

OPEN

# Statistically downscaled climate dataset for East Africa

DATA DESCRIPTOR

Solomon H. Gebrechorkos<sup>1,2</sup>, Stephan Hülsmann<sup>1</sup>  & Christian Bernhofer<sup>2</sup>

Received: 30 October 2018

Accepted: 14 March 2019

Published online: 15 April 2019

For many regions of the world, current climate change projections are only available at coarser spatial resolution from Global Climate Models (GCMs) that cannot directly be used in impact assessment and adaptation studies at regional and local scale. Impact assessment studies require high-resolution climate data to drive impact assessment models. To overcome this data challenge, we produced a station based climate projection (precipitation and maximum and minimum temperature) for Ethiopia, Kenya, and Tanzania using observed daily data from 211 stations obtained from the National Meteorological Agency of Ethiopia and international databases. Moreover, 26 large-scale climate variables derived from the National Centers for Environmental Prediction reanalysis data (1961–2005) and second generation Canadian Earth System Model (CanESM2, 1961–2100) are used. Statistical Down-Scaling Model (SDSM) is used to produce the required high-resolution climate projection by developing a statistical relationship between the large- and local-scale climate variables. The predictors are analysed more than 16458 times and we provided 20 ensembles for the current (1961–2005) and future (2006–2100, under RCP2.6, RCP4.5, and RCP8.5) climate.

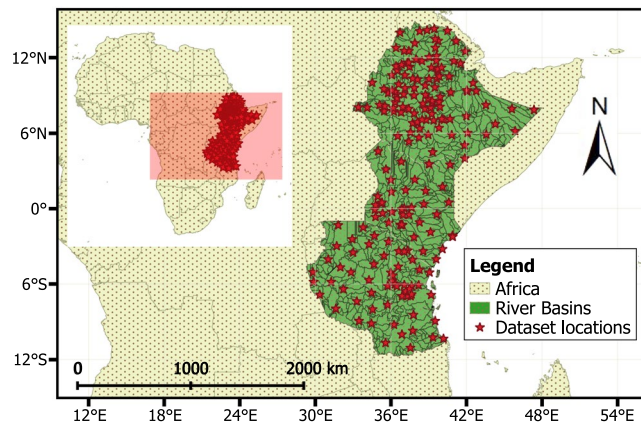
## Background & Summary

Large-scale or global climate models are currently used to advance the scientific knowledge and understanding variabilities and changes in large-scale climate variables<sup>1</sup>. Information obtained from Global Climate Models (GCMs) supports a better understanding of the climate at a global scale<sup>2,3</sup>. The output from GCMs is too coarse (>100 km) to be used in impact assessment studies, adaptation planning, and decision-making process at local or regional scale<sup>4,5</sup>. In addition to the coarse resolution, biases and uncertainties associated with GCMs increase from global to regional and local scales, which limit the suitability and applicability of GCMs in local-scale impact assessment studies<sup>6–9</sup>. Therefore downscaling is required to increase the spatial resolution and reduce biases<sup>1,10</sup> before climate projections can be used for impact assessment and adaptation planning. During the last few decades, two types of downscaling techniques have been introduced to reduce biases and improve the spatial resolution of GCMs. These are dynamical and statistical downscaling methods. Dynamical downscaling (e.g., CORDEX-Africa, <http://www.cordex.org/domains/region-5-africa/>) as a climate modeling process includes local information such as topographic features to produce a high-resolution climate projection, for example 50 km for Africa<sup>11,12</sup>. In addition to the complexity of models and high resource requirements, dynamical models still face biases and sensitivity to the boundary condition of GCMs, which limit their use in local-scale impact assessment and adaptation studies<sup>13–15</sup>. However, compared to dynamical models, statistical downscaling models are fast, simple, effective, and require less computational capacities and expenses.

Statistical models are designed to produce a location based weather series, equivalent to station data, by developing a statistical relationship between local-scale (predictands) and large-scale climate variables (predictors). Considering the overall advantages, statistical downscaling models are widely used in impact assessment studies at local and regional scale in sectors such as water resource and agriculture<sup>10,13,15,16</sup>. Statistical models are classified under three categories based on the statistical approaches used; stochastic weather generator<sup>17</sup>, weather typing<sup>18</sup>, and transfer function<sup>19</sup>.

The Statistical DownScaling Model (SDSM)<sup>19</sup> is one of the most widely used statistical downscaling models, which is developed based on a transfer function and stochastic weather generator. The performance of SDSM was found to be higher than the conventional weather generators<sup>20,21</sup>. In addition, compared to other statistical downscaling models, SDSM possesses better capabilities in capturing rainfall characteristics and maximum and

<sup>1</sup>United Nations University Institute for Integrated Management of Material Fluxes and of Resources (UNU-FLORES), 01067, Dresden, Germany. <sup>2</sup>Faculty of Environmental Sciences, Institute of Hydrology and Meteorology, Technische Universität Dresden, 01062, Dresden, Germany. Correspondence and requests for materials should be addressed to S.H.G. (email: [gebrechorkos@unu.edu](mailto:gebrechorkos@unu.edu))



**Fig. 1** Location map of ground stations and produced datasets (daily precipitation, maximum and minimum temperature) inside river basins (polygons in the map) of Ethiopia, Kenya, and Tanzania.

minimum temperature<sup>19,21</sup>. Therefore, SDSM is used to produce a location-based high-resolution climate projection for regions of East Africa (Ethiopia, Kenya, and Tanzania). Regions of Africa, particularly East Africa, are highly vulnerable to changes in climate and climate extremes and more extreme events such as frequent droughts, floods, and heavy rainstorms are projected in the future<sup>22</sup>. Therefore, considering the observed changes and vulnerability of the region to variability (e.g., seasonal rainfall variability) and changes in climate and climate extremes<sup>22,23</sup> conducting in-depth impact assessment studies at local and regional scale is required to minimize or mitigate impacts in the future through sustainable adaptation measures. However, this type of information is not readily available and producing station based climate projections using SDSM requires observed data with high quality for model calibration and as input to the scenario generator, which is part of SDSM. It is used to generate, after model calibration and validation, an ensemble of synthetic weather series, using daily predictors supplied by a global climate model<sup>15</sup>.

Availability of observed data from ground-based meteorological stations is, however, limited in East Africa due to issues such as limited temporal and spatial coverage, quality, and accessibility (e.g., data sharing policies). For example, from Tanzania, only five stations with maximum coverage of five-years can be provided by the meteorological agency. Moreover, the Kenyan meteorological agency only provides monthly data, which cannot be used in statistical downscaling. Therefore, a combination of datasets, station data obtained from the National Meteorological Agency (NMA) of Ethiopia and daily data available at the National Centers for Environmental Information (NCEI) are used. For areas with no ground observation (remote and data sparse parts of the region), additional datasets from remote sensing and reanalysis based products with high accuracy (compared with ground station data) and covering a large part of the region are used<sup>24</sup>. Compared to previously developed datasets such as CCAFS ([http://www.ccafs-climate.org/data\\_spatial\\_downscaling/](http://www.ccafs-climate.org/data_spatial_downscaling/)), which, based on the Delta method, provides a monthly average of 30 years period at different spatial resolution, we provide point information at a daily time scale. Moreover, compared to CCAFS our dataset can be used without restrictions. In this paper, we present statistically downscaled daily precipitation and maximum and minimum temperature for the current climate conditions (1961–2005) and future climate scenarios (2006–2100) under three Representative Concentration Pathways (RCPs; RCP2.6, RCP 4.5, and RCP 8.5). The data can be used for impact assessment and adaptation studies in Ethiopia, Kenya, and Tanzania (Fig. 1).

## Methods

The observed daily precipitation and maximum and minimum temperature data used in this study are the most comprehensive to date in the statistical downscaling process and for this region. Here, we used only stations with higher quality (e.g., concerning missing values) and temporal coverage in order to identify the most dominant predictors and develop the most accurate future climate scenarios.

**Data Acquisition.** Observed daily precipitation and maximum and minimum temperature during the period of 1961–2005 is obtained from the National Meteorological Agency (NMA) of Ethiopia and National Centers for Environmental Information (NCEI). For data sparse parts of the region, additional daily precipitation and maximum and minimum temperature (T-max and T-min), based on our earlier study on climate data evaluation for East Africa<sup>24</sup>, are used. For regions with limited availability of station data, climate data products with high spatial and temporal resolution can be used to bridge data gaps<sup>25</sup>. For East Africa, we evaluated different daily climate data sources based on climate models, remote sensing, and reanalysis data and the most accurate data sources are identified for application in climate and hydrological studies. From this study, the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)<sup>26</sup> and Observational-Reanalysis Hybrid<sup>27</sup> were identified for precipitation and T-max and T-min, respectively (for more information see<sup>24</sup>). In addition, large-scale climate variables (predictors) for the current climate and future scenarios under the RCPs are used, which is available from the Canadian Climate Data and Scenarios (<http://climate-scenarios.canada.ca/>).

No.	Long name	Short name
1	Mean sea level pressure	mslp
2	Surface airflow strength	p1_f
3	Surface zonal velocity	p1_u
4	Surface meridional velocity	p1_v
5	Surface vorticity	p1_z
6	Surface Wind Direction	p1th
7	Surface divergence	p1zh
8	500 hPa airflow strength	p5_f
9	500 hPa zonal velocity	p5_u
10	500 hPa meridional velocity	p5_v
11	500 hPa vorticity	p5_z
12	500 hPa geopotential height	p500
13	500 hPa Wind Direction	p5th
14	500 hPa divergence	p5zh
15	850 hPa airflow strength	p8_f
16	850 hPa zonal velocity	p8_u
17	850 hPa meridional velocity	p8_v
18	850 hPa vorticity	p8_z
19	850 hPa geopotential height	p850
20	850 hPa Wind Direction	p8th
21	850 hPa divergence	p8zh
22	precipitation	prcp
23	Specific humidity at 500 hPa	s500
24	Specific humidity at 850 hPa	s850
25	Surface specific humidity	shum
26	Mean temperature at 2 m	temp

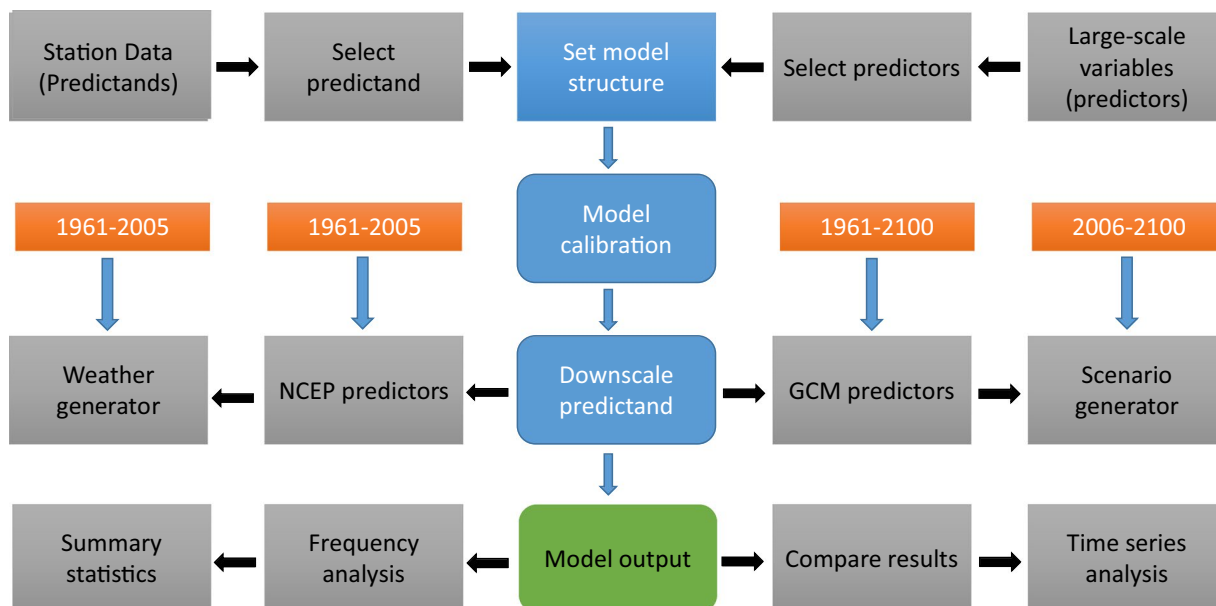
**Table 1.** List of the large-scale climate variables (predictors) used for downscaling.

**Downscaling process.** SDSM is used to downscale the output from a GCM by developing a statistical relationship between the local (predictands) and large-scale climate variables (predictors) using a multi-linear regressions model and stochastic bias correction techniques<sup>10,15</sup>. Observed daily precipitation and maximum and minimum temperature from 211 stations and 26 predictors (Table 1) from the NCEP (National Centers for Environmental Prediction) reanalysis data and CanESM2 (second generation Canadian Earth System Model) are used. Both the NCEP (1961–2005) and CanESM2 (1961–2100) predictors are available at a spatial resolution of about 2.81°, with nearly uniform longitude and latitude. In a single GCM box, 2–17 ground stations are available for the downscaling process (Fig. 1). Compared to GCMs, the NCEP predictors are commonly used due to their accuracy (e.g., high correlation and Nash–Sutcliffe Efficiency) in representing the current climate<sup>15,28</sup>. Therefore, the NCEP and CanESM2 predictors are used for model calibration and validation and future projection, respectively. The predictors derived from the CanESM2 are available under RCP2.6, RCP4.5, and RCP8.5 for downscaling of future climate projections (2006–2100).

After data quality control, predictors are selected for each predictand as shown in Fig. 2. Selection of predictors for a predictand (e.g., maximum temperature) is based on the correlation matrix, partial correlation, and P-value. The selected predictor is further assessed for its accuracy using graphical methods such as a scatterplot. In general, the predictors are analysed more than 5486 (211 stations \* 26 predictors) times for a single and 16458 (211\*26\*3) times for the three predictands (precipitation and maximum and minimum temperature) used in this study.

Using the selected predictors for each predictand, the model is calibrated under unconditional (temperature) and conditional (precipitation) processes on a monthly scale. For stations with a short length of observations, particularly for precipitation, the model is calibrated on seasonal and annual time scales to increase the number of wet days. The calibrated model, using the identified best performing predictors, produces up to 100 ensembles of daily time series and its output is the mean of the ensembles. The model output (ensemble mean) is used to assess the performance of SDSM in reproducing the observed data<sup>10</sup>. The performance is evaluated using a number of statistical parameters (generic and conditional tests) and graphical evaluation methods (e.g., bar plot) included in SDSM. In SDSM, stochastic techniques are included to improve the model performance in reproducing the observed data by artificially inflating the variance of the model output<sup>15</sup>. In addition, optimization techniques such as the ordinary least-square and dual simplex methods are provided in SDSM to control instabilities in regression coefficients<sup>10</sup>. As shown in Fig. 2, the calibrated model is used to generate future scenarios using the CanESM2 predictors available under RCP2.6, RCP4.5, and RCP8.5.

**Data Outputs.** The statistically downscaled daily precipitation (Pr), maximum temperature (Tmax) and minimum temperature (Tmin) for the current climate (1961–2005) and future scenarios under the RCPs (RCP2.6,



**Fig. 2** Schematic overview of the Statistical Down Scaling Model (SDSM). Modified from Wilby *et al.*<sup>19</sup>.

RCP4.5, and RCP8.5) are given in zipped boxes (Box\_1 to Box\_211). For instance in Box\_1, there are 15 text files (.OUT and .PAR) for the three variables as follows;

1. Daily precipitation (Pr)

- Pr.PAR, list of selected predictors for daily precipitation at location one.
- Pr-syn. OUT, model output of daily precipitation for the current period (1961–2005) generated using the weather generator.
- Pr-rcp26. OUT, projected daily precipitation under RCP2.6 (2006–2100).
- Pr-rcp45. OUT, projected daily precipitation under RCP4.5 (2006–2100).
- Pr-rcp85. OUT, projected daily precipitation under RCP8.5 (2006–2100).

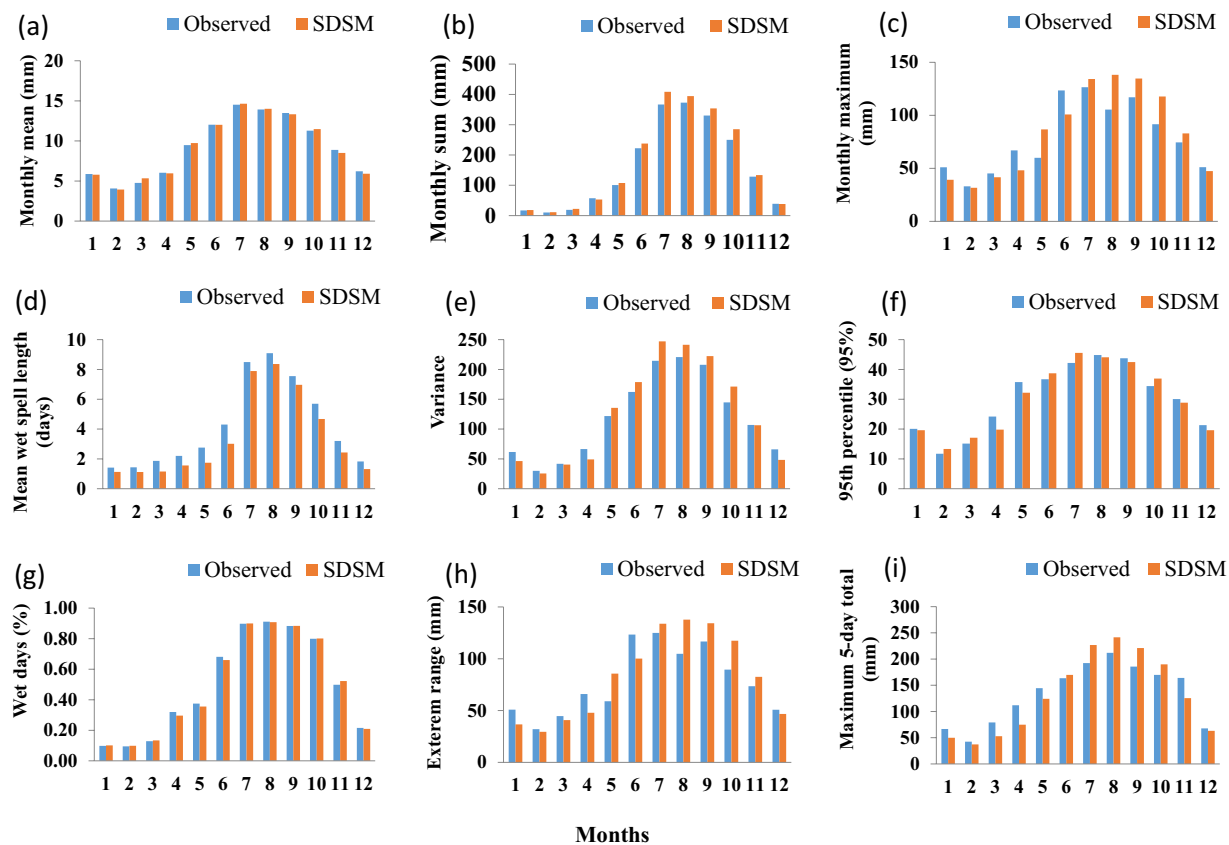
2. Daily maximum temperature (Tmax)

- Tmax.PAR, list of selected predictors for daily maximum temperature at location one.
- Tmax-syn. OUT, model output of daily maximum temperature for the current period (1961–2005) generated using the weather generator.
- Tmax-rcp26. OUT, projected daily maximum temperature under RCP2.6 (2006–2100).
- Tmax-rcp45. OUT, projected daily maximum temperature under RCP4.5 (2006–2100).
- Tmax-rcp85. OUT, projected daily maximum temperature under RCP8.5 (2006–2100).

3. Daily minimum Temperature (Tmin)

- Tmin.PAR, list of selected predictors for daily minimum temperature at location one.
- Tmin-syn. OUT, model output of daily minimum temperature for the current period (1961–2005) generated using the weather generator.
- Tmin-rcp26. OUT, projected daily minimum temperature under RCP2.6 (2006–2100).
- Tmin-rcp45. OUT, projected daily minimum temperature under RCP4.5 (2006–2100).
- Tmin-rcp85. OUT, projected daily minimum temperature under RCP8.5 (2006–2100).

In each file, for example, precipitation (Pr-syn. OUT) at Box\_1, the model output contains 20 ensembles for the current period. The 20 ensembles produced for each predictand show the uncertainty in the projection and this depends on the selected predictors and predictand and length and quality of observed data. The parameter files (.PAR) only provide the short names of the predictors as shown in Table 1. The inclusion of the predictors selected for each station in this dataset enables researchers to identify the large-scale climate variable linked with the local climate. As East Africa is one of the most topographically complex parts of Africa, the predictors vary considerably from location to location. In addition to the data Zip file, location information (latitude (lat) and longitude (lon)) is given as an excel file (Box\_location.csv) for each box.



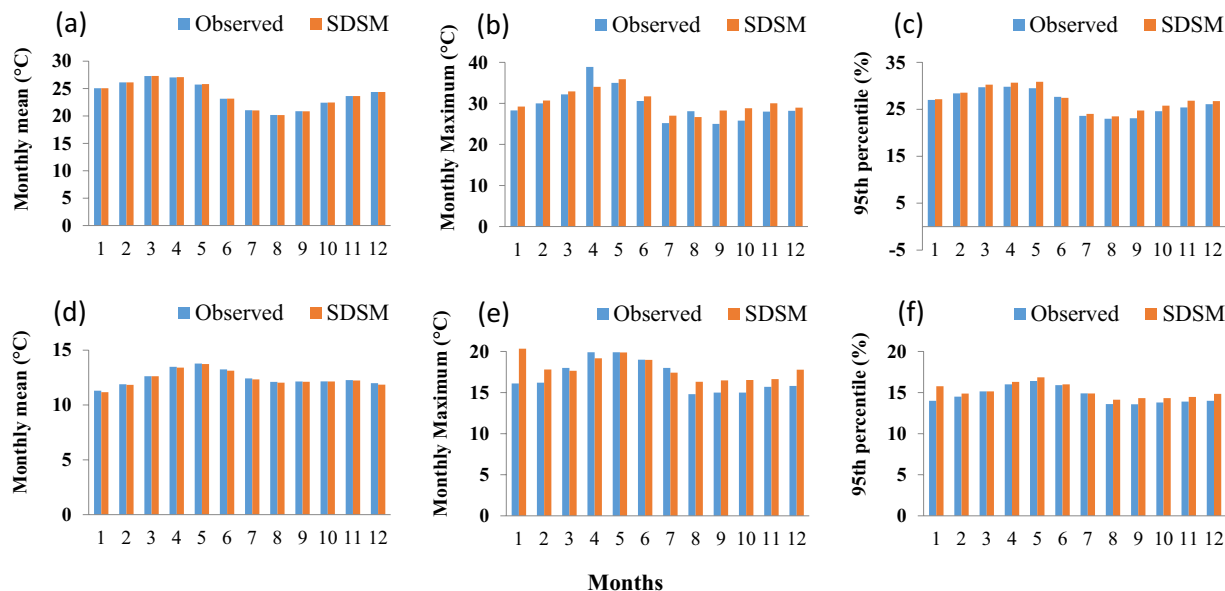
**Fig. 3** An example of the performance of SDSM compared to the observed monthly precipitation characteristics. (a) Mean. (b) Sum. (c) Monthly maximum. (d) Mean wet spell length. (e) Variance. (f) 95<sup>th</sup> Percentile. (g) Percentage of wet days. (h) Extreme range. (i) Maximum 5-day total precipitation.

Variable	Values	R <sup>2</sup>	RMSE	Pbias
Precipitation	Mean	0.997	0.25	0.1
	Maximum	0.90	18.15	6.1
	Variance	0.99	17.66	4.7
	Sum	0.998	18.99	7.8
	Percentage of wet	0.999	0.012	-0.7
	Wet spell	0.99	0.76	-17
	95 <sup>th</sup> Percentile	0.96	2.36	-0.5
Maximum temperature	Mean	0.99	0.03	0.01
	Maximum	0.82	2.26	1.5
	95 <sup>th</sup> Percentile	0.95	0.88	2.6
Minimum temperature	Mean	0.99	0.1	-0.5
	Maximum	0.67	1.66	5.7
	95 <sup>th</sup> Percentile	0.88	0.69	3.5

**Table 2.** An example of model performance for monthly average precipitation and temperature values at station Nekemet during 1961–2005.

### Data Records

For Ethiopia, Kenya, and Tanzania, the daily precipitation and maximum and minimum temperature dataset for the current (1961–2005) and future periods (2006–2100, under the RCPs) are available as a zipped file for download<sup>29</sup>. The zipped file contains 15 files for precipitation and maximum and minimum temperature as explained in the above section (data output). In order to make the data easier for reuse, the data is provided in a text format that can be easily read by different programming languages such as R and Python.



**Fig. 4** Performance of SDSM compared to the observed monthly temperature for station Nekemet. **(a)** Mean of maximum temperature. **(b)** Maximum of maximum temperature. **(c)** 95<sup>th</sup> Percentile of maximum temperature. **(d)** Mean of minimum temperature. **(e)** Maximum of minimum temperature. **(f)** 95<sup>th</sup> Percentile of minimum temperature.

### Technical Validation

Evaluation of the model output for both precipitation and maximum and minimum temperature is carried out using the observed data for each station. In SDSM, multiple model evaluation methods (statistical and graphical) methods are included to assess the performance of the calibrated model in reproducing the observed data. As explained in the above section, the performance of the model depends on the selected predictors for the predictand at a given location. Even though a predictor shows a good correlation and low P-value ( $<0.05$ ) during the screening process, this predictor might not really be the best in reproducing the observed data, which might be due to the presence of outliers. Therefore, a predictor has to be screened first using the correlation matrix, P-value, and scatterplots and the final output is evaluated using the model statistical (e.g., mean, variance and standard deviation) and graphical (bar and line plots) methods. In addition to the statistical parameters available in SDSM, additional methods such as the coefficient of determination ( $R^2$ ), Root Mean Square Error (RMSE), and Percent of bias (Pbias) are used<sup>30</sup> to identify the most accurate predictors. Both  $R^2$  (Eq. 1) and RMSE (Eq. 2) are indicators of goodness of fit, while Pbias (Eq. 3) shows the tendency of the observed data to be over- or underestimated by the model.

$$R^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - X_i)^2}{N}} \quad (2)$$

$$Pbias = \frac{\sum_{i=1}^N (Y_i - X_i)}{\sqrt{\sum_{i=1}^N X_i}} \times 100 \quad (3)$$

where  $X_i$  and  $\bar{X}$  and  $Y_i$  and  $\bar{Y}$  are the observed and model monthly and average data, respectively, of the  $i$ th event in  $N$  number of events.

The overall evaluation methods enabled us to accurately identify the best fit predictor for the 211 stations used in this study. An explanatory example for one station in Ethiopia (Nekemt; latitude = 9.08°N, longitude = 36.46°E) is provided in Fig. 3. Figure 3 shows the performance of SDSM in generating some of the station based precipitation characteristics such as the average monthly mean, sum, maximum, wet spell length, variance, 95<sup>th</sup> percentile, percentage of wet days, extreme range, and maximum 5-day precipitation. For precipitation at station Nekemt, the selected predictors are;

- Mean sea level pressure (mslp),
- Surface divergence (p1zh),
- 850 hPa vorticity (p8\_z), and

- Specific humidity at 850 hPa (s850). This shows that the day to day variabilities in mslp, p1zh, p8\_z, and s850 are useful predictors for precipitation occurrence at station Nakemet compared to the predictors provided in Table 1.

The results (Fig. 3 and Table 2) shows the accuracy of the model in reproducing the observed precipitation characteristics and shows a high  $R^2$  and lower biases and errors. Here, the ensembles mean is used to compare with the observed data. As shown in Table 2, the model shows high values of  $R^2$  ( $>0.96$ ) for the selected precipitation characteristics. Modeling precipitation is one of the most challenging climate variables due to the low predictability of by regional climate forcing<sup>15</sup> and in a topographically complex region.

In addition, the model shows high accuracy for maximum and minimum temperature (Fig. 4 and Table 2). Compared to the mean ( $R^2 > 0.99$ ), the maximum of maximum and minimum temperature are overestimated by 1.5% and 5.7%, respectively (Table 2). For maximum and minimum temperature the selected predictors are mean sea level pressure (mslp), Surface divergence (p1zh), and 850 hPa vorticity (P8\_z) and Surface meridional velocity (p1\_v), 500 hPa geopotential height (p500), Specific humidity at 850 hPa (s850), and Surface specific humidity (shum), respectively. In general, the same approach is used to assess the performance of SDSM and to identify the best performing predictors for all the stations used in this study. For the 211 stations, after quality control, the predictors are evaluated more than 16458 times.

Overall, considering the complexity of the variable, particularly modelling of precipitation, the presence of data gaps, and topography of the region, the results are promising and can be used to drive impact assessment and adaptation studies in this region. In addition, SDSM was also identified as an accurate model in infilling missing values in data-sparse regions such as in Africa and the Middle East<sup>25</sup>. Using the new version of SDSM (SDSM 5.2), the data can be also used to assess the vulnerability of location-based adaptation measures and develop climate change scenarios without the dependency of GCMs<sup>31</sup>.

### Code Availability

SDSM version 4.2, freely available (<https://sdsml.org.uk/software.html>), is used to statically downscale the projection from the second generation Canadian Earth System Model (CanESM2). The predictors derived from CanESM2 and the NCEP reanalysis data<sup>32</sup> are exported into SDSM directory for model calibration and projection. The CanESM2 is one of the GCMs used in the Coupled Model Inter-comparison Project Phase 5 (CMIP5). A free code written in R (mean-R.txt) is provided to compute the ensembles mean for a single predictand.

### References

1. Dixon, K. W. *et al.* Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Clim. Change* **135**, 395–408 (2016).
2. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) WG1, 1007 (Cambridge Univ. Press, 2007).
3. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2013: The Physical Science Basis* (eds Stocker *et al.*) WG1, 1535 (Cambridge Univ. Press, 2013).
4. Gutmann, E. D. *et al.* A Comparison of Statistical and Dynamical Downscaling of Winter Precipitation over Complex Terrain. *J. Clim.* **25**, 262–281 (2012).
5. Meenu, R., Rehana, S. & Mujumdar, P. P. Assessment of hydrologic impacts of climate change in Tunga–Bhadra river basin, India with HEC-HMS and SDSM. *Hydrol. Process.* **27**, 1572–1589 (2013).
6. Lutz, A. F. *et al.* Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *Int. J. Climatol.* **36**, 3988–4005 (2016).
7. Joetzier, E., Douville, H., Delire, C. & Ciais, P. Present-day and future Amazonian precipitation in global climate models: CMIP5 versus CMIP3. *Clim. Dyn.* **41**, 2921–2936 (2013).
8. Knutti, R. & Sedláček, J. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change* **3**, 369–373 (2013).
9. Gebrechorkos, S. H., Hülsmann, S. & Bernhofer, C. Regional climate projections for impact assessment studies in East Africa. *Environ. Res. Lett.* (2019).
10. Tavakol-Davani, H., Nasser, M. & Zahraie, B. Improved statistical downscaling of daily precipitation using SDSM platform and data-mining methods. *Int. J. Climatol.* **33**, 2561–2578 (2012).
11. Endris, H. S. *et al.* Assessment of the Performance of CORDEX Regional Climate Models in Simulating East African Rainfall. *J. Clim.* **26**, 8453–8475 (2013).
12. Dosio, A. & Panitz, H.-J. Climate change projections for CORDEX-Africa with COSMO-CLM regional climate model and differences with the driving global climate models. *Clim. Dyn.* **46**, 1599–1625 (2016).
13. Brown, C., Greene, A. M., Block, P. J. & Giannini, A. *Review of Downscaling Methodologies for Africa Climate Applications*. (International Research Institute for climate and Society, 2008).
14. Hamlet, A., Salathe, E. & Carrasco, P. Statistical Downscaling Techniques for Global Climate Model Simulations of Temperature and Precipitation with Application to Water Resources Planning Studies. Chapter 4. In *Final Report for the Columbia Basin Climate Change Scenarios Project* (University of Washington, Seattle, 2010).
15. Wilby, R. L. & Dawson, C. W. SDSM 4.2-A Decision Support Tool for the Assessment of Regional Climate Change Impacts *Version 4.2 User Manual*. 1–94 (Lancaster University, 2007).
16. Khan, M. S. & Coulibaly, P. Assessing Hydrologic Impact of Climate Change with Uncertainty Estimates: Bayesian Neural Network Approach. *J. Hydrometeorol.* **11**, 482–495 (2009).
17. Semenov, M. A. & Barrow, E. M. Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change* **35**, 397–414 (1997).
18. Anandhi, A. *et al.* Examination of change factor methodologies for climate change impact assessment. *Water Resour. Res.* **47** (2011).
19. Wilby, R. L., Dawson, C. W. & Barrow, E. M. sdsml — a decision support tool for the assessment of regional climate change impacts. *Environ. Model. Softw.* **17**, 145–157 (2002).
20. Hashmi, M. Z., Shamseldin, A. Y. & Melville, B. W. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. *Stoch. Environ. Res. Risk Assess.* **25**, 475–484 (2011).
21. Hassan, Z., Shamsudin, S. & Harun, S. Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature. *Theor. Appl. Climatol.* **116**, 243–257 (2013).

22. Niang, I. *et al.* *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects* (eds Barros, V. R. *et al.*) WG2, 1199–1265 (Cambridge Univ. Press, 2014).
23. Gebrechorkos, S. H., Hülsmann, S. & Bernhofer, C. Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. *Int. J. Climatol.* **39**, 18–30 (2018).
24. Gebrechorkos, S. H., Hülsmann, S. & Bernhofer, C. Evaluation of multiple climate data sources for managing environmental resources in East Africa. *Hydrol. Earth Syst. Sci.* **22**, 4547–4564 (2018).
25. Wilby, R. L. & Dawson, C. W. The Statistical DownScaling Model: insights from one decade of application. *Int. J. Climatol.* **33**, 1707–1719 (2013).
26. Funk, C. *et al.* The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* **2**, 150066 (2015).
27. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *J. Clim.* **19**, 3088–3111 (2006).
28. Sachindra, D. A., Huang, F., Barton, A. & Perera, B. J. C. Statistical downscaling of general circulation model outputs to precipitation—part 2: bias-correction and future projections. *Int. J. Climatol.* **34**, 3282–3303 (2014).
29. Gebrechorkos, S. H., Hülsmann, S. & Bernhofer, C. Statistically downscaled climate dataset for East Africa. *figshare*. <https://doi.org/10.6084/m9.figshare.c.4282226> (2019).
30. Legates, D. R. & McCabe, G. J. Evaluating the use of ‘goodness-of-fit’ Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241 (1999).
31. Wilby, R. L., Dawson, C. W., Murphy, C., O’Connor, P. & Hawkins, E. The Statistical DownScaling Model - Decision Centric (SDSM-DC): conceptual basis and applications. *Clim. Res.* **61**, 259–276 (2014).
32. Kalnay, E. *et al.* The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).

## Acknowledgements

We would like to thank the National Meteorological Agency (NMA) of Ethiopia for providing adequate observed station data.

## Author Contributions

Solomon H. Gebrechorkos carried out the downscaling process and wrote the manuscript with substantial contribution from Stephan Hülsmann and Christian Bernhofer.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019