Enhancing ANACIM High-Resolution Merged Historical Datasets by Generating Additional Meteorological Parameters

Rija Faniriantsoa | James Hansen

Workshop Report

CGIAR

AICCRA
Accelerating the Impact of CGIAR Climate Research for Africa

Enhancing ANACIM high-resolution merged historical datasets by generating additional meteorological parameters

Workshop Report

Accelerating Impact of CGIAR Climate Research for Africa (AICCRA)

December 21, 2022

Rija Faniriantsoa
James Hansen

COLUMBIA CLIMATE SCHOOL
INTERNATIONAL RESEARCH INSTITUTE
FOR CLIMATE AND SOCIETY

To cite this workshop report
Faniriantsoa R, Hansen JW, 2022. Enhancing ANACIM high-resolution merged historical datasets by generating additional meteorological parameters. AICCRA Workshop Report. Wageningen, the Netherlands: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).

About AICCRA reports
Titles in this series aim to disseminate interim climate change, agriculture, and food security research and practices and stimulate feedback from the scientific community.

About AICCRA
Accelerating Impacts of CGIAR Climate Research in Africa (AICCRA) is a project that helps deliver a climate-smart African future driven by science and innovation in agriculture. It is led by the Alliance of Bioversity International and CIAT and supported by a grant from the International Development Association (IDA) of the World Bank. Explore AICCRA's work at aiccra.cgiar.org

Contact us
Accelerating Impacts of CGIAR Climate Research for Africa (AICCRA). Email: aiccra@cgiar.org

Acknowledgements

The Accelerating Impacts of CGIAR Climate Research for Africa (AICCRA) project is supported by a grant from the International Development Association (IDA) of the World Bank. IDA helps the world's poorest countries by providing grants and low to zero-interest loans for projects and programs that boost economic growth, reduce poverty, and improve poor people's lives. IDA is one of the largest sources of assistance for the world's 76 poorest countries, 39 of which are in Africa. Annual IDA commitments have averaged about $21 billion over circa 2017-2020, with approximately 61 percent going to Africa.

About the Authors

Rija Faniriantsoa is a Staff Associate at the International Research Institute for Climate and Society (IRI), Columbia Climate School, Columbia University, New York. Contact: rijaf@iri.columbia.edu

James W. Hansen is a Senior Research Scientist at the International Research Institute for Climate and Society (IRI), Columbia Climate School, Columbia University, New York. Contact: jhansen@iri.columbia.edu

Acronyms

ADT	Automated Weather Data Tool
AICCRA	Accelerating the Impact of CGIAR Climate Research for Africa project
ANACIM	*Agence Nationale de l'Aviation Civile et de la Météorologie*
AWS	Automatic Weather Station
ENACTS	Enhancing National Climate Services initiative
JRA-55	Japanese 55-year Reanalysis
MERRA-2	Modern-Era Retrospective analysis for Research and Applications Version 2
NASA-POWER	NASA Prediction of Worldwide Energy Resources

Introduction

ANACIM already has a high resolution ENACTS datasets composed of rainfall and temperature data covering the period from 1981 to the present. In order to be able to produce a high-resolution evapotranspiration dataset computed with the Penman-Monteith method, a work has been carried out, with ANACIM staff from November 14th to 22nd 2022 at ANACIM headquarters in Dakar, to produce a high-resolution meteorological dataset required to compute the evapotranspiration.

Activities and Results

The additional meteorological datasets generated are presented in Table 1. To generate the additional high-resolution meteorological datasets, several tasks have to be worked out step by step. Each step is described in the following sections.

Table 1. New merged data sets generated at ANACIM

Datasets	Temporal resolution	Temporal coverage	Spatial resolution
Relative humidity	daily	2019-01-01 to 2022-11-15	0.0375° x 0.0375°
Wind data (speed and direction)	daily	2019-01-01 to 2022-11-15	0.0375° x 0.0375°
Total solar radiation	daily	2019-01-01 to 2022-11-15	0.0375° x 0.0375°
Surface air pressure	daily	2019-01-01 to 2022-11-15	0.0375° x 0.0375°

Inventory available station data

The first step of the dataset production starts with the inventory of all available stations data (both manual observation and from automatic weather stations). The stations data and metadata gathering were challenging, the staff working on collecting the data does not have a direct access to ANACIM database for the manual observation, therefore, we have been forced to use only the stations data available on their personal computers which only cover the period from 2019-01-01 to 2022-11-15. The installation of the Automatic Weather Station Data Tool (ADT) was supposed to be done before proceeding to the production of the high-resolution meteorological datasets, but ANACIM staff working on the data was not available if we followed the established schedule, so the production of the additional meteorological datasets has been moved up two weeks earlier which made it complicated to access to the automatic weather station data and metadata. All data collected has been formatted to a standard format to facilitate their exploitation. The number of stations from which the data were used is summarized in Table 2.

Table 2. Number of observing stations used to generate new merged data at ANACIM.

Variables	Manual stations	Automatic weather stations
Humidity data	24	24
Sunshine duration/radiation data	21	24
Wind data	25	16
Surface air pressure		17
Temperature data		24
Precipitation		226

Figures 1 and 2 show the temporal and spatial distribution of data availability for manual observation and automatic weather stations data.

Station data quality control

Quality control of the stations observed data (manual and automatic) has been carried out to ensure the quality of data to be used for the merging process. The quality check starts with the verification of the station locations. A visual verification was first performed for the time series from the manual station to detect the outliers, then an automatic spatial outlier check at daily time scale was performed. A limit check at a hourly time scale was performed on the time series from the automatic stations followed by a spatial check of the hourly data. All erroneous values were discarded from the time series.

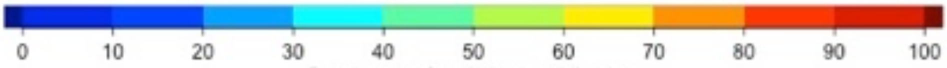
Aggregation to daily time step

After quality control the times series from the automatic weather stations were aggregated to daily temporal resolution to match the manual observation data. The sunshine duration from manual observation were converted into radiation data by using the Angstrom formula.

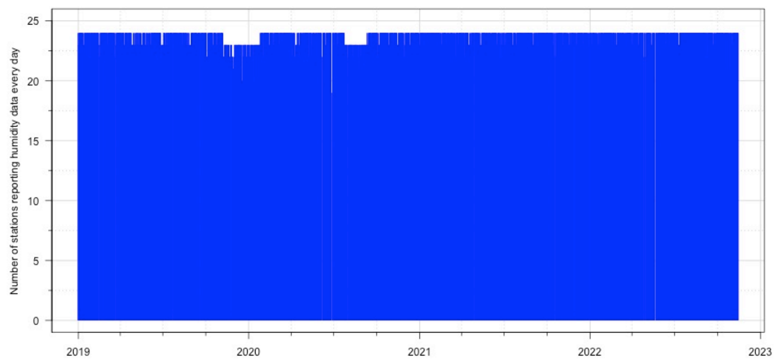
To make sure that times series from AWS are well calibrated and close to the manual stations, a comparison between the daily times series from the automatic and manual stations were carried out. The pair of times series were selected from automatic stations located at or near to (within a radius of 30 km) the manual station sites. We found 3 AWS for radiation and 2 AWS for humidity data where the difference from the times series was very big, these stations have not been used. If the AWS is located at the manual station site, the data from this AWS was used to fill the missing values from the manual; if the AWS is located near to a manual station and does not present a big difference in terms of observed values, the times series from this AWS will be used for the merging.

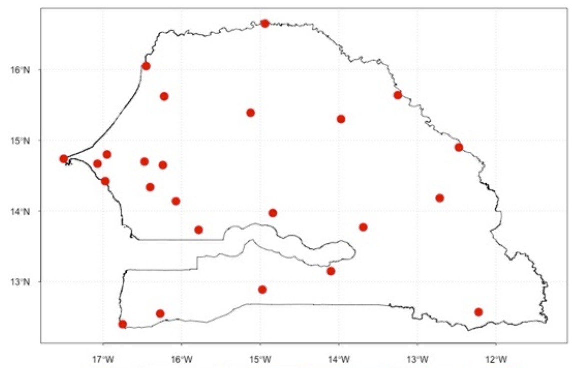
Figure 1. Availability and spatial distribution of data from manual observation stations

Daily availability (number of stations reporting)	Spatial distribution The color bar represents the percentage of non-missing data

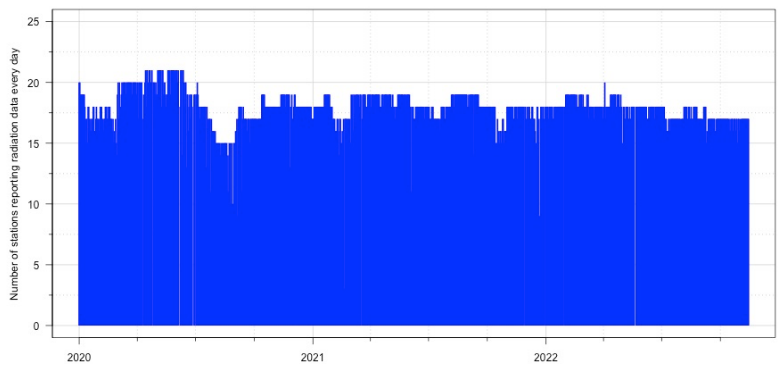


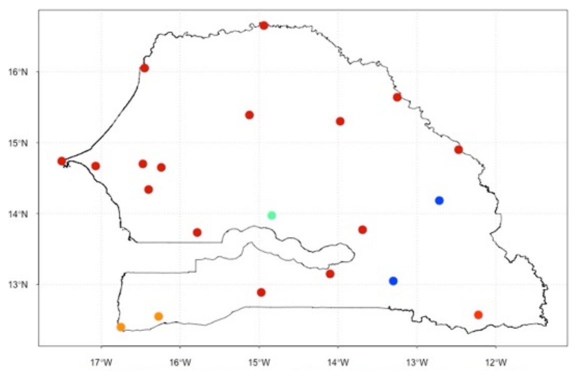
Relative humidity (2019-01-01/2022-11-15)



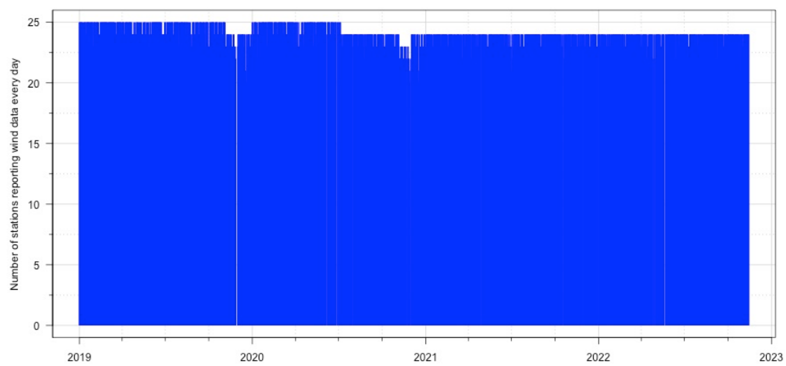


Bright sunshine duration (2020-01-01/2022-11-15)





Wind (2019-01-01/2022-11-15)



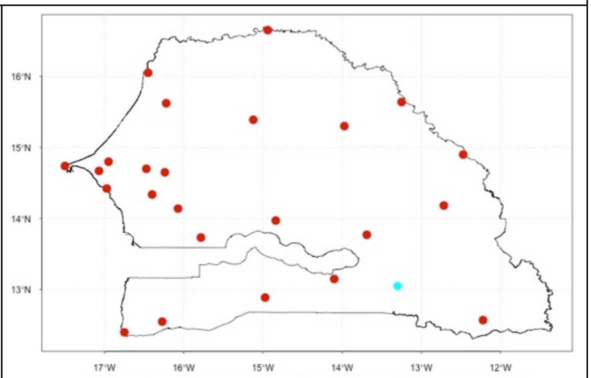
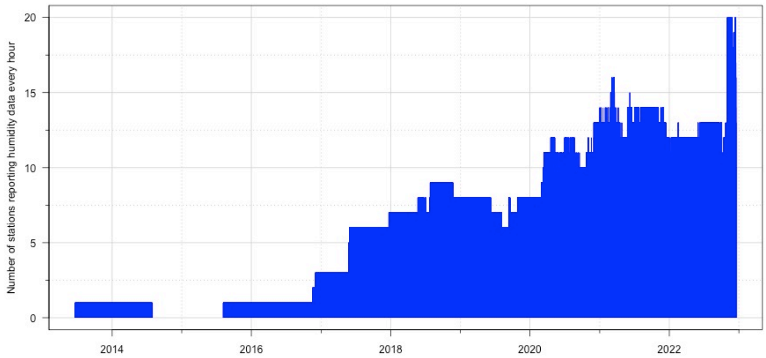
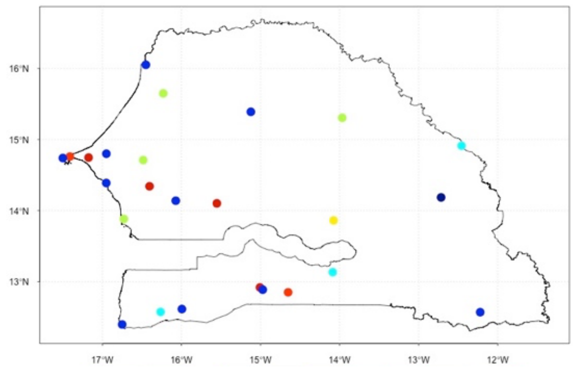


Figure 2. Availability and spatial distribution of data from automatic weather stations (AWS)

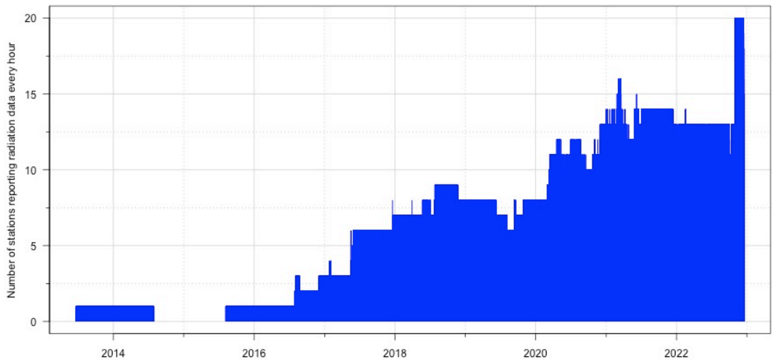
Daily availability (number of stations reporting)	Spatial distribution The color bar represents the percentage of non-missing data

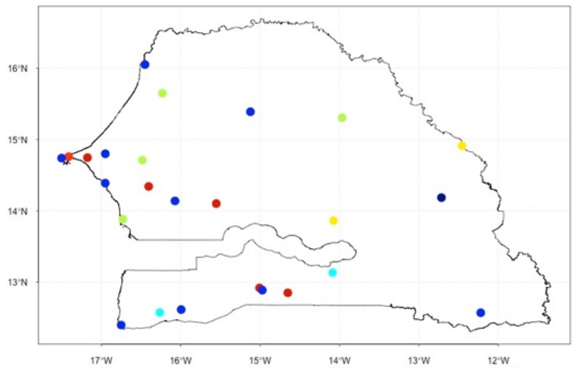
Relative humidity (2019-01-01/2022-11-15)



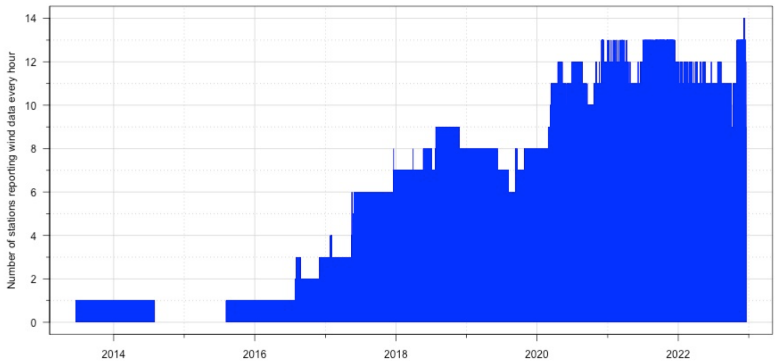


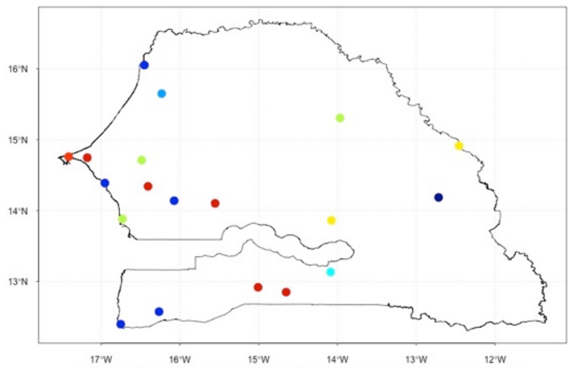
Solar radiation (2019-01-01/2022-11-15)



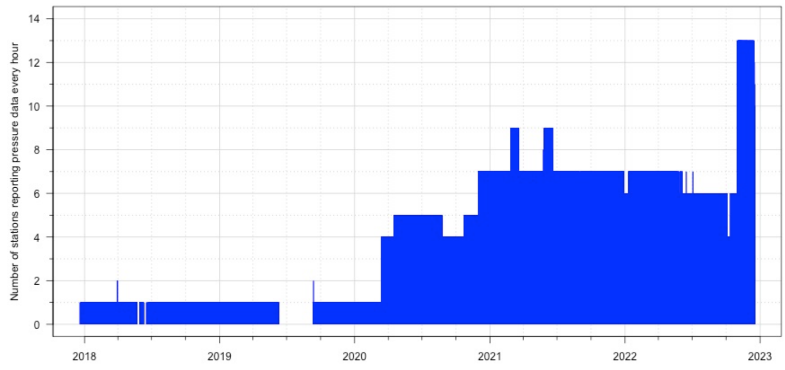
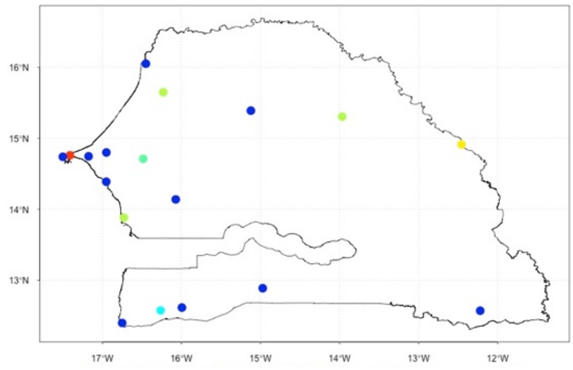
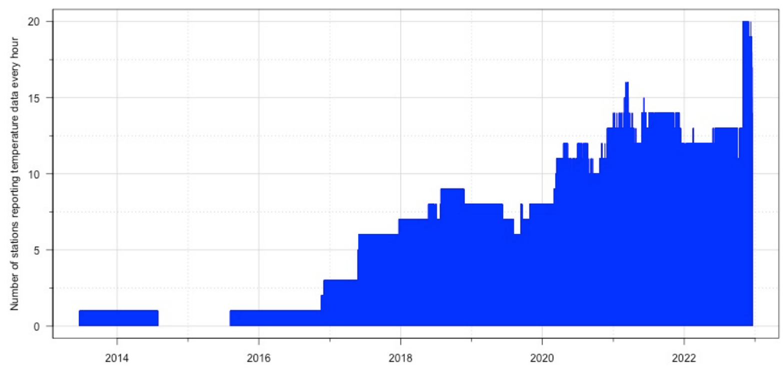
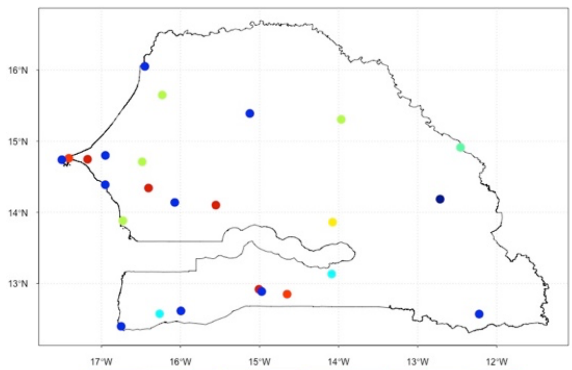
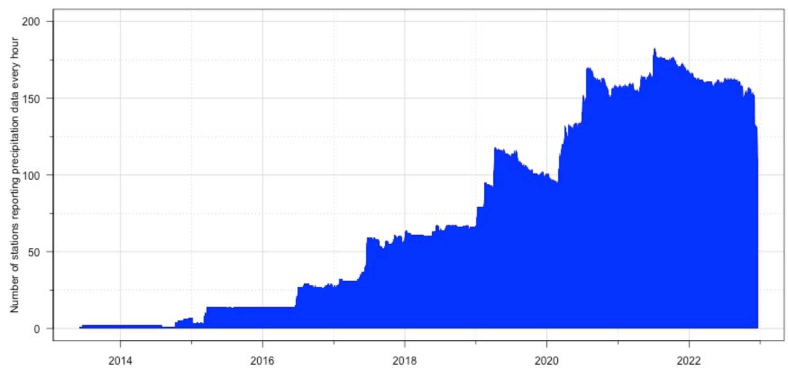
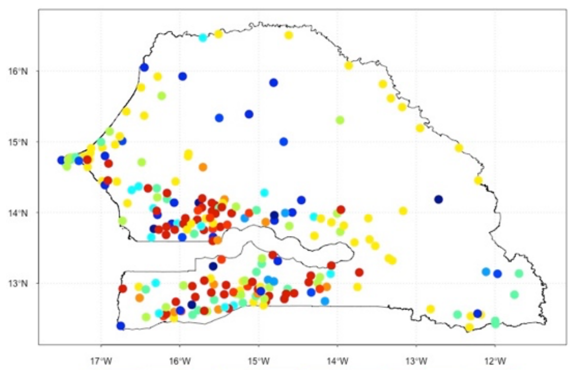


Wind (2019-01-01/2022-11-15)





Surface air pressure (2019-01-01/2022-11-15)

<table>
<tr><th>Daily availability (number of stations reporting)</th><th>Spatial distribution
The color bar represents the percentage of non-missing data</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Temperature (2019-01-01/2022-11-15)</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Precipitation (2019-01-01/2022-11-15)</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Reanalysis data evaluation and selection

The next step was to select the reanalysis data to merge with the station data. Two reanalysis products were downloaded, the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) and the Japanese 55-year Reanalysis (JRA-55), an additional source of data, the NASA Prediction of Worldwide Energy Resources (NASA-POWER) was also downloaded in order to be compared with the 2 reanalysis products. To perform the comparison, the gridded data was extracted at the station locations, then the series pairs (station and extracted data) will be used to compute a statistical indicator to measure the performance of the 3 gridded data. The correlation, multiplicative bias, root mean square error and Nash-Sutcliffe efficiency were computed. Table 3 shows the computed statistics of the daily mean relative humidity using all data from the stations and gridded data extracted at the corresponding station locations.

Table 3. Goodness-of-fit statistics between station and reanalysis daily mean relative humidity data

Statistic	JRA-55	MERRA-2	NASA-POWER
Correlation	0.844	0.784	0.809
Multiplicative bias	1.309	1.559	1.508
Root Mean Square Error	18.286	23.228	22.941
Nash-Sutcliffe Efficiency	0.259	-0.345	-0.166

JRA-55 perform much better than MERRA-2 and NASA-POWER. It is important to note that NASA-POWER meteorological parameters are derived from MERRA-2 and GEOS 5.12.4 FP-IT, NASA-POWER data has been improved, therefore it performs better than MERRA-2.

Figure 3 shows the scatter plot of daily mean relative humidity for the period 2019-2022 from stations versus gridded data extracted at the corresponding station locations. There is a clear overestimation of the gridded data compared to the station data. Figures 4-6 show the correlation, multiplicative bias and root mean square error of the daily mean relative humidity for the period 2019-2022. The statistics of each station were computed separately. All the computed statistics, for the relative humidity, radiation, wind speed data whether calculated using all data combined or for each station, show a better performance of JRA-55. Therefore, we used JRA-55 as a proxy to merge with the stations data.

As the reanalysis data show obvious overestimation, the next step would obviously be the bias correction of the reanalysis data, but we did not have a long series of data at our disposal we skipped this step.

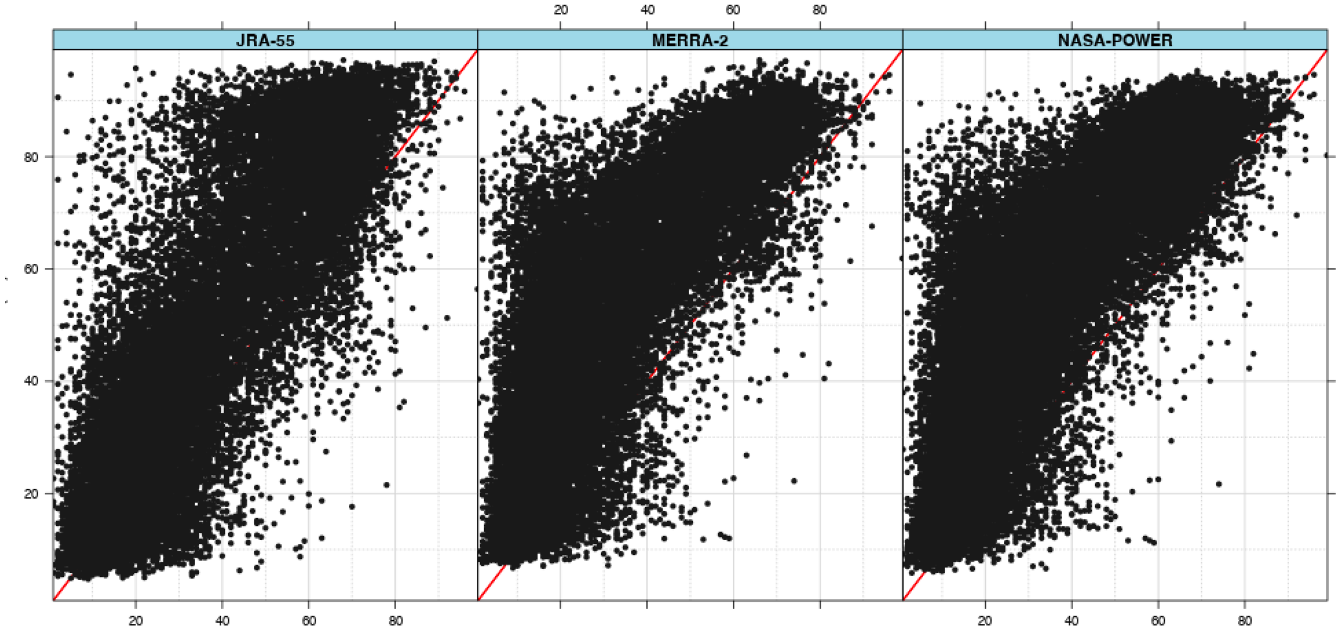


Figure 3. Scatter plot of daily mean relative humidity from stations vs. merged data at station locations, for the period 2019-2022

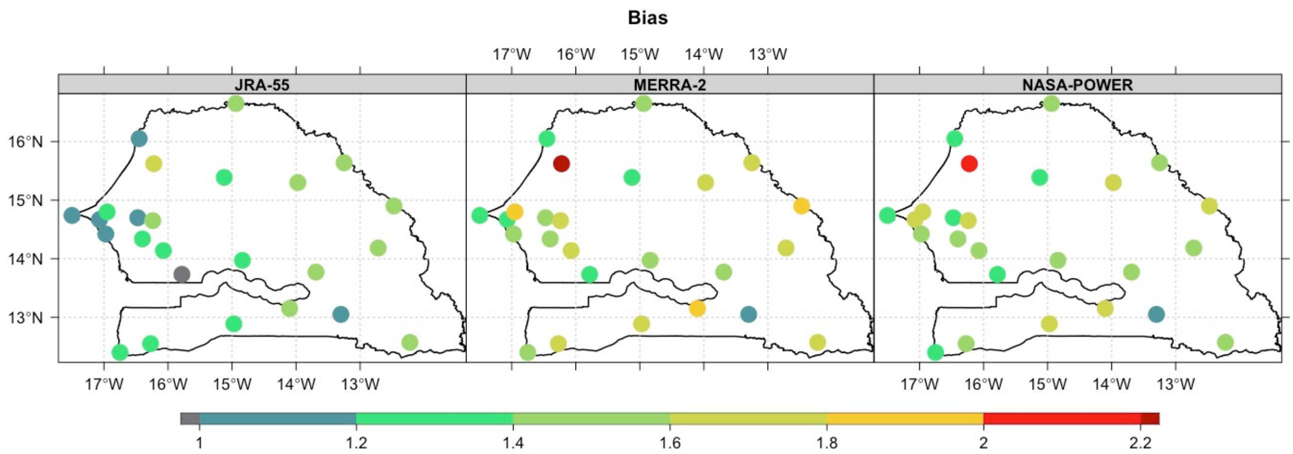


Figure 4. Mean bias error of daily mean relative humidity between stations and merged data at station locations, for the period 2019-2022

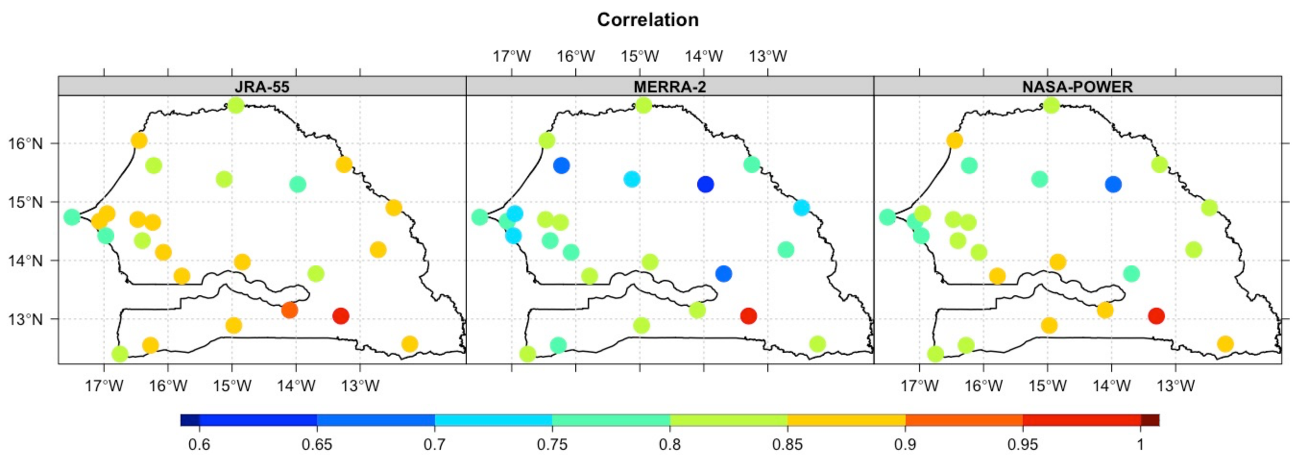


Figure 5. Correlation of daily mean relative humidity between stations and merged data at station locations, for the period 2019-2022

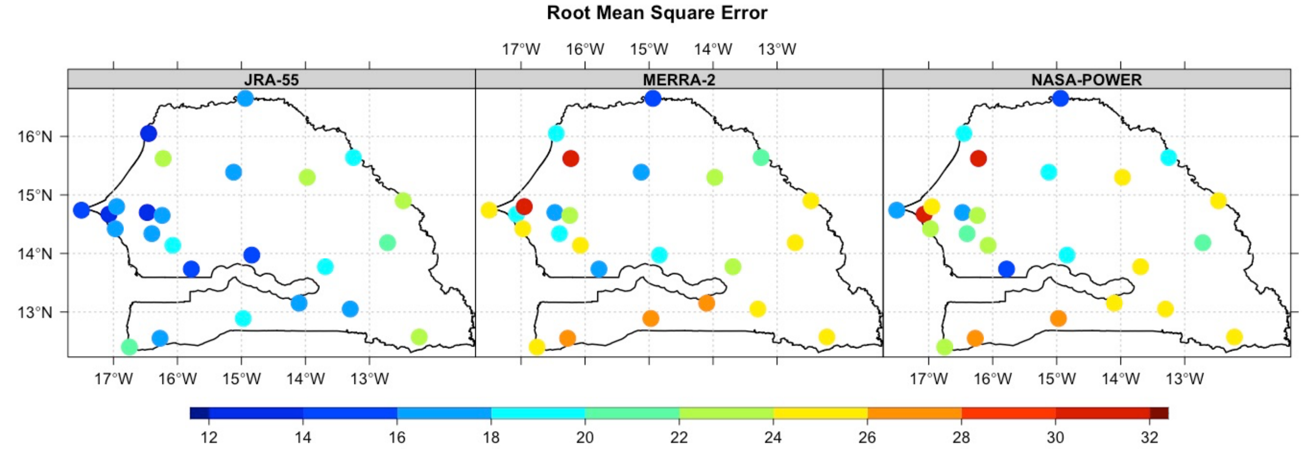


Figure 6. Root mean square error of daily mean relative humidity between stations and merged data at station locations, for the period 2019-2022

Merged data generation

When the all the data required to merge the station data and reanalysis were ready, we proceeded to the selection of the methods to interpolate and merge the data. We performed several merging runs using all possible combinations of the merging and interpolation methods. The merging was performed over one year using a daily data for all meteorological parameters. We selected the merging and interpolation method that best reproduces the stations data when the merged data are extracted at the station locations. The methods selected were used to produce the high-resolution meteorological datasets presented in Table 1.

Conclusions and Recommendations

To have a long series of high-resolution merged historical datasets of those meteorological parameters, it is strongly recommended that ANACIM staff undertakes to reproduce all the steps described above after gathering all the historical meteorological parameters from manual observation stations, and for at least covering a period of 30 years. The steps to be followed can be summarized as follows:

- Gathering historical data from manual stations
- Formatting the station data
- Performing a quality control
- Downloading JRA-55 reanalysis
- Adjusting the bias of JRA-55
- Merging the bias corrected JRA-55 with the station data using the merging selected methods.

An evaluation of the merged data must be performed. A data catalog at ANACIM Data Library should be created to facilitate the access of the merged data.

Appendix

Activity Participants

No	M/F	Organization	Location
1	M	Civil Aviation and Meteorology Service in Senegal (ANACIM)	Dakar
2	M	Civil Aviation and Meteorology Service in Senegal (ANACIM)	Dakar
3	M	Civil Aviation and Meteorology Service in Senegal (ANACIM)	Dakar
4	M	Civil Aviation and Meteorology Service in Senegal (ANACIM)	Dakar
5	M	IRI, AICCRA	New York

The CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) brings together some of the world's best researchers in agricultural science, development research, climate science and Earth system science, to identify and address the most important interactions, synergies and tradeoffs between climate change, agriculture and food security. For more information, visit us at https://ccafs.cgiar.org/.

Titles in this series aim to disseminate interim climate change, agriculture and food security research and practices and stimulate feedback from the scientific community.

AICCRA is led by:

Alliance

Bioversity International

CIAT

AICCRA is supported by the International