

Research Article

Comparison of Parametric and Nonparametric Methods for Analyzing the Bias of a Numerical Model

Isaac Mugume,¹ Charles Basalirwa,¹ Daniel Waiswa,¹ Joachim Reuder,²
Michel d. S. Mesquita,³ Sulin Tao,⁴ and Triphonja J. Ngailo⁵

¹Department of Geography, Geoinformatics and Climatic Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda

²Geophysical Institute, University of Bergen, Allegaten 70, 5007 Bergen, Norway

³Uni Research Climate, Bjerknes Centre for Climate Research, Bergen, Norway

⁴School of Applied Meteorology, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 21004, China

⁵Department of General Studies, Dar es Salaam Institute of Technology, P.O. Box 2958, Dar-es-Salaam, Tanzania

Correspondence should be addressed to Joachim Reuder; joachim.reuder@uib.no

Received 17 February 2016; Accepted 4 April 2016

Academic Editor: Aiguo Song

Copyright © 2016 Isaac Mugume et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Numerical models are presently applied in many fields for simulation and prediction, operation, or research. The output from these models normally has both systematic and random errors. The study compared January 2015 temperature data for Uganda as simulated using the Weather Research and Forecast model with actual observed station temperature data to analyze the bias using parametric (the root mean square error (RMSE), the mean absolute error (MAE), mean error (ME), skewness, and the bias easy estimate (BES)) and nonparametric (the sign test, STM) methods. The RMSE normally overestimates the error compared to MAE. The RMSE and MAE are not sensitive to direction of bias. The ME gives both direction and magnitude of bias but can be distorted by extreme values while the BES is insensitive to extreme values. The STM is robust for giving the direction of bias; it is not sensitive to extreme values but it does not give the magnitude of bias. The graphical tools (such as time series and cumulative curves) show the performance of the model with time. It is recommended to integrate parametric and nonparametric methods along with graphical methods for a comprehensive analysis of bias of a numerical model.

1. Introduction

The models are used in many fields such as engineering, agriculture, health, business, and weather and climate for simulation and prediction. They help to understand the different subprocesses underlying a given process and have undergone tremendous improvements due to developments in computing technology. These models range from simple (e.g., linear regression models) to complex models (e.g., weather and climate prediction models); Glahn and Lowry [1] categorized the models as dynamical and statistical. A combination of dynamical and statistical models is also used in operational forecasting especially using statistical techniques to correct output from a dynamical model.

The national meteorological services usually operate high resolution numerical weather prediction models so as to

give accurate guidance to users of weather information [2]. The accuracy of a given model is the measure of how close the model predicted fields are compared to independently observed atmospheric fields [3, 4] but it can be affected by errors in initial conditions, imperfections in the model, and inappropriate parameterizations. When a model agrees with observations, the confidence in using the model is higher [5] but the present agreement does not necessarily guarantee the skill for the future model prediction.

The main advantage of models is their objectivity [1]. However, the presence of systematic errors is due to bias [6] which occurs due to difference in model response to external forcing [7] such as errors in initial conditions. This bias can manifest as overprediction or underprediction and is defined by the World Meteorology Organization as the mean difference between forecast values and mean actual

TABLE 1: Statistical bias measures of actual and model simulation for maximum temperatures.

Measure	arua	ebb	ksse	jinja	mbra	gulu
RMSE	2.19	7.50	4.41	8.86	2.47	2.37
MAE	2.01	7.22	4.05	8.60	2.24	1.98
ME	-2.01	-7.22	-4.05	-8.60	-1.92	-1.91
Rel. bias	-0.06	-0.26	-0.12	-0.28	-0.07	-0.06
BES	-2.06	-7.13	-4.03	-8.69	-2.21	-1.94
Skewness	0.80	-0.17	0.34	0.12	0.89	0.10
STM	-0.97	-1.00	-0.94	-1.00	-0.81	-0.81

observations [8] while Haerter et al. [9] define bias as time independent component of error in model output.

A couple of methods have been proposed to correct for the bias. Maraun [10] used quantile-quantile method and found that uncorrected regional climate models underestimated precipitation and produced many drizzle cases. Durai and Bhradwaj [11] investigated four statistical bias correction methods (namely, best easy systematic method, lagged linear regression, nearest neighbor, and running mean removal) and noted that the running mean and nearest neighbor methods improved the forecast skill. These methods attempt to reduce the bias in the next forecast using the information from the bias of the previous forecast [12]; however they influence the model output if prediction is based on bias corrected data [8] and they cannot correct improper representation of processes producing the model output [9].

Many studies have employed the parametric methods such as RMSE [13–15], MAE [14, 15], and ME [16] relative error [13, 16] to analyze the bias of numerical models but have put less emphasis on graphical tools as well as the nonparametric method. In the present study, we investigate the performance of the bias analysis methods on actual January 2015 temperature data and simulated temperature data using the Weather Research and Forecast (WRF) model (Tables 1 and 2). The rest of the paper is organized as follows: Section 2 describes the data sources, Section 3 presents overview of the methods of bias analysis, Section 4 presents results and discussion, and Section 5 gives summary and conclusion.

2. Data

We simulate January 2015 temperature using WRF model version 3.7 [17], with parameterizations schemes: WRF single moment 6-class scheme microphysics, the Kain-Fritsch cumulus parameterization, the Asymmetric Convective Model option for planetary boundary layer, the Rapid Radiative Transfer Model for longwave radiation, and the Dudhia scheme for shortwave radiation. This data is compared with observed January 2015 temperature (maximum and minimum temperature) data obtained from the Uganda National Meteorological Authority (UNMA). We use six stations (namely, Arua (arua), Entebbe (ebb), Kasese (ksse), Jinja (jinja), Mbarara (mbra), and Gulu (gulu)). For a given

TABLE 2: Statistical bias measures of actual and model simulation for minimum temperatures.

Measure	arua	ebb	ksse	jinja	mbra	gulu
RMSE	5.59	2.77	1.58	3.90	2.83	2.19
MAE	5.31	2.37	1.25	3.31	2.50	1.72
ME	5.31	-1.69	0.78	3.25	-2.39	0.21
Rel. bias	0.46	-0.09	0.05	0.21	-0.15	0.01
BES	5.23	-1.86	0.66	3.31	-2.29	0.11
Skewness	0.32	0.39	0.44	-0.06	-0.16	0.09
STM	1.00	-0.55	0.35	0.87	-0.94	-0.03

day and station, the maximum simulated temperature is compared with the maximum observed temperature and the minimum simulated temperature is compared with the minimum observed temperature.

3. Methods of Bias Analysis

In order to comprehensively investigate the performance of numerical models, it is important to evaluate them on many metrics other than using a single method [5]. In this section, we present the popular methods for analyzing bias of numerical models. The parametric methods are presented in Sections 3.1–3.6 while the nonparametric method considered is described in Section 3.7.

3.1. The Difference Measures. Willmott et al. [3] suggested a difference variable, D , given by the difference between the model predicted value, M , and observed value, O , that is,

$$D = M - O. \quad (1)$$

This is appropriate for point measurements. It is this measure that gives rise to other measures like the *root mean square error* (RMSE), the *bias* or *mean error* (ME), and the *mean absolute error* (MAE).

For a model j , with time-ordered data set $\{M_i\}$ we define the difference D_{ij} as follows:

$$D_{ij} = M_{ij} - O_i, \quad (2)$$

where i is the i th data point and O_i is the corresponding i th observed value from time-ordered actual observed data set $\{O_i\}$. A positive (negative) value indicates that model output is higher (lower) than the actual values.

3.2. The RMSE. The RMSE is the square root of the average squared differences (D_{ij}^2) and is a popular statistical measure for the performance of numerical model in atmospheric research [15]. For a model, j , the RMSE is thus defined as follows:

$$\text{RMSE}_j = \left[\frac{1}{n} \sum_{i=1}^n D_{ij}^2 \right]^{1/2}. \quad (3)$$

The RMSE is a good criteria to classify the accuracy of a model and a low index indicates higher accuracy.

3.3. The MAE. The MAE is the average of the magnitudes of differences (D_{ij} taken as positive) and is also a popular index for estimating bias in atmospheric studies. For a model, j , the MAE is defined as follows:

$$\text{MAE}_j = \frac{1}{n} \sum_{i=1}^n |D_{ij}| \quad (4)$$

and, just like RMSE, a low index indicates higher accuracy.

3.4. The Bias. The bias, also known as the mean error (ME), is obtained by averaging the differences (D_{ij}) over the number of cases. For a given model output, M_j , the ME is calculated from

$$\text{ME}_j = \frac{1}{n} \sum_{i=1}^n D_{ij}. \quad (5)$$

The magnitude of ME is equal to the MAE if all the predicted values of the model are higher (or lower) than the actual values. A value of bias close to zero indicates that model values are in fair agreement with actual values with zero implying no bias.

The relative bias is another bias measure suggested by Christakis et al. [16] in which ME is divided by average observations and given as follows:

$$\text{Rel. bias}_j = \frac{\text{ME}_j}{\bar{O}} \times 100\%. \quad (6)$$

The bias given by (5) and (6) gives both the direction and probable magnitude of the error.

3.5. The Skewness Coefficient. The skewness coefficient is a moment measure based on symmetry [18]. Having obtained the differences between the model and actual values (D_{ij}), positive (or negative) skewness indicates that model outputs are largely lower (or higher) than actual observations. The skewness coefficient is defined as follows:

$$\gamma_j = \frac{(1/(n-1)) \sum_{i=1}^n (D_{ij} - \bar{D}_j)^3}{s_j^3} \quad (7)$$

with s_j as the standard error of the sample biases forming a distribution $\{D_{1j}, D_{2j}, \dots, D_{nj}\}$ and calculated as follows:

$$s_j = \sqrt[2]{\left[\frac{1}{n-1} \sum_{i=1}^n (D_{ij} - \bar{D}_j)^2 \right]}. \quad (8)$$

3.6. The Bias Easy Systematic Method. The bias easy systematic (BES) method considers location measures (especially quartiles) and is given by Durai and Bhadravaj [11] as follows:

$$\text{BES} = \frac{q_1 + 2q_2 + q_3}{4}, \quad (9)$$

where q_1 , q_2 , and q_3 are the sample lower quartile, median, and upper quartile, respectively, of the differences, D_{ij} , and it is commended for its robustness for taking care of extreme values by Woodcock and Engel [12].

3.7. Sign Test Method. The sign test method (STM) is a nonparametric method based on assigning a score, $\hat{\theta}_{ij}$, that compares the prediction, M_{ij} , and observation, O_i , at a given point. If the model predicts higher values than observation ($M_{ij} > O_i$), we assign positive one (i.e., $\hat{\theta}_{ij} = +1$), if the model prediction is equal to observed value ($M_{ij} = O_i$), we assign zero (i.e., $\hat{\theta}_{ij} = 0$), and if the model predicts a value lower than observation ($M_{ij} < O_i$), we assign negative one (i.e., $\hat{\theta}_{ij} = -1$); thus

$$\hat{\theta}_{ij} = \begin{cases} +1: & \text{if } M_{ij} > O_i \\ 0: & \text{if } M_{ij} = O_i \\ -1: & \text{if } M_{ij} < O_i. \end{cases} \quad (10)$$

For a model j forming a distribution of scores, $\{\hat{\theta}_{ij}\}$, of size n , such that $\{\hat{\theta}_{1j}, \hat{\theta}_{2j}, \dots, \hat{\theta}_{nj}\}$, the mean is computed as follows:

$$\bar{\Theta}_j = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{ij}. \quad (11)$$

If the mean score, $\bar{\Theta}_j$, for a given model is positive, the model is generally considered to overpredict; if it is negative then the model underpredicts. Otherwise there is no significant bias. We suggest the hypothesis as

$$\begin{aligned} H_0: \bar{\Theta}_j &= 0, \\ H_1: \bar{\Theta}_j &\neq 0, \end{aligned} \quad (12)$$

and consider for unbiased model (i.e., zero bias)

$$\bar{\Theta}_j = E(\hat{\theta}_{ij}) = 0. \quad (13)$$

For a distribution of sample size less than 30 ($n < 30$), we propose the use of Student's t -distribution and make approximation to normal distribution for large samples ($n > 30$). The standard error is computed using

$$S(\hat{\Theta}_j) = \sqrt[2]{\left[\frac{1}{n-1} \sum_{i=1}^n (\hat{\theta}_{ij} - \bar{\Theta}_j)^2 \right]}. \quad (14)$$

The nonparametric statistic for measuring bias is then corrected and calculated using

$$Z(\hat{\theta}_{ij}) = \begin{cases} \frac{\bar{\Theta}_j - 1}{S(\hat{\Theta}_j)}: & \text{if } \bar{\Theta}_j > 0 \\ 0: & \text{if } \bar{\Theta}_j = 0 \\ \frac{\bar{\Theta}_j + 1}{S(\hat{\Theta}_j)}: & \text{if } \bar{\Theta}_j < 0. \end{cases} \quad (15)$$

We can then test this for a given significance level and make statistical inferences.

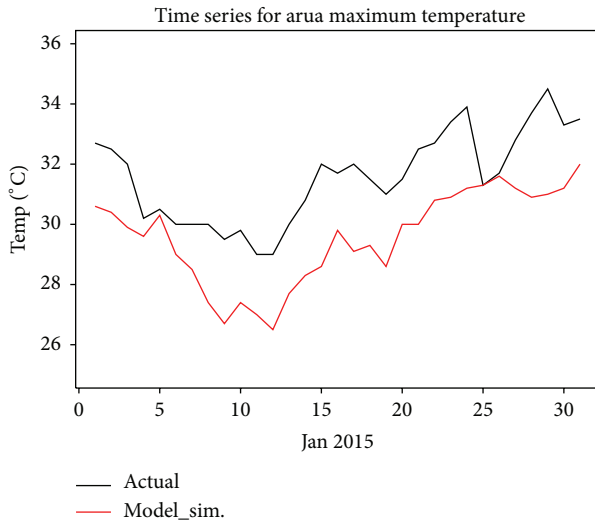


FIGURE 1: Arua: max_temp.

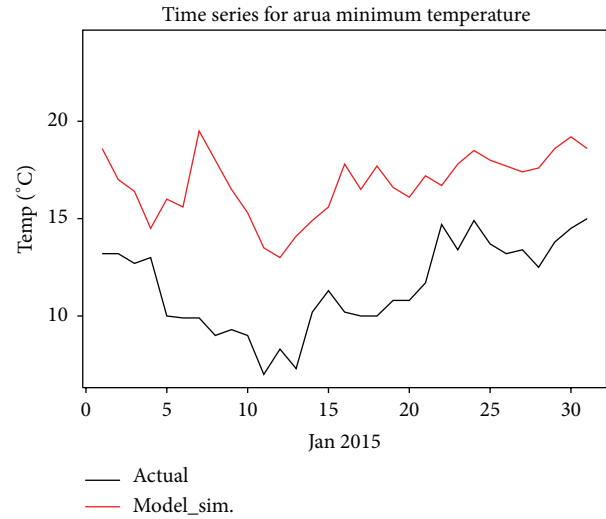


FIGURE 2: Arua: min_temp.

4. Results and Discussion

In comparing model results with observations, we assume that observed values are accurate and that it is the model predicted values that contain error because, as explained by Piani et al. [19], the models have inconsistencies that are sometimes not solved by bias correction. This thus brings the necessity of clearly determining the direction and magnitude of the bias. The magnitude of the bias can be affected by other factors, namely, the geographical location and season [11]. These factors are not considered in the study but it is possible to compare spatial and temporal bias using the different bias analysis methods.

Table 1 presents bias estimation using maximum temperatures as simulated by WRF model and actual observed values for maximum temperature while Table 2 presents bias estimation using model simulated values and actual observed values for minimum temperature. These tables help to explore the different possible cases and we obtain a negative bias for all maximum temperatures (Table 1) and some stations have positive bias for some minimum temperatures (Table 2). These cases are also presented using time series figures (Figures 1–12). The time series figures help to investigate how the biases change with time and the greater the departure from the curves (model simulated curve and observed curve), the greater the bias. For Gulu, (Figures 11 and 12) the model and actual observations follow roughly the same trend. For Kasese (Figure 6) there is high variability for actual minimum temperatures compared to those presented by model. For Jinja (Figure 7) actual observations have increasing trend while model values have a decreasing trend over the period (20–30 days). These results imply that a given model can have varying performance in different geographical regions, hence bias.

4.1. Traditional Bias Analysis Methods. The popular traditional parametric bias analysis tools were presented in Section 3. A discussion of these methods is presented below.

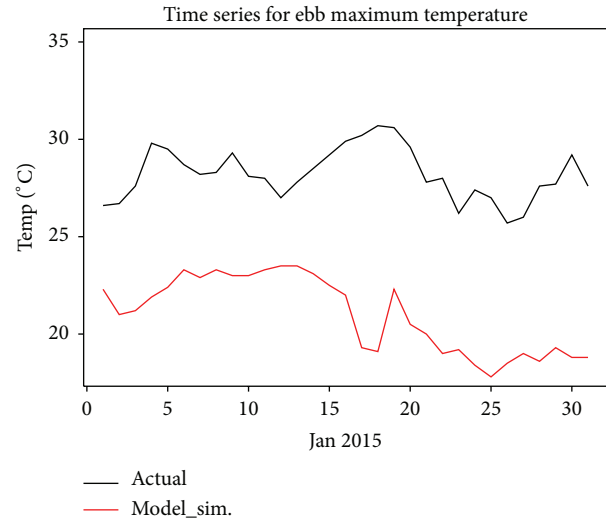


FIGURE 3: Entebbe: max_temp.

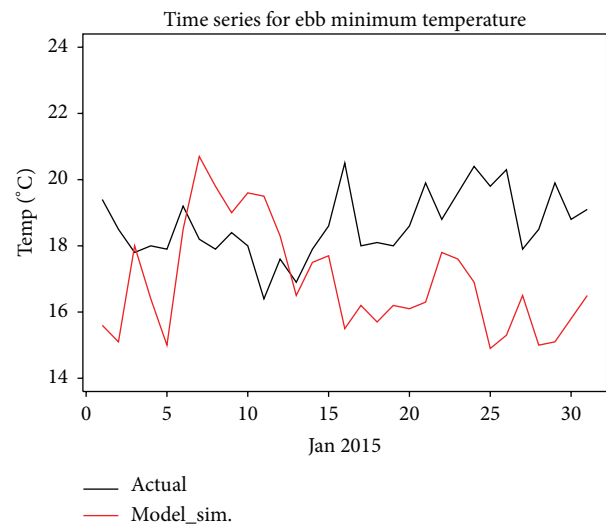


FIGURE 4: Entebbe: min_temp.

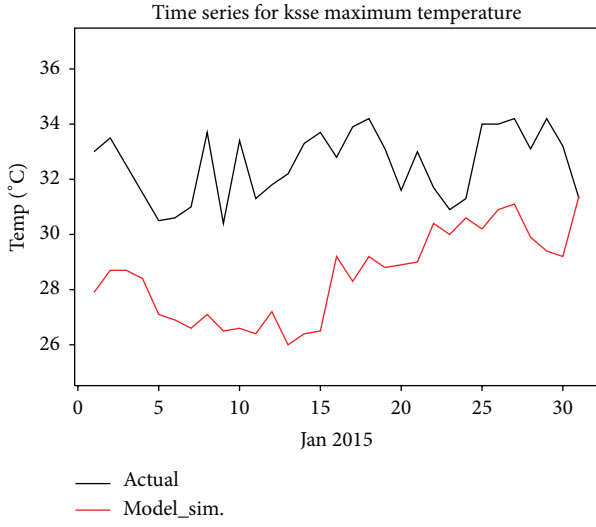


FIGURE 5: Kasese: max_temp.

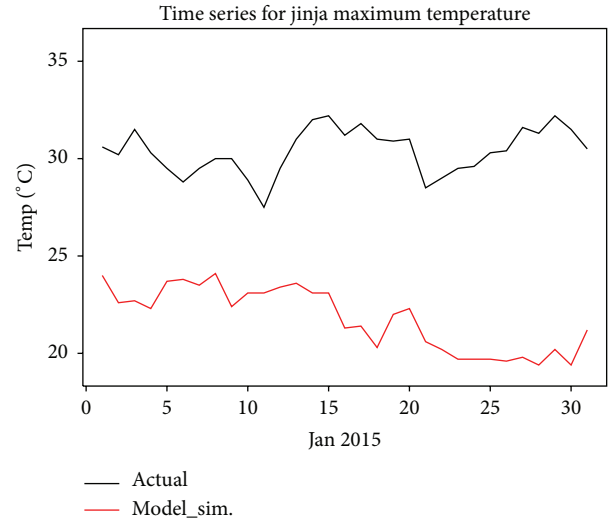


FIGURE 7: Jinja: max_temp.

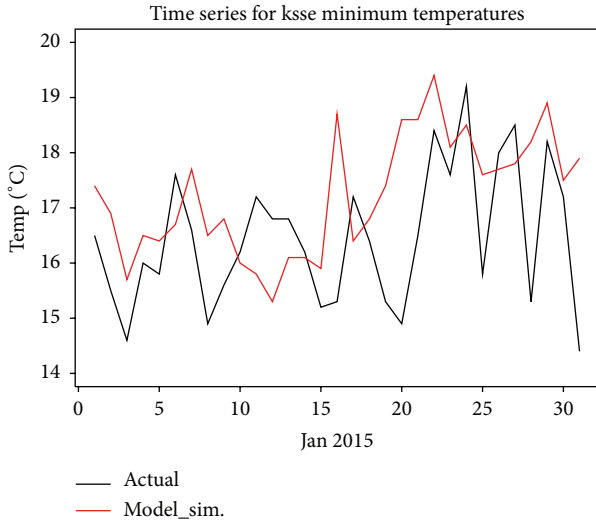


FIGURE 6: Kasese: min_temp.

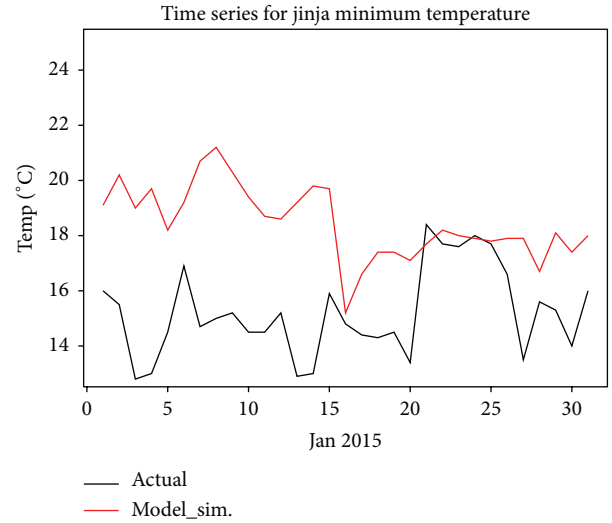


FIGURE 8: Jinja: min_temp.

The RMSE and MAE vary with both magnitude of error and sample size [15]. If an extreme event happens and is not correctly predicted (simulated) by the model, a big error will result and can manifest as outliers, thus distorting the index. The problem of estimating bias using the RMSE and MAE is as follows: (i) it does not show the direction of bias and (ii) it treats all the biases in one direction, thus amplifying the bias. The bias given by (5) and the relative bias defined by (6) are of great importance as they suggest both magnitude and possible direction of the bias. This is helpful as it indicates whether the model overpredicts or underpredicts the field being predicted, but, as explained by Knutti et al. [5], simple averaging (e.g., bias or ME) is not effective as it is affected by extremes and biases in different directions canceling. The BES is however a location measure and is less affected by extreme values.

4.2. The Sign Test Method (STM). In this method, we assign positive (or negative) one depending on the direction of the bias and then compute the mean of the assigned values. A value of mean greater (less) than zero indicates positive (negative) bias. By STM, we believe that the direction of the bias is preserved while not being influence by extreme values which occur rarely. For example, a model can have many drizzle days when in reality the days are dry but underpredicts a heavy rainfall event [8]. Aggregating these results using traditional bias estimation methods can lead to confusing results suggesting that the model has no or less bias than should be expected.

If the number of biases in opposite directions is equal, the STM will give a zero score. Although this may appear to be a drawback, its meaning is easily understood. It simply means that the model can equally overpredict or underpredict; however, it rarely occurs in numerical models. On the contrary,

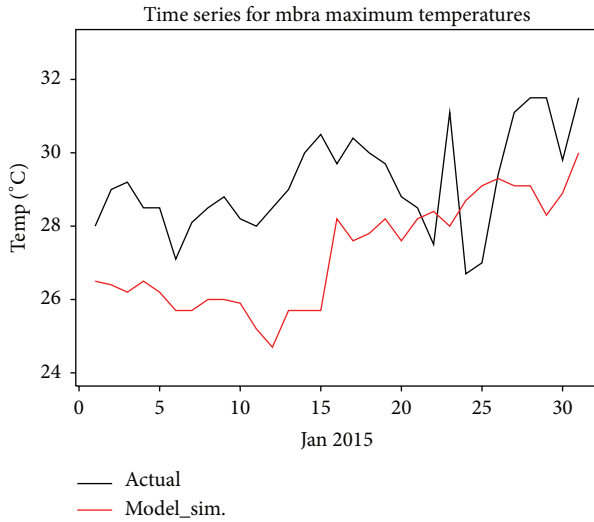


FIGURE 9: Mbarara: max_temp.

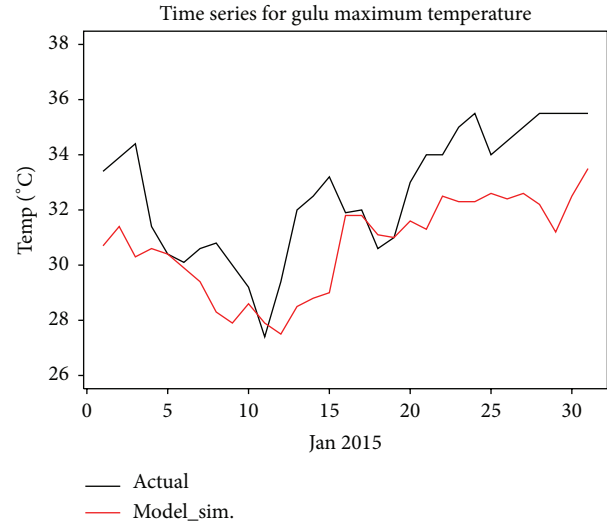


FIGURE 11: Gulu: max_temp.

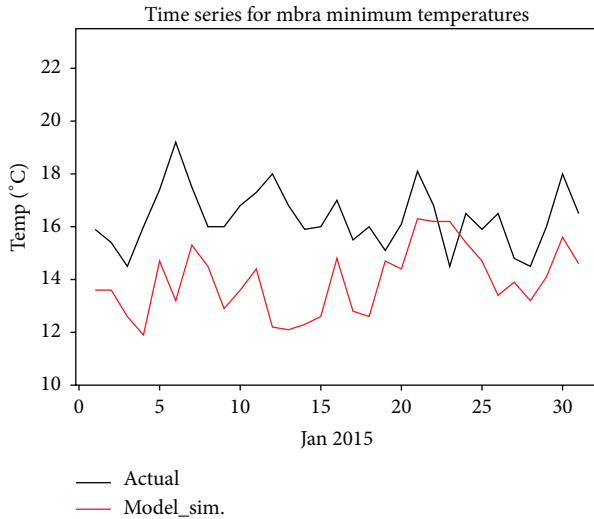


FIGURE 10: Mbarara: min_temp.

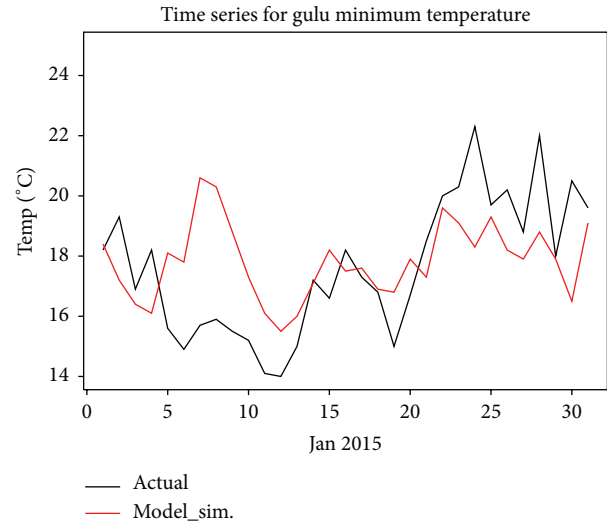


FIGURE 12: Gulu: min_temp.

if the other methods gave zero, the meaning would not immediately imply that the number of biases in one direction is exactly equal to the number of biases in the other direction and that there has been an offset. It could imply that the model is unbiased which could be misleading. The inferences made using STM statistic are based on general assumption that lead to some function of the sample observation whose sampling distribution can be determined without knowledge of the specific distribution function underlying the population [20]. The STM is also less concerned with the distribution of the population which is why it is noted to be affected by extreme values.

It is possible for the STM and the parametric methods to disagree (Table 2). In results presented by Table 2, for gulu, the STM gives a negative index while the ME gives a positive index. By STM, it means that the model had more value underpredicted than overpredicted which, unfortunately, was

weakly resolved by the ME. This probably means that there are cases of partial cancelation of values by the ME which is why it is giving a positive bias.

In principle, we believe that the direction presented by the STM should approach the direction presented by ME for a large sample of values.

5. Summary and Conclusions

The numerical models normally have both systematic and nonsystematic errors. The systematic errors manifest as bias in the model which may lead to either overprediction or underprediction. In the study, we analyzed the parametric methods of analyzing bias and compared them with STM but have not considered spatial bias and methods of correcting the bias.

The parametric methods are based on difference measure and the STM is based on assigning a score of +1 to positive biases and −1 to negative biases and then getting an average of these scores. We believe that STM is ideal for estimating bias in prediction or simulation of scalar geophysical variables (e.g., wind speed, rainfall amount, and temperature) by numerical models and that it is reliable and robust because the values presented are clear to understand as far as determining the direction of the bias is needed. The direction of the bias is needed in order to tune the model to correct for future biases. By STM a value of +1 (−1) indicates that all the values are higher (lower) than the actual ones. The STM can be used in inferences, thus reducing uncertainty, and is also based on a simple algorithm.

However, we do not suggest neglecting other measures but propose a complement because, in order to get a complete analysis of the data, it is important to compare both parametric and nonparametric tools [3]. We also recommend the use of graphical tools especially the density plots and investigating the skewness as well as tail properties. The time series plots can be used to investigate the performance of the model for extended period of time, with an intention of ascertaining whether the model worsens or improves with time. Lastly, while assigning ± 1 loses the magnitude of the bias, STM only helps to determine the direction of the bias.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors appreciate the WIMEA-ICT project for the support and the Uganda National Meteorological Authority for availing the temperature data used for model comparison. They also express sincere thanks to Godfrey Mujuni for organizing the temperature data used.

References

- [1] H. R. Glahn and D. A. Lowry, "The use of model output statistics (MOS) in objective weather forecasting," *Journal of Applied Meteorology*, vol. 11, no. 8, pp. 1203–1211, 1972.
- [2] M. Baldauf, A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, "Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities," *Monthly Weather Review*, vol. 139, no. 12, pp. 3887–3905, 2011.
- [3] C. J. Willmott, S. G. Ackleson, R. E. Davis et al., "Statistics for the evaluation and comparison of models," *Journal of Geophysical Research: Oceans*, vol. 90, no. C5, pp. 8995–9005, 1985.
- [4] M. Niu, S. Sun, J. Wu, and Y. Zhang, "Short-term wind speed hybrid forecasting model based on bias correcting study and its application," *Mathematical Problems in Engineering*, vol. 2015, Article ID 351354, 13 pages, 2015.
- [5] R. Knutti, R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, "Challenges in combining projections from multiple climate models," *Journal of Climate*, vol. 23, no. 10, pp. 2739–2758, 2010.
- [6] T. M. Smith, P. A. Arkin, J. J. Bates, and G. J. Huffman, "Estimating bias of satellite-based precipitation estimates," *Journal of Hydrometeorology*, vol. 7, no. 5, pp. 841–856, 2006.
- [7] C. Deser, A. Phillips, V. Bourdette, and H. Teng, "Uncertainty in climate change projections: the role of internal variability," *Climate Dynamics*, vol. 38, no. 3–4, pp. 527–546, 2012.
- [8] U. Ehret, E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert, "Should we apply bias correction to global and regional climate model data?" *Hydrology and Earth System Sciences*, vol. 16, no. 9, pp. 3391–3404, 2012.
- [9] J. O. Haerter, S. Hagemann, C. Moseley, and C. Piani, "Climate model bias correction and the role of timescales," *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 1065–1079, 2011.
- [10] D. Maraun, "Bias correction, quantile mapping, and downscaling: revisiting the inflation issue," *Journal of Climate*, vol. 26, no. 6, pp. 2137–2143, 2013.
- [11] V. R. Durai and R. Bhadraraj, "Evaluation of statistical bias correction methods for numerical weather prediction model forecasts of maximum and minimum temperatures," *Natural Hazards*, vol. 73, no. 3, pp. 1229–1254, 2014.
- [12] F. Woodcock and C. Engel, "Operational consensus forecasts," *Weather and Forecasting*, vol. 20, no. 1, pp. 101–111, 2005.
- [13] S. Tao, S. Shen, Y. Li, Q. Wang, P. Gao, and I. Mugume, "Projected crop production under regional climate change using scenario data and modeling: sensitivity to chosen sowing date and cultivar," *Sustainability*, vol. 8, no. 3, p. 214, 2016.
- [14] R. Shrivastava, S. K. Dash, R. B. Oza, and D. N. Sharma, "Evaluation of parameterization schemes in the WRF model for estimation of mixing height," *International Journal of Atmospheric Sciences*, vol. 2014, Article ID 451578, 9 pages, 2014.
- [15] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [16] N. Christakis, T. Katsaounis, G. Kossioris, and M. Plexousakis, "On the performance of the WRF numerical model over complex terrain on a high performance computing cluster," in *Proceedings of the IEEE International Conference on High Performance Computing and Communications, IEEE 6th International Symposium on Cyberspace Safety and Security, IEEE 11th International Conference on Embedded Software and Systems (HPCC, CSS, ICESS '14)*, pp. 298–303, Paris, France, August 2014.
- [17] W. Wang, C. Bruyere, M. Duda et al., ARW Version 3 Modeling System User's Guide. Mesoscale and Microscale Meteorology Division. National Center for Atmospheric Research, June, 2015, http://www2.mmm.ucar.edu/wrf/users/docs/user-guide_V3/ARWUsersGuideV3.pdf.
- [18] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, vol. 100, Academic Press, 2011.
- [19] C. Piani, G. P. Weedon, M. Best et al., "Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models," *Journal of Hydrology*, vol. 395, no. 3–4, pp. 199–215, 2010.
- [20] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, Springer, Berlin, Germany, 2011.

