(In Global Climate Models, we trust?)
An Introduction to Trusting Global Climate Models and Bias Correction

Omar C. Gates & Richard B. Rood

Many tools, such as observing instruments and models, help climate scientists understand Earth's changing climate. Global climate models (GCMs) provide guidance for the future. Credibility is defined as having the attribute of scientific adequacy, and is necessary for determining the usability of GCM information. The credibility of GCMs is based, in part, on the ability of the GCM to represent past weather and climate. This evaluation with the past reveals strengths and weaknesses and contributes to the uncertainty description.

When considering GCMs as a source of knowledge about the future, it is important to understand its uncertainties. Models are built for purpose, and global climate models are expected to represent the global average surface temperature with some robustness. When looking at regional and local information, the uncertainty increases. The GCMs would not be expected to represent, with quantitative accuracy, the details of a region even as large as the Great Lakes. Hence, information about regions of interest to planners and managers require tailoring of information from multiple sources to be relevant or salient. This influences the uncertainty description.

Statistical comparison of simulations to observations of the past quantify how well the models represent both averages and variability. The average of the deviations of the GCM from an observational average is defined as bias. These biases can have a substantial impact on the interpretation of the future, and the ability to use the models in planning. Intuitively, if the bias is "small," it might be corrected in the tailoring process. However, climate model biases are often not small, and a question arises: in the presence of a large bias, what is the usability of model information? One of the most difficult aspects of climate model simulations to tailor for stakeholders is bias.

This white paper explores and frames the role of bias and bias correction in the usability of information from global climate models. In addition to credible results, GCMs must have plausible results. Plausible results are consistent with the known physical processes of a system. A GCM's ability to represent physical processes can be determined by evaluation of features such as weather systems (Hewitson et al., 2014). These values, credibility and plausibility, help establish trustworthiness and to determine the usability of these models.

For this narrative, model bias is defined as the average of the difference between a GCM's simulations and some baseline values. The baseline values can come from observations or, perhaps, a benchmark model simulation. Bias indicates a systematic difference, and bias relative to reliable observations is interpreted as "error." Presently, all climate models have bias, and those biases vary by location and season (Maraun, 2016). Model bias is used by scientists to diagnose errors and develop ways to correct them. For a practitioner using the GCM simulation, bias is a barrier to usability.

To better understand the role of model bias in usability, a distinction between predictive and derived model outputs is used. Predictive outputs include variables, such as temperature and precipitation, directly simulated within a climate model. Derived outputs, such as temporal averages and seasonal cycles, are calculated from primary variables. Stakeholders often use derived outputs, which may not, in fact, be consistent with the purpose of the original model design.

GCMs are complex and bias occurs from the accumulation of errors in the many smaller parts that make up the GCM. It is difficult to connect the cause of a bias to a specific deficiency in the GCM output. Biases in derived model outputs are more difficult to quantify because the required, additional calculations further obscure the sources of bias. Bias in model simulations is one of the most daunting challenges in improving the quality of simulations.

If a bias is small, it is intuitive that it might be corrected and used by practitioners. What is small? This depends upon the parameter, and perhaps, the state of the science of both observational and simulation capability. In some cases, a relative error of 10% as being small makes sense. This would be consistent with, for example, the development of the approximations of both theoretical and simulation foundations of climate models. In other cases, an absolute measure might be a more appropriate measure. For example, the temperature is well known and well simulated on weather scales, and absolute errors of 2°C, might be considered large; such an error would be much smaller than 10% on an absolute temperature scale.

There are other ways to consider bias errors, which are especially useful for derived quantities. A derived quantity of importance to the Great Lakes is the amount of water entering or leaving the Great Lakes – net basin supply. This relies on knowledge of precipitation, evaporation, and how much water runs off on the land. Calculations rely on both observations and climate models, leading to the fact that these two sources of information are rarely independent of each other.

Quantities such as net basin supply rely on calculation of budgets; in principle, a simple counting problem. However,the accounting is complex and has compounded errors. When parts the of budget are added together to calculate the water budget of the lake, our experience is that the outcomes are biased relative to observations. Hence, a measure of a small bias might be that it is small, for example <10%, relative to others parts of the budget. Another more absolute measure, perhaps, is relative to the sum of those terms.

Bias, and whether it is small or large, is a challenge to define and describe. It is safe to say, however, that in the case of many GCM products, both predictive and derived model outputs, bias is large. It is not rare to see biases greater than 100%, especially for a derived quantity, such as water supply for the Great Lakes.

There is a fundamental question of how to use model information in the presence of so much bias. The adage, “all models are wrong, but some models are useful,” comes into play. Some would maintain that the biases are so large that their use is suspect or perhaps even not ethical (Hewitson et al., 2014). Others would maintain that they are the best source of information we have, and we have problems we address. An answer is required. Therefore, we need some way to use the numbers form models, and that requires some sort of bias correction to the model output.

Bias correction acts as a statistical fix to a physical problem within the climate model. For the GCM to have plausibility, as defined above, a climate model must have the relevant atmosphere, land, and ocean physics represented credibly within the region of interest (Maraun, 2016). According to Maraun, 2016, “bias correction adjusts selected simulated statistics to match observations during a present-day calibration period.” However, bias correction should not be used to obtain plausible true future values (Maraun, 2016). For example, bias correction may add a lake effect to model-generated output; however, bias correction would not add lake interactions into a GCM that neglects these interactions in its initial runs.

Spatial downscaling is often used to tailor model simulations to localities. Statistical downscaling aims to tailor GCM output by generating statistically realistic variations at fine spatial scales and, possibly, day-to-day variations (Maraun, 2016). Bias correction is often used with statistical downscaling to adjust the statistics before mapping to finer resolutions. If the bias correction is large, then though it renders the climate information usable in the impact model, the bias correction does not, in fact, correct the scientific and structural deficiencies of the underlying GCM. This challenges both the interpretation of the results and the ethics of the use of knowledge whose underlying foundation has been altered.

Dynamical (regional) downscaling is a method that is also used to tailor GCM simulations to finer spatial and temporal scales. The GCM simulation is provided to a regional model which then evolves under the influence of the GCM's information. A major motivation for this approach is to incorporate local physical effects providing and inproved scientific and structural foundation. Though dynamical downscaling is expected to reduce some biases, particularly those related to the resolution of dynamical processes or topography (White, 2013), many sources of bias remain. Indeed, there in no necessity that bias will be reduced relative to global simulations.

First, the regional model inherits of biases from the GCM used to constrain the regional model. Then, the regional model will have its own set of biases associated with the same categories of model biases present in the GCM. Statistical bias correction would likely, still, be required for direct use in an impact model.

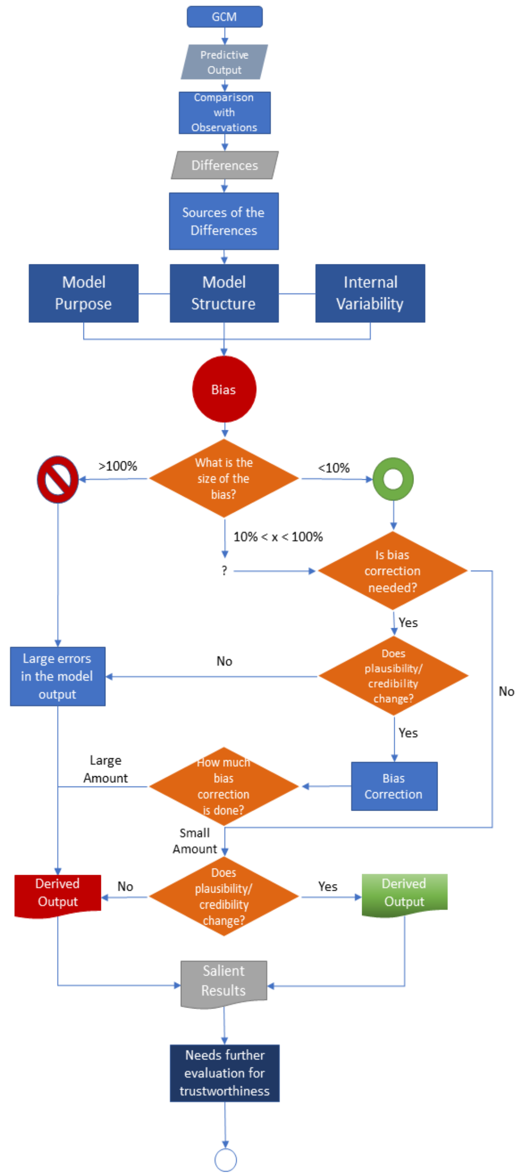


Figure 1: A predictive output's decision flow chart where practitioners will answer questions in the decision tree to determine whether the salient results are trustworthy.

Frameworks to Analyze the Role of Bias Correction in Adaptation Applications

Methods of bias correction vary based on the purpose and desired outcome. The delta method reduces the difference between the GCM-created values and observations by adjusting averages. Uniform variables such as temperature are addressed with this method. Another technique is the quantile method which removes inherent bias within the climate model, and the non-uniform variables such as precipitation are addressed with this approach.

Figure 1 provides a framework to discuss bias, and represents an important subset of the potential sources of bias; Model Purpose, Model Structure, and Internal Variability. The model's purpose is based on its intended application, and the model structure prioritizes the design to meet the needs of that application. Design decisions include grid resolution, spatial coverage, temporal scale, and physical processes that are included. The internal variability represents the dynamic range of the climate from the average and includes sources like the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO). In general, climate models are designed to represent internal variability in, at best, a statistical sense.

Although applying bias correction might produce salient results, that is, they fit into some historical ranges, the GCM's plausibility and credibility can be questioned. The deficiencies associated with fit for purpose, structure, and internal variability are not fixed by bias correction. They are hidden from the end user.

The flow chart in Figure 1 describes how outputs of predictive quantities can lead to salient results. Once the differences are identified, practitioners may see the varying amounts of bias: spatial variation, seasonal variation, variation from model to model, and variation from variable to variable. A natural question: if a variable is well represented in winter and not in summer, is there reason to assume winter is more correct for the right reasons?

Practitioners determine the necessity for bias correction. In this example we have presumed that there is a path for small, incremental, bias correction, where the bias is less than 10 percent. We also presume that if there is a 100 percent error, then there is a suggestion that the model's credibility and plausibility are of dubious merit. The right side of the decision tree in Figure 1 shows the multiple options available with small bias and bias between 10% and 100%. Small amounts of bias corrections can lead to derived outputs that create salient results compared to the original model. However, very large bias (greater than 100 percent) and large bias correction, as shown on the left side of the decision tree, can create derived outputs which produce salient results not representative of the model's purpose. The salient results will require more evaluation by looking at the derived outputs' process of dealing with bias.

Derived quantities, such as those from Regional Climate Models (RCMs), follow a more complicated decision tree for their application (Figure 2). RCMs improve simulation of physical processes by bridging the resolution gap between GCMs' simulations and local observations (Maraun, 2016). When compared to observations, these derived quantities highlight the behaviors of the GCM. The inherited bias from the parent GCM approximations occurs as a result of the differences from observations not being addressed, even with bias correction. Although the predictive output may contain bias, the RCM used for downscaling a GCM may introduce bias through the recalculations. The quality of the observational record determines whether a derived product accurately represents the local climate. Another interpretation for the quality involves the assurance of the observations being representative of the local climate.

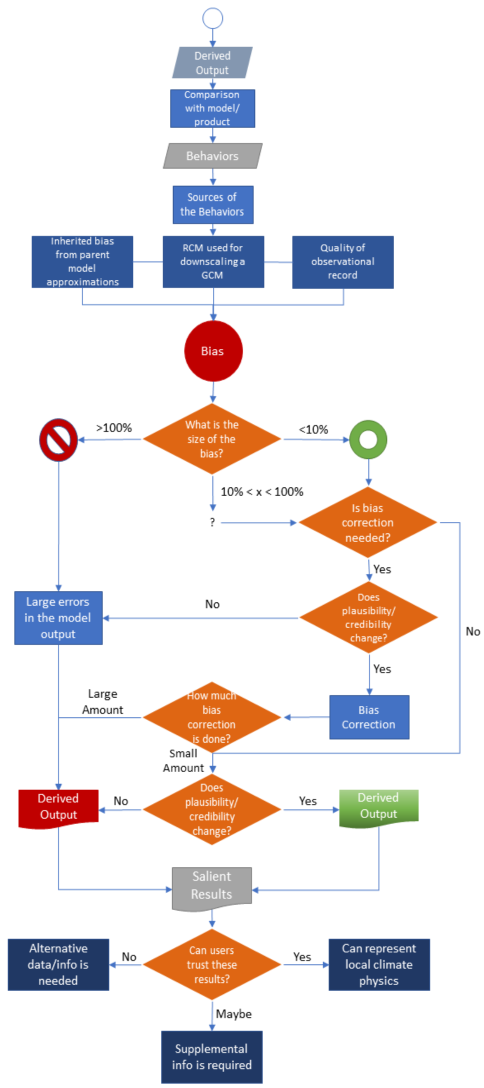


Figure 2: The further evaluation of a derived output's trustworthiness through understanding the model's behavior leading to salient results.

Practitioners will need to think about a similar series of questions to determine whether a derived output's use is justified. Figure 2 provides a flow chart of the steps and the questions practitioners will need to answer for a model's trustworthiness. Questions of plausibility and credibility will be present whether bias correction is done or not. Again, corrections of large bias will reduce both plausibility and credibility.

The final question practitioners must ask is whether they can trust the results from the GCMs derived or predictive products. The bottom of Figure 2 offers three pathways for answering this question. Practitioners can fully trust the data if they think the information gives a defensible answer to their research question. If the results do not give reliable and trustworthy information, then alternative data is needed for this situation. For practitioners unsure of whether to trust the salient results, integrating supplemental information from different sources can fill in the gaps where the models fall short of the mark.

RegCM4 Case Study

The Regional Climate Model Version 4 (RegCM4) is a collection of six dynamically downscaled GCMs available from the University of Wisconsin Nelson Institute. This is a primary souce of information used in GLISA.

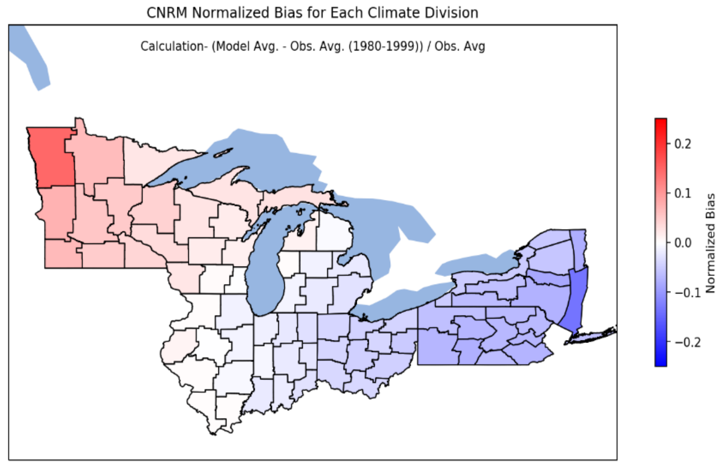


Figure 3: A normalized bias for each U.S. Climate Division based on the UW-RegCM4 model run for the CNRM Climate Model Version 5. (Preliminary Figure courtesy of Alexia Prosperi)

For this case study, the RegCM4's credible representation of the historical climate in the Great Lakes region is compared to the National Centers for Environmental Information (NCEI) Climate Divisions' observations. Figure 3 shows an example of the bias found in the Centre National de Recherches Meterologiques Coupled Global Climate Model Version Five (CNRM-CM5) historical model run compared to the observations. The states of Minnesota and New York have climate divisions (i.e., MN-D01 and NY-D05) with the highest amounts of bias in the region. Even with dynamical downscaling, there is bias within the historical model runs of the UW-RegCM4, and possible sources may come from an inherited bias of the parent GCM models or the quality of the historical observations. Further bias correction would not be useful for understanding the local climate in some areas, and the gathering of additional local information can aid in filling the knowledge gaps.

There are limitations in bias correction's ability to fix the physical errors within a climate model. It cannot fix a fundamental problem, one associated with the GCM (Maraun, 2016). Bias correction does not improve the skill nor add information absent from the model, and it can only adjust the marginal aspects of a climate model (Haerter et al., 2011 and Maraun, 2016).

The ultimate result of bias correction is creating salient results that may fit into the stakeholders' algorithm, but the scientific foundation and reasoning may be weak. Multiple sources provide various ways of creating a sensible climate action plan; however, planners and managers hesitate to use multiple sources due to the additional complications from the work and results (Hewitson et al., 2014). Working groups may rely on a single source of information for making informed decisions, and the resulting analysis may have errors leading to poor decision-making due to misinterpretations. Drawing on multiple sources demands achieving a good understanding of local changes even if there are no concrete action items created in the process (Hewitson et al., 2014). Clear communication between climate scientists and partners will provide an avenue to actionable steps towards sound decisions.

GLISA's Approach

The Great Lakes Integrated Sciences and Assessments (GLISA) center strives to provide regional partners with the most usable climate information for the Great Lakes region (Briley et al., 2021). We aim to assure that this information is plausible, defensible, and actionable (Hewitson et al., 2014; Briley et al., 2021).

The RegCM4 case study is an example of downscaling providing a better local simulation. Than global models, of the Great Lakes region. There is no doubt that the physical integrity, i.e. plausibility, is better than GCMs or statistical downscaling. However, the six RCM simulations still inherited the bias behaviors from their respective parent GCMs. They were a source of their own internal biases. Bias was not eliminated by the more robust physical representations.

GLISA strives to rely on of the underlying weather features and climate processes. An important part of GLISA's approach to the evaluation of climate models, both global and regional, is to be mindful of the representation of local geographical and meteorological details. For example, in the Great Lakes region, the lakes need to be represented and atmosphere-lake-land interactions are important to credibility, that is, scientific adequacy. A meteorological detail of importance is the lake effect, the roles of the lakes in modifying the temperature and precipitation in large regions surrounding the lakes. This brings attention to important parameters such as lake water temperature and the presence or absence of lake ice.

As an example, we determine that models misses a fundamental aspect of lake-effect precipitation such as formation of lake ice. We do not advocate correcting the associated biases in precipitation then moving forward as if those corrections are relevant for the future. These challenges credibility and plausibility, and raises ethical considerations about the decisions associated with such bias corrections.

Our approach to bias correction has been to document and report the bias. For quantitative purposes we rely on the delta method and adjust model-mean outputs on a regional, rather than a local scale. In this case, we strive to maintain the physical relationships between temperature and moisture that are represented, broadly and physically, within the global and regional models. With this form of bias correction, the ranges of variability suggested by the models are used as one estimate of uncertainty for which we can maintain a justifiable scientific basis.

To manage uncertainty, GLISA frames climate narratives around the observed historical changes for a local area. We develop narratives, scenarios, for the future by evaluating how the historical changes fit into the information provided by model simulations as well as constraints provided by analysis of the physical consistency, defined as plausibility, of the possible future climate states (Briley et al., 2017, 2021). That is, we use the model information as guidance, in concert with other sources of information.

GLISA uses scenario planning to work with partners and practitioners on climate adaptation projects. From a climate science perspective, we pose a range of plausible futures (Briley et al., 2017). The foundation of the scenarios includes observations and simulations of temperature, which provide our most robust knowledge. We, then, consider the primary ways that precipitation and hydrology plausibly respond to temperature. The details of the temperature response are guided by model simulations that represent, for example, higher or lower temperature changes in the future.

From these initial scenarios, another set of scenarios for derived variables, such as lake level or lake ice, is developed. These scenarios are based on historical variability, with the assumption that such modes of variability will continue in the future. We analyze whether there are physical reasons to expect these modes of variability will realize new extremes. For example, during sustained drought, high temperatures might support record low lake levels. During sustained wet periods, high temperatures support more precipitation and perhaps record-high lake levels. Analysis of historical use cases provides insight into how the atmosphere-lakes-land interact, which provides focus on like processes within the models.

This scenario approach keeps explicit the realistic prospect that the influence of climate change on a phenomenon, for example lake levels, is unlikely to be a single outcome. As in the past, the phenomenon has variable behavior responding to many influences. Climate change will change all aspects of the system, manifesting as the variability that is different from past variability.

Practitioners, then, consider the climate change scenarios in concert with management and planning scenarios that consider competing tensions.

Some practitioners do require quantitative input for their planning and decision making. In these cases, bias correction is, perhaps, required to allow their calculations, simply, to be performed. The bias and bias correction procedure need, minimally, to be identified and described. The bias correction procedure changes the uncertainty description in fundamental ways, and likewise, the foundation on which to interpret any results. GLISA guides these practitioners to data products suitable for their calculations. GLISA assures the data user is aware that the errors that cause the bias have not been corrected and limit the ability of interpretations (Briley et al., 2021).

Conclusion

Climate simulations are a critical part of climate science and the ability to plan for climate change. They provide guidance that in concert with other information allows framing of plausible futures. Interpretation of simulations as deterministic predictions is unwarranted. Even in the case of probabilistic interpretations, based on many simulations by many models, quantitative use of climate model output is hampered by many sources of uncertainty. Model bias is an indication of error in the model, and removal of bias has the potential of obscuring the errors and their existence.

However, simulations are only one part of the portfolio of information that is available for planning. As is shown every day in weather prediction, biased models provided usable information for making forecasts. They provide the skilled interpreter with a tool to understand complex relations among correlated processes. This tool, in combination with observations, underlying scientific principles, and a portfolio of models designed for interpretive purposes, allows us to develop plausible, defensible, and actionable scenarios. These scenarios help to manage, explicitly, uncertainty from known and unknown sources, as well as to frame the uncertainty of our knowledge of climate change relative to other uncertainties that decision-makers face.

References

Briley, L. J., Ashley, W. S., Rood, R. B., & Krmenec, A. (2017). The role of meteorological processes in the description of uncertainty for climate change decision-making. *Theoretical and Applied Climatology*, *127*(3–4), 643–654. https://doi.org/10.1007/s00704-015-1652-2

Briley, L. J., Rood, R. B., & Notaro, M. (2021). Large lakes in climate models: A great lakes case study on the usability of CMIP5. *Journal of Great Lakes Research*, (xxxx). https://doi.org/10.1016/j.jglr.2021.01.010

Climate Change, Agriculture, and Food Security (2014). *Bias Correction.* https://www.ccafs-climate.org/bias_correction/

Dee, D. Fasullo J., Shea, D., Walsh J. (2016, December 16). *The Climate Data Guide: Atmospheric Reanalysis: Overview and Comparison Tables.* University Corporation for Atmospheric Research (UCAR). https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables

International Research Institute for Climate and Society. (n.d.) *Climate Variability.* https://iri.columbia.edu/our-expertise/climate/climate-variability/

Haerter, J. O., Hagemann, S., Moseley, C., & Piani, C. (2011). Hydrology and Earth System Sciences Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci*, *15*, 1065–1079. https://doi.org/10.5194/hess-15-1065-2011

Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., & Jack, C. (2014). Interrogating empirical-statistical downscaling. *Climatic Change*, *122*(4), 539–554. https://doi.org/10.1007/s10584-013-1021-z

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., … Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3), RG3003. https://doi.org/10.1029/2009RG000314

Maraun, Douglas. (2013). Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *Journal of Climate*, *26*(6), 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

Maraun, D, & Widmann, M. (2015). The representation of location by a regional climate model in complex terrain. *Hydrol. Earth Syst. Sci*, *19*, 3449–3456. https://doi.org/10.5194/hess-19-3449-2015

Maraun, Douglas. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, *2*(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun, Douglas, & Widmann, M. (2018). Cross-validation of bias-corrected climate simulations is misleading. *Hydrol. Earth Syst. Sci*, *22*, 4867–4873. https://doi.org/10.5194/hess-22-4867-2018

Notaro, M., Bennington, V., & Lofgren, B. (2015). Dynamical downscaling-based projections of great lakes water levels. *Journal of Climate*, *28*(24), 9721–9745. https://doi.org/10.1175/JCLI-D-14-00847.1

Notaro, M., Bennington, V., & Vavrus, S. (2015). Dynamically downscaled projections of lake-effect snow in the Great Lakes basin. *Journal of Climate*, *28*(4), 1661–1684. https://doi.org/10.1175/JCLI-D-14-00467.1

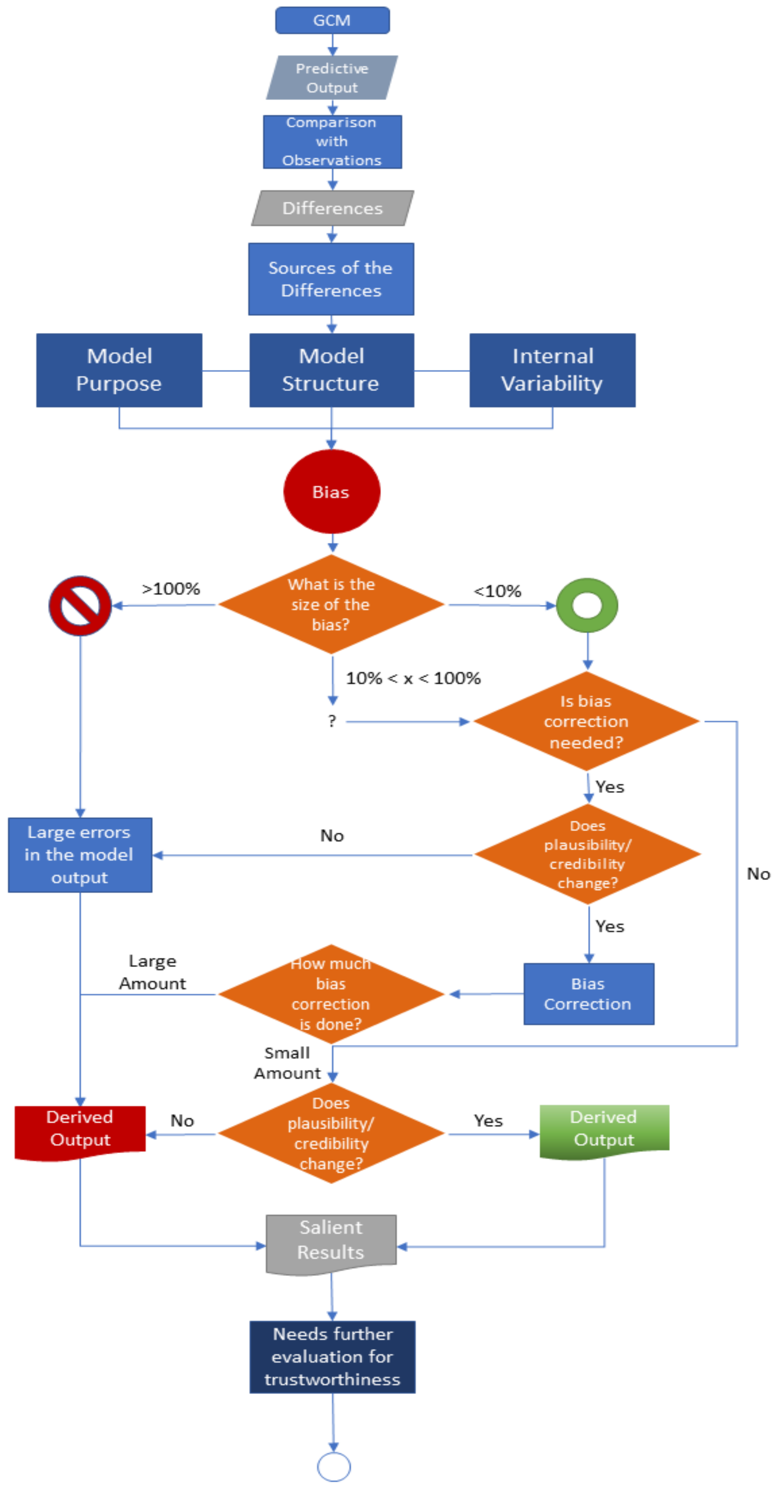
Stocker, T.F., Qin D., Plattner G.-K. Tignor M., Allen S.K., Boschung J., Nauels A., Xia Y., Bex V., and Midgley P.M. (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment of the Intergovernmental Panel on Climate Change.* IPCC. https://www.ipcc.ch/report/ar5/wg1/

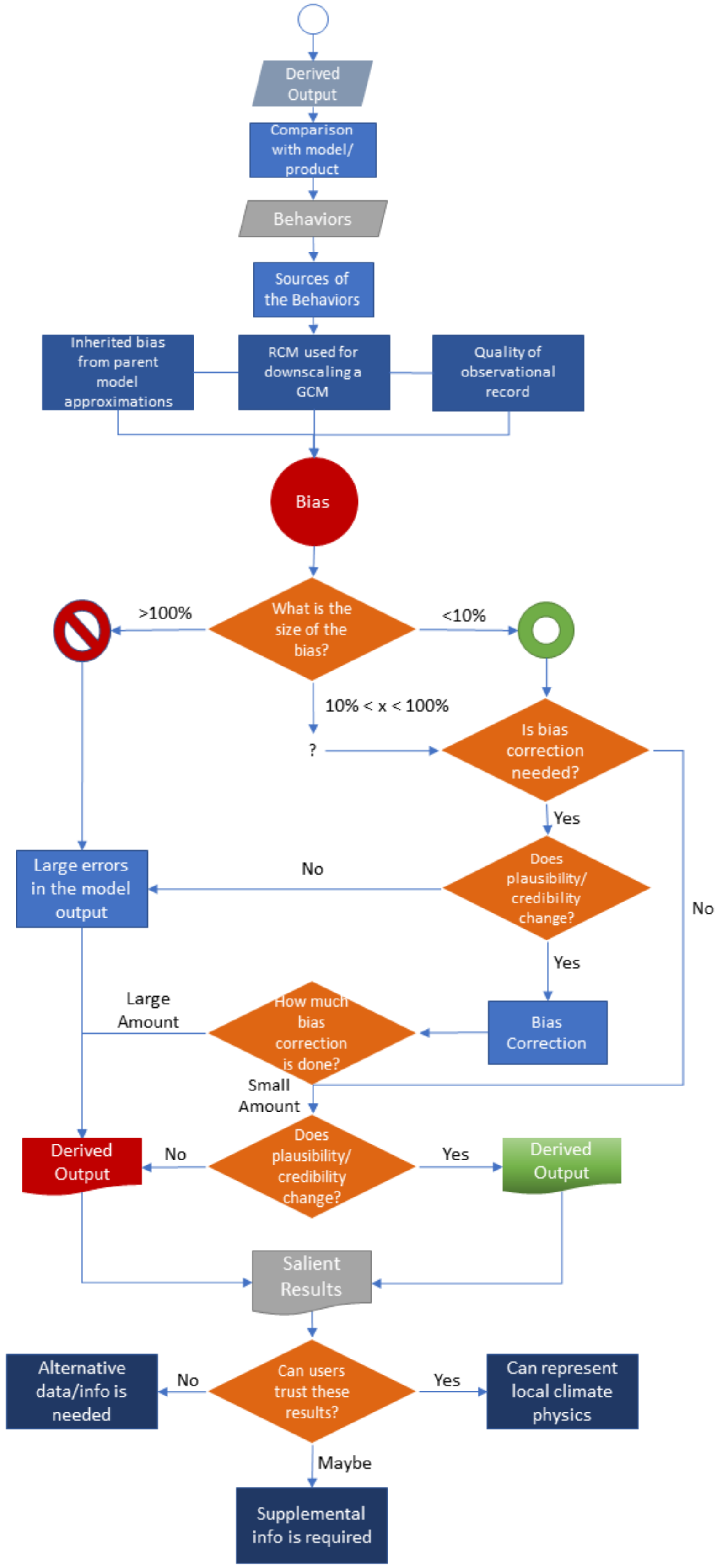
The Climate Data Factory (2018), *Technical Report: Bias Adjusting Climate Model Projections* Cochard, F., Loukos, H., and Noël, T. (2018) The Climate Data Factory.

White, R. H., & Toumi, R. (2013). The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters*, *40*(12), 2907–2912. https://doi.org/10.1002/grl.50612

Appendix

Larger images of Figures 1 and 2




Derived Output
Comparison with model/ product
Behaviors
Sources of the Behaviors
Inherited bias from parent model approximations
RCM used for downscaling a GCM
Quality of observational record
Bias
What is the size of the bias?
>100%
<10%
10% < x < 100%
?
Is bias correction needed?
Yes
No
Does plausibility/ credibility change?
No
Large errors in the model output
Yes
Bias Correction
How much bias correction is done?
Large Amount
Small Amount
Does plausibility/ credibility change?
No
Yes
Derived Output
Derived Output
Salient Results
Can users trust these results?
No
Yes
Maybe
Alternative data/info is needed
Can represent local climate physics
Supplemental info is required