convection scheme and impact of improved convection on the global circulation and climate. *Journal of Geophysical Research.* 118 (2), 395–415.

[18] Zhang, T., Li, L., Lin, Y. Xue, W., Xie, F., Xu, H. & Huang, X. (2015). An automatic and effective parameter optimization method for model tuning. *Geoscientific Model Development*, 8, 3579–3591.

[19] Zou, L., Qian, Y. Zhou, T. &Yang, B. (2014). Parameter tuning and calibration of RegCM3 with MIT–Emanuel cumulus parameterization scheme over CORDEX East Asia domain. *Journal of Climate*, 27, 7687–7701.