Data in Brief 54 (2024) 110455

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Daily bias-corrected weather data and daily simulated growth data of maize, millet, sorghum, and wheat in the changing climate of sub-Saharan Africa

Seyyedmajid Alimagham[a], Marloes P. van Loon[a], Julian Ramirez-Villegas[a,b], Herman N. C Berghuijs[a,c], Martin K. van Ittersum[a,d,*]

[a] *Plant Production Systems Group, Wageningen University & Research, P.O. Box 430, 6700AK Wageningen, the Netherlands*
[b] *Bioversity International, Via di San Domenico 1, Rome, Italy*
[c] *Wageningen Environmental Research, Wageningen University & Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands*
[d] *Department of Crop Production Ecology, Swedish University of Agricultural Sciences (SLU), P.O. Box 7043, Uppsala 75007, Sweden*

A R T I C L E I N F O

Article history:
Received 29 February 2024
Revised 16 April 2024
Accepted 17 April 2024
Available online 23 April 2024

Dataset link: Daily bias-corrected weather data and daily simulated growth data of maize, millet, sorghum, and wheat in the changing climate of sub-Saharan Africa (Original data)

Keywords:
Africa
Bias
Climate change
Cereal

A B S T R A C T

Crop models are the primary means by which agricultural scientists assess climate change impacts on crop production. Site-based and high-quality weather and climate data is essential for agronomically and physiologically sound crop simulations under historical and future climate scenarios. Here, we describe a bias-corrected dataset of daily agro-meteorological data for 109 reference weather stations distributed across key production areas of maize, millet, sorghum, and wheat in ten sub-Saharan African countries. The dataset leverages extensive ground observations from the Global Yield Gap Atlas (GYGA), an existing climate change projections dataset from the Inter-Sectoral Model Intercomparison Project (ISIMIP), and a calibrated crop simulation model (the WOrld FOod Studies –WOFOST). The weather data were bias-corrected using the delta method, which is widely

* Corresponding author at: Plant Production Systems Group, Wageningen University & Research, P.O. Box 430, 6700AK Wageningen, the Netherlands.
E-mail address: martin.vanittersum@wur.nl (M.K. van Ittersum).

https://doi.org/10.1016/j.dib.2024.110455

used in climate change impact studies. The bias-corrected dataset encompasses daily values of maximum and minimum temperature, precipitation rate, and global radiation obtained from five models participating in the Sixth Phase of the Coupled Model Intercomparison Project (CMIP6), as well as simulated daily growth variables for the four crops. The data covers three periods: historical (1995–2014), 2030 (2020–2039), and 2050 (2040–2059). The simulation of daily growth dynamics for maize, millet, sorghum, and wheat growth was performed using the daily weather data and the WOFOST crop model, under potential and water-limited potential conditions. The crop simulation outputs were evaluated using national agronomic expertise. The presented datasets, including the weather dataset and daily simulated crop growth outputs, hold substantial potential for further use in the investigation of future climate change impacts in sub-Saharan Africa. The daily weather data can be used as an input into other modelling frameworks for crops or other sectors (e.g., hydrology). The weather and crop growth data can provide key insights about agro-meteorological conditions and water-limited crop output to inform adaptation priorities and benchmark (gridded) crop simulations. Finally, the weather and simulated growth data can also be used for training machine learning techniques for extrapolation purposes.

Specifications Table

Subject	Global and Planetary Change
Specific subject area	Climate change; Agro-meteorology; Crop modeling.
Type of data	Table, Processed, Simulated
Data collection	Comparing the ISIMIP data with the measured weather data from the Global Yield Gap Atlas (GYGA; www.yieldgap.org) revealed systematic bias, with substantial effect on crop simulations. We used the delta method to correct the bias in the ISIMIP weather data. Reference weather station (RWS) data was gathered from the GYGA. Daily simulated crop growth data was generated using the bias-corrected weather data and the WOFOST crop model. GYGA provided other required inputs for running the WOFOST crop model.
Data source location	Plant Production Systems Group, Wageningen University & Research
Data accessibility	Repository name: Mendeley Data DOI: 10.17632/7s4frszjmz.2 https://data.mendeley.com/datasets/7s4frszjmz/2
Related research article	Climate change impact and adaptation of rainfed cereal crops in sub-Saharan Africa https://authors.elsevier.com/sd/article/S1161-0301(24)00058-3

1. Value of the Data

- The dataset harbors substantial potential for use by the broader crop modelling community because (1) it covers all key weather inputs for crop simulation in ten countries and in areas of importance for agricultural production; (2) it is based on ground-truth data collected over the course of nearly a decade by the GYGA project team; and (3) it extends the weather data with crop state variables derived from crop simulations with WOFOST (WOrld FOod STudies);

a well-established model for four cereal crops (maize, millet, sorghum, wheat) that are of importance for food security.

- This dataset includes bias-corrected daily values of maximum and minimum temperatures, precipitation rate, and global radiation obtained from five Global Circulation Models (GCMs) under two shared socio-economic pathway (SSP) scenarios, SSP3-7.0 and SSP5-8.5, for three time periods: historical (1995–2014), 2030 (2020–2039), and 2050 (2040–2059). The utilization of data from five GCMs and two SSP scenarios facilitates the implementation of multi-model ensemble analysis, which is widely recognized as crucial in climate projection and climate risk analysis [2].
- Under potential and water-limited conditions, the daily simulated variables include phenological stage, dry matter, root depth, leaf area index, and evapotranspiration (Table 2). These simulations allow exploring a wide range of plausible crop productivity futures, providing insights on growth dynamics under future climate scenarios. Both the weather data and the daily crop growth data can serve three key purposes: (1) the analysis of climate change impacts on agriculture in SSA; (2) simulation and benchmarking of crop growth responses with other crop models; and (3) training of machine learning meta-models for spatiotemporal extrapolation purposes. Our dataset is the first site-specific weather and crop growth dataset with multi-crop and multi-country coverage that focuses on the CMIP6 climate model ensemble in SSA.

2. Background

Gridded climate scenario data for climate change impact assessments in agriculture is available from a variety of data sources. The Inter-Sectoral Model Intercomparison Project (ISIMIP) dataset was used as a source of daily weather for crop simulation in several existing studies e.g., [3,4]. Systematic bias in the ISIMIP dataset, however, precludes robust crop simulations at the site-specific scale. Thus, secondary bias correction of the ISIMIP dataset is necessary [5,6]. However, such process is challenging in regions such as sub-Saharan Africa (SSA) due to limited measured weather data. We leveraged the extensive Global Yield Gap Atlas (GYGA) network of collaborations and data gathering efforts to produce a site-specific and second bias-corrected version of ISIMIP for 109 sites of relevance to maize (105 sites), millet (69 sites), sorghum (72 sites), and wheat (16 sites) production in ten countries in SSA. Moreover, the inputs for WOFOST were gathered from a comprehensive management and soil dataset from the GYGA project, which were verified for each station by local experts. The GYGA dataset specifies cultivars, planting calendars, and soil hydrological properties, which allowed for realistic simulations of potential and water-limited potential yield across the 10 sub-Saharan African (SSA) countries.

The focus of the published paper is on presenting results obtained under water-limited condition. In addition to the simulated results under water-limited conditions, this dataset also includes daily simulations for potential conditions. Furthermore, the dataset encompasses both the weather data and simulated outcomes for the time horizon of 2030, which are not present in the published paper.

3. Data Description

The dataset presented herein is unique in that it has multi-country, multi-location, multi-crop, and multi-scenario (GCM, SSP, period) coverage. It has been produced by leveraging a publicly available dataset (the ISIMIP dataset) and several high-quality local datasets from the Global Yield Gap Atlas. We foresee three key uses of the data:

(1) The analysis of climate change impacts on agriculture in SSA. The data presented covers multiple climate and agricultural futures and therefore can be used to derive key insights on agro-meteorological conditions for maize, millet, sorghum, and wheat across

multiple countries in SSA. Importantly, the daily resolution of the dataset allows technical and scientific enquiry of agronomic performance in tandem with meteorological and eco-physiological analysis.

(2) Simulation and benchmarking of crop growth responses with other crop models. Our simulations use a well-established and calibrated crop model (WOFOST), and compare well with field observations of water-limited yield potential [1]. The weather input datasets we provide can be used as input for other crop modeling frameworks, whereas the simulated crop growth can be used to benchmark simulations from other models. Such kind of benchmarking can help reduce uncertainties and establish crop-climate model ensembles (e.g., [2,7,8]).

(3) Training of machine learning meta-models for spatiotemporal extrapolation purposes. The combination of weather and crop status variables presented here allows building machine learning models that capture key growth dynamics and therefore overcome the extrapolation challenges typical in empirical models.

The dataset described in this article includes daily weather data and simulated crop growth data for 109 sites across ten countries in SSA. Whereas the weather data covers all sites, the simulated growth data for the four crops (maize, millet, sorghum, wheat) covers subsets of the sites depending on the geographic distribution of cultivation areas for the given crop (Fig. 1). Table 1 offers a full list of dataset specifications. Table 2 presents additional information regarding the data stored in the repository.

3.1. Geographic coverage of the dataset

Fig. 1 illustrates the spatial distribution of weather stations for each crop in the ten countries. The stations can represent 65 % of the harvested area for maize, 90 % for millet, 83 % for sorghum, and 59 % for wheat; For the four crops together 72 % of the total harvested area in those countries can be represented (Table 3). When determining coverage areas, it was assumed that a station with a specific climate can serve as representative for the climate zone in which the weather station is located.

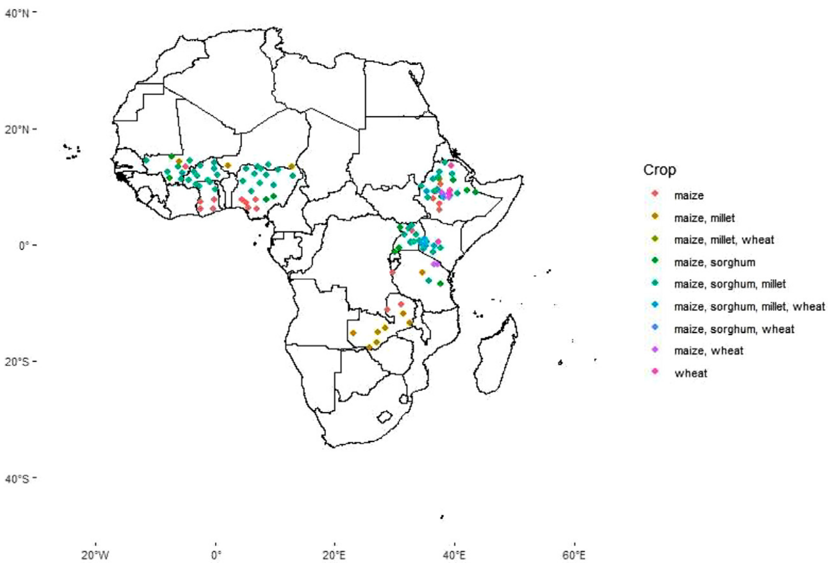

Fig. 1. The spatial distribution of weather stations for each crop.

Table 1
General technical specifications of the dataset.

Characteristic	Description
Continent	Africa
Countries covered	Burkina Faso, Ethiopia, Ghana, Kenya, Mali, Niger, Nigeria, Tanzania, Uganda, Zambia
Number of sites	109
Crops covered	Maize (n = 105 sites), millet (n = 69 sites), sorghum (n = 72 sites), wheat (n = 16)
Periods	Historical (1995–2014), 2030 (2020–2039), 2050 (2040–2059)
Shared Socioeconomic Pathways (SSPs)	SSP3-7.0, SSP5-8.5
General Circulation Models (GCMs)	GFDL-ESM4 (GFDL), IPSL-CM36A-LR (IPSL), MPI-ESM1-2-HR (MPI), MRI-ESM2-0 (MRI), and UKESM1-0-LL (UKESM)
Climate variables	Precipitation, maximum temperature, minimum temperature, downwards shortwave solar radiation
Crop model	WOrld FOod STudies (WOFOST)
Growth variables	Development stage, total above ground dry matter, dry matter for grain, root depth, leaf area index, and evapotranspiration (see also Table 2).
Temporal resolution	Daily

Table 2
The data information stored in the repository.

Data group	Variable	Time step	unite	Abbreviation
Weather	Minimum temperature	daily	°C	TMIN
	Maximum temperature	daily	°C	TMAX
	Precipitation	daily	mm	RAIN
	Radiation	daily	kJ	IRRAD
	Wind*	daily	m/s	WIND
	Vapor pressure**	daily	kPa	VPA
Agronomic (potential and water-limited potential conditions)	Development stage	daily	–	DVS***
	Total above ground dry matter	daily	kg/ha	TAGP
	Dry matter for leaf	daily	kg/ha	TWLV
	Dry matter for stem	daily	kg/ha	TWST
	Dry matter for root	daily	kg/ha	TWRT
	Dry matter for grain	daily	kg/ha	TWSO
	Root depth	daily	cm	RD
	Leaf area index	daily	–	LAI
	Evapotranspiration	daily	cm	ET
	Transportation	daily	cm	TRA
	Evaporation	daily	cm	EVS
	Water stress index	daily	–	RFWS****

* The wind data in this study are sourced from ISIMIP and have not undergone bias correction.
** The methodology employed by [9] was used to compute daily vapor pressure data using bias-corrected maximum and minimum temperatures.
*** For further details, refer to [10].
**** It ranges from 0 to 1, with 1 indicating no water stress and 0 indicating no available water for crops.

3.2. Data table structure

Besides the variables in Table 2, there are extra columns in the data table explained in Table 4.

Table 3
Number of reference weather stations for each crop and the percentage of national crop area coverage by the climate zones that are represented by the weather stations in each country.

Country	Number of sites for doing simulations				The harvested area coverage by selected sites for each crop (%)**			
	maize	millet	sorghum	wheat	maize	millet	sorghum	wheat
Burkina Faso	8	8	8	0	87	99	98	–
Ethiopia	26	13	19	9	71	79	60	62
Ghana	6	2	2	0	90	93	94	–
Kenya	9	8	8	6	50	46	61	24
Mali	9	8	8	0	92	97	94	–
Niger	4	4	2	**0***	60	93	97	**97***
Nigeria	17	10	12	**0***	68	81	79	**89***
Tanzania	6	2	3	1	34	59	54	29
Uganda	12	8	10	**0***	86	88	67	**54***
Zambia	8	6	**0***	**0***	94	88	**62***	**94***
Total	105	69	72	16	65	90	83	59

* While we did not simulate crop growth for these specific crop and country combinations, it is possible to cover the harvested area of these crops with selected stations designated for other crops within the country (see **).

** It was assumed that the climate in the buffer zone of the stations can serve as representative for the climate zone in which the weather station is located.

Table 4
Data tables' columns.

Data group	variable	Description
Weather	country_name	The name of the country
	GCM	The name of the GCM
	Station_id	An ID for a given station. This ID serves as a connector between different files.
	lon	Longitude of the station
	lat	Latitude of the station
	DAY	Date (YYYYMMDD)
	SNOWDEPTH	The depth of snow which is 0
Agronomic (potential and water-limited potential conditions)	run-id	An ID for each combination (country*station*soil*cycle*crop*time horizon*SSP*GCM*year)
	country	The name of the country
	Station_id	An ID for the given station. This ID serves as a connector between different files
	lon	Longitude of the station
	lat	Latitude of the station
	soil	A code for the soil type. The information of each soil type is presented in excel file named “soil data” in the repository.
	cycle	Mono cropping (1) or double cropping (1 and 2) at a given station
	crop	The name of the crop
	SSP	Future climate change scenario. If it is HIS, it means it is for the historical condition.
	GCM	The name of the GCM.
	Irrigation_status	It shows if the simulation is for potential (Irrigated) or water-limited (Rainfed) conditions

4. Materials and methods

4.1. Weather station selection

The data of 105 weather stations for maize, 69 for millet, 72 for sorghum and 16 for wheat representing the key production areas of these rainfed crops (Table 3) were extracted from the GYGA dataset for the ten countries in SSA (https://www.yieldgap.org). The GYGA protocol selects the key climate zones (designated climate zones, DCZ) and crop combination for each country based on harvested area and information from local agronomists. Within those DCZs, local weather stations are identified. Next, a 100-km radius buffer area surrounding each weather station is created and clipped by the borders of the DCZ and the country to ensure that the buffer zone is within a unique climate zone and country combination [11].

4.2. Reference weather station data

Ground-based weather station (also referred to as reference weather station, RWS) data were extracted from the GYGA dataset. GYGA gathered these datasets from local partners, including National Meteorological Agencies and National Agricultural Research Organizations. The data included daily maximum and minimum temperatures, precipitation, and solar radiation for the period 2000–2019 (https://www.yieldgap.org). When there is no long-term weather data for a given RWS, the GYGA project uses a method to create long-term weather series based on a few years of observed daily temperature. The approach is referred to as data propagation (see [12]). The propagated data are comprised of uncorrected gridded solar radiation from the Prediction of Worldwide Energy Resource dataset from the National Aeronautics and Space Administration (NASA–POWER), rainfall from the Tropical Rainfall Measuring Mission (TRMM) dataset, and site-specific calibration of NASA–POWER maximum and minimum temperatures using a limited amount of observed daily temperature data. For further details, the reader is referred to [12].

4.3. CMIP6 weather data

The weather data of the historical, 2030, and 2050 time periods of the Coordinated Modelling Intercomparison Project-Phase 6 (CMIP6) were gathered from the ISIMIP project (https://www.isimip.org). ISIMIP provides global gridded daily weather data for our variables of interest (precipitation, maximum and minimum temperature, solar radiation, wind). We downloaded ISIMIP data from the ISIMIP data repository.[1] ISIMIP provides data for five different GCMs, including GFDL-ESM4 (GFDL), IPSL-CM36A-LR (IPSL), MPI-ESM1-2-HR (MPI), MRI-ESM2-0 (MRI), and UKESM1-0-LL (UKESM). We gathered data for the five GCMs and two shared socioeconomic pathways (SSPs), namely, SSP3-7.0 and SSP5-8.5. The ISIMIP project provides these datasets with a daily temporal resolution, and a regular $0.5^\circ \times 0.5^\circ$ global grid spacing ($\sim$50 $\times$ 50 km pixel size at the equator). The data are produced by primary bias correcting raw CMIP6 data using quantile mapping approach and a bias-adjustment method, with the observational W5E5 v.1.0 dataset as a reference [13]. For each reference weather station (Fig. 1; Table 3), the ISIMIP climate data were extracted from the grid cells corresponding to the locations of the reference weather stations. Hereafter, we use the term ’raw ISIMIP data‘ to describe this dataset.

[1] https://data.isimip.org/

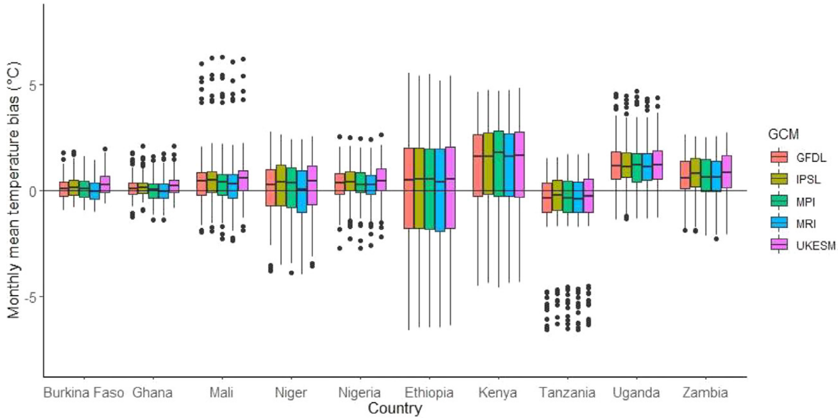

Fig. 2. The monthly mean temperature delta value (the ISIMIP historical data minus the measured data) in all reference weather stations for maize, millet, sorghum, and wheat areas in SSA using five different GCMs (different boxplot colors). The boxes contain monthly bias values for all stations situated within the country under historical conditions. The box covers 25–75 % of the data, the thick horizontal line corresponds to the median, and the whiskers extend to 5–95 % of the data points.

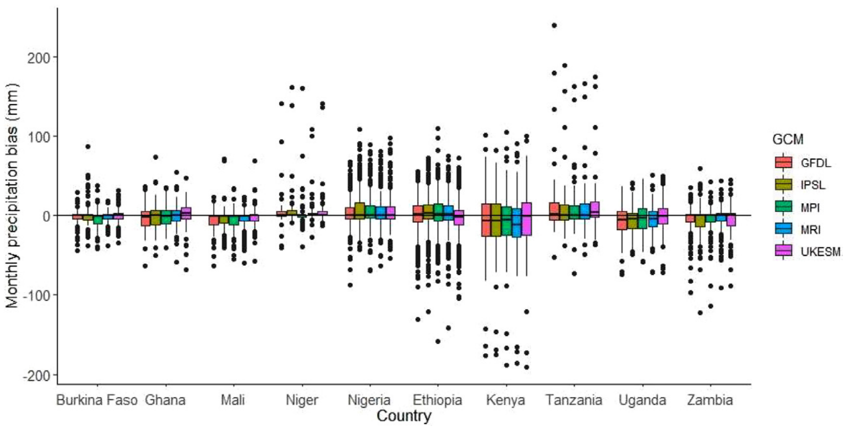

Fig. 3. The monthly precipitation bias value (the ISIMIP historical data minus the measured data) in all reference weather stations for maize, millet, sorghum, and wheat in SSA using five different GCMs (different boxplot colors). The boxes contain monthly bias values for all stations situated within the country under historical conditions. The box covers 25–75 % of the data, the thick horizontal line corresponds to the median, and the whiskers extend to 5–95 % of the data points.

4.4. Bias in the historical data

The monthly average of historical weather data from the raw ISIMIP data was compared with the monthly average of measured weather data to examine the systematic bias in the 'raw ISIMIP data' (Eq. (1) and (2); Fig. 2). Additionally, a comparison was conducted between the simulated crop growth results based on the measured data and the historical raw ISIMIP data. A substantial bias was observed in the raw ISIMIP data for some RWSs in some countries (Figs. 2 and 3). The size and effect of this bias in the crop model simulations suggests additional bias correction is needed.

$$Vbias_{m.g} = V_{isimip(m.g)} \quad - V_{mes(m)} \tag{1}$$

Vbias $_{m.g}$: the monthly bias in the raw ISIMIP temperature data for GCM g in month m
V $_{mes(m)}$: the monthly mean, maximum, or minimum temperature of the measured data (mes) in month m
V $_{isimip(m.g)}$: the monthly mean, maximum, or minimum temperature of the raw ISIMIP data for GCM g in month m

$$Xbias_{m.g} = \frac{X_{isimip(m,\ g)} - X_{mes(m)}}{X_{mes(m)}} * 100 \tag{2}$$

Xbias $_{m.g}$: the monthly bias in the raw ISIMIP precipitation or radiation data for GCM g in month m
X $_{mes(m)}$: the monthly mean precipitation or radiation of the measured data (mes) in month m
X $_{isimip\ (m.g)}$: the monthly mean precipitation or radiation of the raw ISIMIP data for GCM g in month m

4.5. Secondary bias correction of the historical data

We used the delta method as the form of bias correction of weather data, which is the simplest form of error removal for climate change simulation data [14,15]. In our case, it is used as the secondary bias correction method to correct the systematic bias of 'the raw ISIMIP data' [5,6] identified (see Section 4.4). This bias correction of the raw ISIMIP data was labelled as the secondary bias correction [6] because it was applied after the primary bias correction, i.e., the quantile mapping approach of the ISIMIP climate data. We assume that this secondary bias correction addresses remaining errors in the ISIMIP dataset not addressed by the original correction, while also addressing scale differences between site (GYGA weather data) and grid cell (raw ISIMIP data). It is worth noting that the methodology employed by [9] was used to compute daily vapor pressure data, employing second bias corrected maximum and minimum temperatures. Due to a lack of measured wind data, the ISIMIP wind data was not corrected.

To perform secondary bias correction, the mean bias of the raw ISIMIP data was calculated for each GCM and for each month in the period from 1995 to 2014. We used the absolute differences for the minimum and maximum temperatures (Eq. (3)) and the relative differences for precipitation and radiation (Eq. (4)). The use of relative changes for precipitation and radiation avoids arriving at negative values when applying delta values onto the historical raw ISIMIP weather data [14].

$$V_{isimip\_sbc(m,\ g)} = V_{isimip(m,\ g)} + (\bar{v}_{mes(m)} - \bar{v}_{isimip(m,\ g)}) \tag{3}$$

$$X_{isimip\_sbc(m,\ g)} = X_{isimip(m,\ g)} * \left(\frac{\bar{x}_{mes(m)}}{\bar{x}_{isimip(m,\ g)}}\right) \tag{4}$$

V: daily minimum or maximum temperature for GCM g in month m
isimip_sbc: ISIMIP secondary bias-corrected data,
isimip: the raw ISIMIP data,
mes: the measured data,
$\bar{v}$: the 20-year mean of maximum or minimum temperature for GCM g in month m
X: daily radiation or precipitation for GCM g in month m
$\bar{x}$: the 20-year mean of radiation or precipitation for GCM g and in month m

4.6. Secondary bias correction of the future weather data

The delta method was also used to correct the bias of the future weather data [14]. According to this method, the anomaly (delta change) for future periods was computed by comparing them to the historical climate of the same GCM for each variable and month.

$$\Delta V_{(m,\ g)} = V_{isimipF(m,\ g)} - V_{isimipH(m,\ g)} \tag{5}$$

$$\Delta X_{(m,\ g)} = \frac{X_{isimipF(m,\ g)} - X_{isimipH(m,\ g)}}{X_{isimipH(m,\ g)}} \tag{6}$$

ΔV: the delta change of minimum or maximum temperature for GCM g in month m
V: the 20-year mean of the minimum or maximum temperature for GCM g in month m
isimipH: the raw ISIMIP data for the historical conditions
isimipF: the raw ISIMIP data for the future periods (2030 or 2050)
ΔX: the delta change of precipitation or radiation for GCM g in month m
X: the 20-year mean of the precipitation or radiation for GCM g in month m

It is worth noting that the delta value in Eq. (6) was infinite for the months without precipitation in SSA during the dry season. To avoid this, we set a threshold of 0.001 mm per month for both the current and future GCM values, which prevents indetermination in Eq. (6).

The variables calculated by using Eqs. (3)–(6) were used to generate the second bias-corrected data for the future conditions:

$$V_{F(m,\ g)} = V_{isimip\_sbc(m,\ g)} + \ \Delta V_{(m,\ g)} \tag{7}$$

$$X_{F(m,\ g)} = X_{isimip\_sbc(m,\ g)} * \left(1 + \Delta X_{(m,\ g)}\right) \tag{8}$$

V_F: the daily second bias-corrected minimum or maximum temperature for GCM g in month m
X_F: the daily second bias-corrected precipitation or radiation for GCM g in month m

4.7. Daily simulated crop growth data

The daily growth dataset for the four crops was produced using the Python Crop Simulation Environment version of the crop model WOFOST (https://pcse.readthedocs.io/en/stable/). This crop model takes into account phenological development, leaf development, light interception, CO_2 assimilation, root growth, transpiration, respiration and partitioning of assimilates [16]. Daily weather data, crop cultivar parameters, soil parameters, and management data are needed to run the model. All these data except weather data (see previous sections for origin) were extracted from the GYGA dataset for each crop and weather station (https://www.yieldgap.org). The cultivar, soil, and management data in GYGA have been collected and evaluated through an extensive network of agronomists and crop modelers that collaborate with the GYGA project (https://www.yieldgap.org). The cultivar is specified using thermal times derived from locally available planting, flowering, and maturity times. Soil data were obtained from the Africa Soil Information Service (AfSIS; [17]). Management inputs (crop calendars, rainfed or irrigated) were gathered in each country by expert agronomists. The WOFOST model was executed using the daily second-bias corrected weather data for both historical and future conditions. We recorded daily simulation output for development stage, dry matter, root depth, leaf area index, evapotranspiration (Table 2).

It is worth noting that the input data for the crop model in the GYGA project were station-specific and have been verified by local experts for each station. The WOFOST model has been used in the GYGA project to simulate water-limited potential yields, and phenology related results have been evaluated by the agronomists and experts of the ten countries [18,19]. In addition, the crop model was evaluated against the highest yielding treatments from rainfed field experiments conducted in SSA under diverse climatic conditions, sourced from published articles which showed a robust simulation of the water-limited potential yield of the four crops in SSA [1]. Due to the extensive nature of our dataset in terms of climate and agronomic variables, countries, and within-country location (see Table 3), and the known quality of the simulated output, we propose that it is a useful benchmark for other simulation studies seeking to understand climate change impacts in SSA.

Limitations

The data is valid and limited to the buffer zones around the identified reference stations in the ten countries for target crops. Significant crop area changes require addition of new reference weather stations and measured weather data. It is important to acknowledge that the availability of weather data collected from ground-based sources in SSA is restricted due to institutional restrictions from the originating national partners (mostly National Meteorological Agencies). However, the provision of the ISIMIP dataset with secondary bias correction circumvents these restrictions by providing an improved version of the publicly available ISIMIP dataset. The delta method relies on two assumptions: (i) bias in future data is the same as in historical data, and (ii) within each month, future weather data exhibit the same variability as historical weather data. Therefore, the monthly distribution shape remains unchanged, while the delta modifies the values. Additionally, this approach is unsuitable for extreme events [20].

Ethics Statement

The proposed data does not involve any human subjects, animal experiments, or data collected from social media platforms. To the best of the knowledge of the authors, the data presented or methods used do not carry any racial, gender, or socioeconomic bias.

Data Availability

Daily bias-corrected weather data and daily simulated growth data of maize, millet, sorghum, and wheat in the changing climate of sub-Saharan Africa (Original data) (Mendeley Data).

CRediT Author Statement

Seyyedmajid Alimagham: Conceptualization, Methodology, Formal analysis, Writing – original draft, Data curation; **Marloes P. van Loon:** Methodology, Writing – review & editing, Data curation; **Julian Ramirez-Villegas:** Conceptualization, Writing – original draft; **Herman N. C Berghuijs:** Methodology, Writing – review & editing; **Martin K. van Ittersum:** Supervision, Conceptualization, Writing – original draft.

Acknowledgments

M.K.v.I and M.v.L. acknowledge the financial contribution of the NWO-WOTRO Strategic Partnership NL-CGIAR and support from the Excellence in Agronomy (EiA) 2030 CGIAR Initiative. JR-V acknowledges support from the Excellence in Agronomy (EiA) 2030 initiative, the Climate Resilience (ClimBeR) initiative of the CGIAR, and the Agriculture Adaptation Atlas project (https://adaptationatlas.cgiar.org).

Declaration of Competing Interest

The authors declare that they have no known competing financial or personal interests that could have influenced the work presented.

References

[1] S.M. Alimagham, M.P. van Loon, J. Ramirez-Villegas, S. Adjei-Nsiah, F. Baijukya, A. Bala, R. Chikowo, J.V. Silva, A.M. Soulé, G. Taulya, F.A. Tenorio, K. Tesfaye, M.K. van Ittersum, Climate change impact and adaptation of rainfed cereal crops in sub-Saharan Africa, Eur. J. Agron. (2024), doi:10.1016/j.eja.2024.127137.

[2] F. Tao, R.P. Rötter, T. Palosuo, C. Gregorio Hernández Díaz-Ambrona, M.I. Mínguez, M.A. Semenov, K.C. Kersebaum, C. Nendel, X. Specka, H. Hoffmann, F. Ewert, Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments, Glob. Chang Biol. 24 (2018) 1291–1307, doi:10.1111/gcb.14019.

[3] S. Minoli, J. Jägermeyr, S. Asseng, A. Urfels, C. Müller, Global crop yields can be lifted by timely adaptation of growing periods to climate change, Nat. Commun. 13 (2022) 7079, doi:10.1038/s41467-022-34411-5.

[4] J. Jägermeyr, C. Müller, A.C. Ruane, J. Elliott, J. Balkovic, O. Castillo, B. Faye, I. Foster, C. Folberth, J.A. Franke, K. Fuchs, Climate impacts on global agriculture emerge earlier in new generation of climate and crop models, Nat. Food 2 (2021) 873–885, doi:10.1038/s43016-021-00400-y.

[5] D.L. Liu, B. Wang, J. Evans, F. Ji, C. Waters, I. Macadam, X. Yang, K. Beyer, Propagation of climate model biases to biophysical modelling can complicate assessments of climate change impact in agricultural systems, Int. J. Climatol. 39 (2018) 424–444, doi:10.1002/joc.5820.

[6] Y. Yang, D.L. Liu, M.R. Anwar, G. O'Leary, I. Macadam, Y Yang, Water use efficiency and crop water balance of rainfed wheat in a semi-arid environment: sensitivity of future changes to projected climate changes and soil type, Theor. Appl. Climatol. 123 (2016) 565–579, doi:10.1007/s00704-015-1376-3.

[7] C. Zhao, B. Liu, S. Piao, X. Wang, D.B. Lobell, Y. Huang, M. Huang, Y. Yao, S. Bassu, P. Ciais, J.L Durand, Temperature increase reduces global yields of major crops in four independent estimates, Proc. Natl. Acad. Sci. 114 (2017) 9326–9331, doi:10.1073/pnas.1701762114.

[8] D.B. Lobell, S. Asseng, Comparing estimates of climate change impacts from process-based and statistical crop models, Environ. Res. Lett. 12 (2017) 015001, doi:10.1088/1748-9326/aa518a.

[9] Soltani, A. and Sinclair, T.R. Modeling physiology of crop development, growth and yield. CABi. 2012.

[10] A.J.W. De Wit, H.L. Boogaard, I. Supit, M van den Berg, System Description of the WOFOST 7.2, Cropping Systems Model, Wageningen University and Research, 2020.

[11] L.G. Van Bussel, P. Grassini, J. Van Wart, J. Wolf, L. Claessens, H. Yang, H. Boogaard, H. de Groot, K. Saito, K.G. Cassman, M.K. van Ittersum, From field to atlas: upscaling of location-specific yield gap estimates, Field Crops Res. 177 (2015) 98–108, doi:10.1016/j.fcr.2015.03.005.

[12] J. Van Wart, P. Grassini, H. Yang, L. Claessens, A. Jarvis, K.G. Cassman, Creating long-term weather data from thin air for crop simulation modeling, Agric. For. Meteorol. 209 (2015) 49–58, doi:10.1016/j.agrformet.2015.02.020.

[13] M. Cucchi, G.P. Weedon, A. Amici, N. Bellouin, S. Lange, H. Müller Schmied, H. Hersbach, Buontempo, CWFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data 12 (2020) 2097–2120, doi:10.5194/essd-12-2097-2020.

[14] C. Navarro-Racines, J. Tarapues, P. Thornton, A. Jarvis, J. Ramirez-Villegas, High-resolution and bias-corrected CMIP5 projections for climate change impact assessments, Sci. Data 7 (2020) 1–14, doi:10.1038/s41597-019-0343-8.

[15] P. Laux, R.P. Rötter, H. Webber, D. Dieng, J. Rahimi, J. Wei, B. Faye, A.K. Srivastava, J. Bliefernicht, O. Adeyeri, J. Arnault, To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data, Agric. For. Meteorol. 304 (2021) 108406, doi:10.1016/j.agrformet.2021.108406.

[16] A. De Wit, H. Boogaard, D. Fumagalli, S. Janssen, R. Knapen, D. van Kraalingen, I. Supit, R. van der Wijngaart, K van Diepen, 25 years of the WOFOST cropping systems model, Agric. Syst. 168 (2019) 154–167, doi:10.1016/j.agsy.2018.06.018.

[17] J.G. Leenaars, L. Claessens, G.B. Heuvelink, T. Hengl, M.R. González, L.G. van Bussel, N. Guilpart, H. Yang, K.G. Cassman, Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa, Geoderma 324 (2018) 18–36, doi:10.1016/j.geoderma.2018.02.046.

[18] J.I. Rattalino Edreira, J.F. Andrade, K.G. Cassman, M.K. van Ittersum, M.P. van Loon, P. Grassini, Spatial frameworks for robust estimation of yield gaps, Nat. Food 2 (2021) 773–779, doi:10.1038/s43016-021-00365-y.

[19] M.K. Van Ittersum, L.G. Van Bussel, J. Wolf, P. Grassini, J. Van Wart, N. Guilpart, L. Claessens, H. De Groot, K. Wiebe, D. Mason-D'Croz, H. Yang, Can sub-Saharan Africa feed itself? Proc. Natl. Acad. Sci. 113 (2016) 14964–14969, doi:10.1073/pnas.1610359113.

[20] M. Mendez, B. Maathuis, D. Hein-Griggs, L.F. Alvarado-Gamboa, Performance evaluation of bias correction methods for climate change monthly precipitation projections over Costa Rica, Water (Basel) 12 (2020) 482, doi:10.3390/w12020482.