EMI R&D PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2026
Project Title:	A High-Resolution Hindcast for the Sahel Region Based on ERA5 data and the MOLOCH model
Computer Project Account:	SPITCAPE
Principal Investigator(s):	Valerio Capecchi
Affiliation:	Consorzio LaMMA, Laboratorio di Meteorologia e Modellistica Ambientale per lo sviluppo sostenibile
Name of ECMWF scientist(s) collaborating to the project (if applicable)	
Start date of the project:	The project was submitted in 2025 as Late Request for Special Project; here we report the activities carried out during 2026
Expected end date:	2028

Computer resources allocated/used for the current year and the previous one

(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	4.0M	3.9M	10.0M	4.4M
Data storage capacity	(Gbytes)				

Summary of project objectives (10 lines max)

The Special Project is part of the “SLAPIS Sahel” initiative (https://climateservices.it/progetto/slapis/), funded by the Italian Agency for Development Cooperation and implemented by the Polytechnic University of Turin in collaboration with the Institute for Bioeconomy of the National Research Council of Italy (IBE-CNR), the National Meteorology Directorate (DMN) of Niger, and the National Meteorology Agency (ANAM) of Burkina Faso. One of the main objectives of the project is to produce a long-term, high-resolution hindcast for the Sahel region. This is achieved through the dynamical downscaling of the ERA5 reanalysis using the MOLOCH limited-area model. The resulting dataset is intended to support climate services for climatological, agricultural, and hydrological applications. In addition, it will serve as a reference dataset for training a regional, data-driven machine learning (ML) model within the Anemoi framework.

Summary of problems encountered (10 lines max)

The preprocessing and model integration phases were completed without issues. However, a bug in the post-processing code caused problems in extracting variables on pressure levels. Despite discussions with the MOLOCH model developers, we have not been able to resolve the issue so far. As a result, to date only surface variables have been successfully extracted and archived.

Summary of plans for the continuation of the project (10 lines max)

For the continuation of the project, we intend to complete the 5-year period 2014-2018. We plan to extract and archive only the surface variables, since the issue related to the extraction of variables on pressure levels has not yet been resolved. However, since the extraction of pressure-level variables is crucial for the training of data-driven models, we aim to resolve the bug in the post-processing procedure. So far, we have also archived the soil variables and plan to archive them for the period 2014-2018. In fact, such variables are needed should we decide to rerun the simulations.

List of publications/reports from the project with complete references

None. However, we plan to assess the quality of the hindcast estimates by comparing model outputs with observational data. If the evaluation is satisfactory, we intend to submit the dataset for publication in a peer-reviewed scientific journal once the hindcast is complete or, at a minimum, after at least 15 years of simulated data have been produced.

Summary of results

A 7-year period of the MOLOCH hindcast, covering 2019–2025 over the African domain shown in Figure 1, has been completed. Since the hydrologically relevant season extends from May to October, only these months were simulated, while the remaining months were excluded.

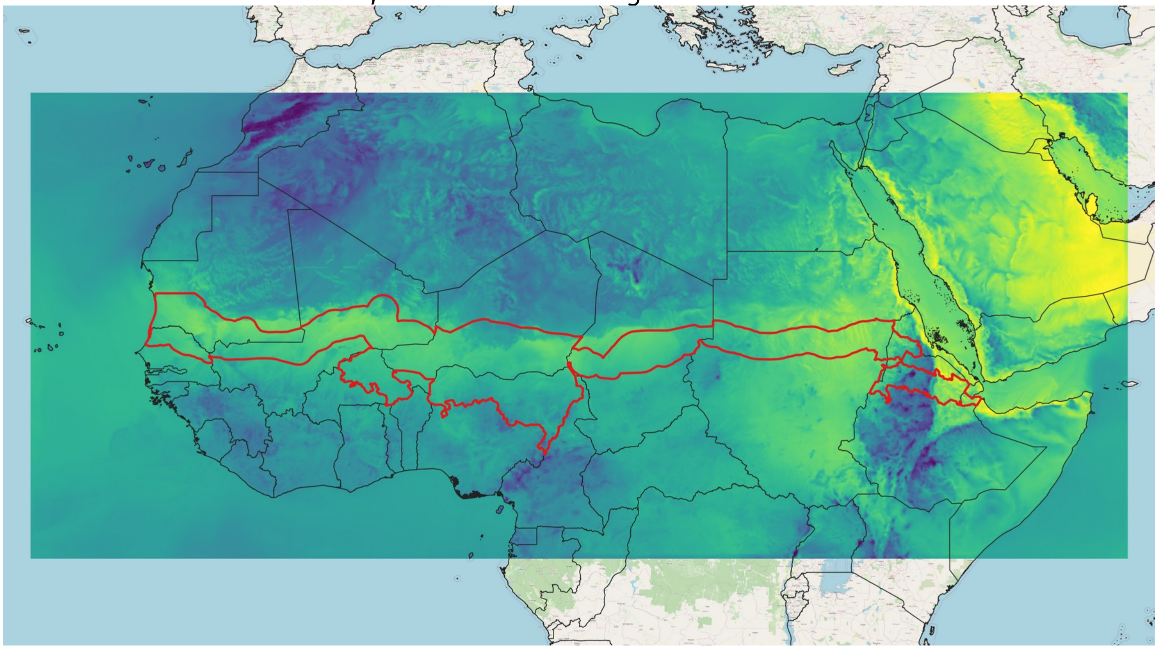


Figure 1: sample data of the MOLOCH hindcast. The red line indicates the Great Green Wall

In the following section, we report the **geographical settings** of the integration domain: grid points 1922x802, grid spacing ~4 km, latitude=(0.0,33.3), longitude=(-25.0, 53.7), vertical levels 70. Model physical settings are detailed in the documentation available on the gitlab repository (only in Italian): https://gitlab.com/isac-meteo/globo-bolam-moloch/-/blob/main/modelli_ISAC_corso_utenti.pdf?ref_type=heads.

The **experimental design** is as follows: initial conditions are provided by the ERA5 data with boundary conditions updated every 3 hours. The MOLOCH simulation starts each day at 18:00 UTC and has a forecast length equal to 30 hours. The MOLOCH model produces outputs every hour over the domain of integration shown in Figure 1. The daily data of the MOLOCH hindcast are built using the last 24 h of the model simulation, while the first 6 hours of integration are considered as spin-up time and thus discarded.

Near surface variables were extracted and archived in the ECFS. The list of archived variables is provided below, with units in brackets:

- Pressure reduced to MSL [Pa];
- total precipitation [kg m-2];
- Total snowfall [kg m-2];
- total cloud [%];
- downward short-wave radiation flux [W m-2];
- evaporation [kg m-2];
- snow depth water equivalent [m];
- lapse rate [Kg m-1];
- soil temperature [K];
- Sensible heat net flux [W m-2];
- Latent heat net flux [W m-2];
- 2-metre above ground temperature [K];
- 2-metre above ground relative humidity [%];
- 2-metre above ground specific humidity [Kg Kg-1];
- 2-metre above ground dew point [K];
- 10-metre, 50-metre, 80-metre and 100-metre above ground u-component of wind [m s-1];
- 10-metre, 50-metre, 80-metre and 100-metre above ground v-component of wind [m s-1];
- 10-metre above ground wind gust [m s-1].

The computational cost of performing a single-day simulation, broken down by job type, is as follows:

JOBs

pre-processing (QoS = nf)
    # Consumption : ~4 SBUs
    # ElapsedRaw : ~800 sec
model run (QoS = np)
    # Consumption : ~6000 SBUs
    # ElapsedRaw : ~800 sec
post-processing (QoS = nf)
    # Consumption : ~1 SBUs
    # ElapsedRaw : ~130 sec

This yields an approximate computational cost for one simulation year, corresponding in practice to a 6-month simulation period, of:
total computational cost = 183 X (4+6000+1) = 1.1 M SBUs
Based on the numbers reported above, the wall-clock time needed to perform a single day simulation is approximately:
1-day wall-time = (800+800+130) = 1730 sec ~ 29 min
Since the daily simulations must be performed sequentially, the wall-clock time required to complete the 6-month simulation period for each year is approximately 91 days. This is because each daily simulation relies on the soil-state output from the previous day, *d−1*, to initialise the soil state for day *d*. Therefore, a new daily simulation can only be started after the previous one has completed. This procedure avoids abrupt soil-state initialisation that would otherwise result from using global data directly.

Output samples
The figures below show selected average fields derived from the model simulations completed to date, covering the 7-year period 2019–2025. A more comprehensive analysis, including a formal assessment of model accuracy, will be performed once simulations spanning at least 15 years have been completed.
However, to provide a preliminary indication of the MOLOCH hindcast output, the figures show modelled precipitation data compared with observed data from selected weather stations in the Sahel region. The observations are taken from the GSOD dataset - Global Surface Summary of the Day - available at https://www.ncei.noaa.gov/data/global-summary-of-the-day/access/, and refer to daily precipitation amounts in July. Locations are indicated in Figure 2.

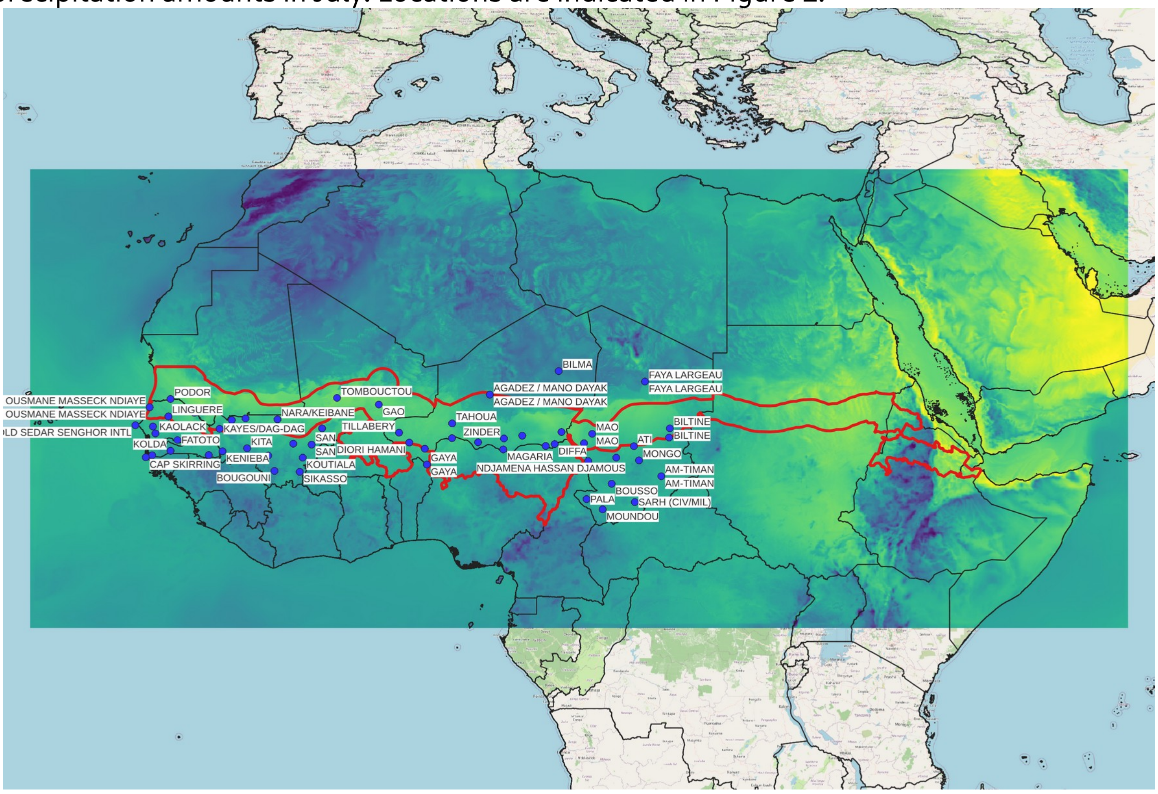
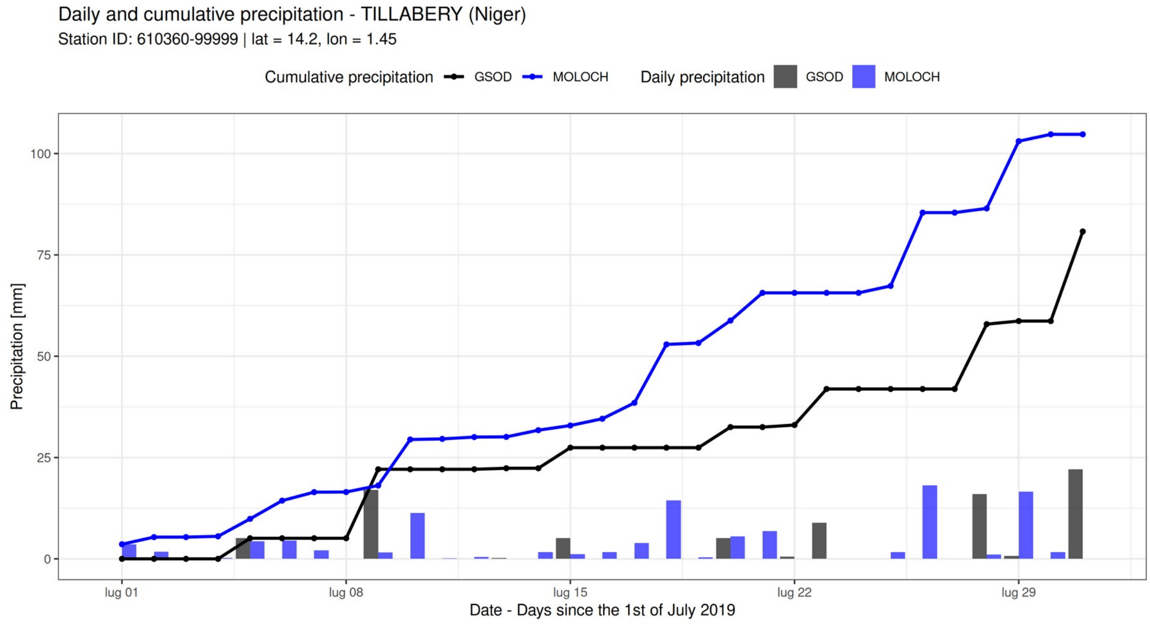
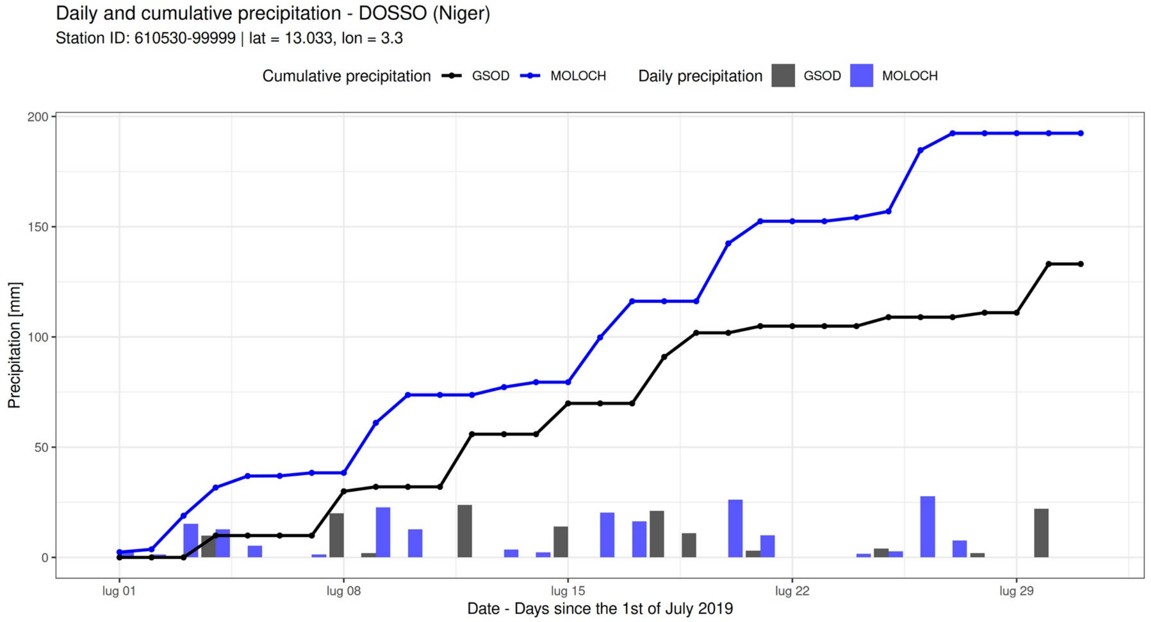
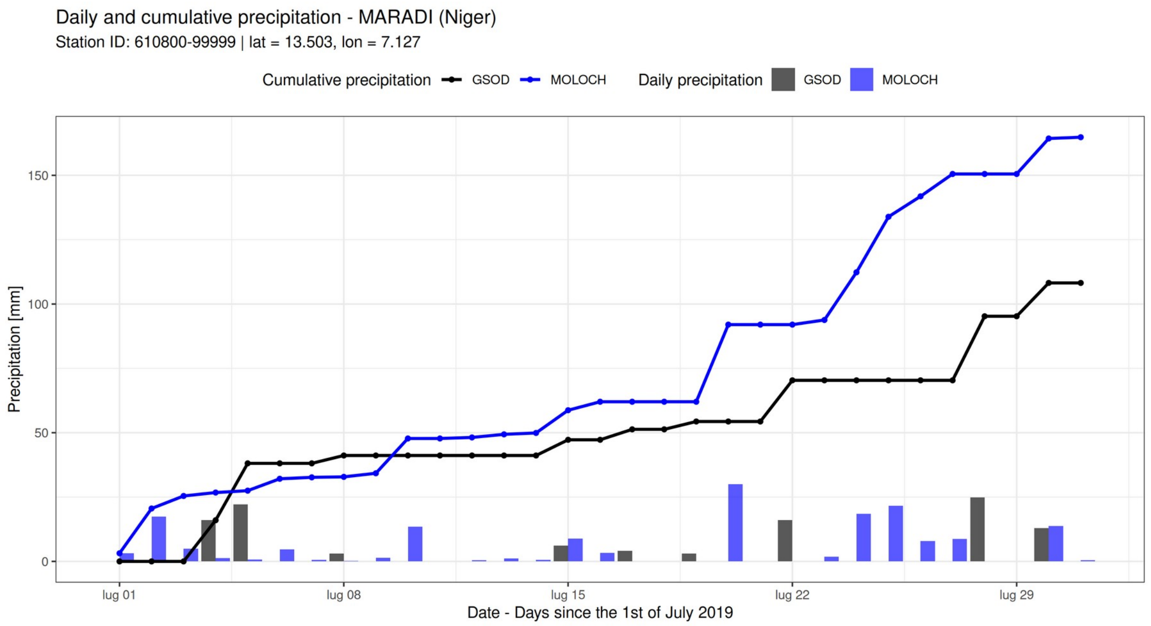


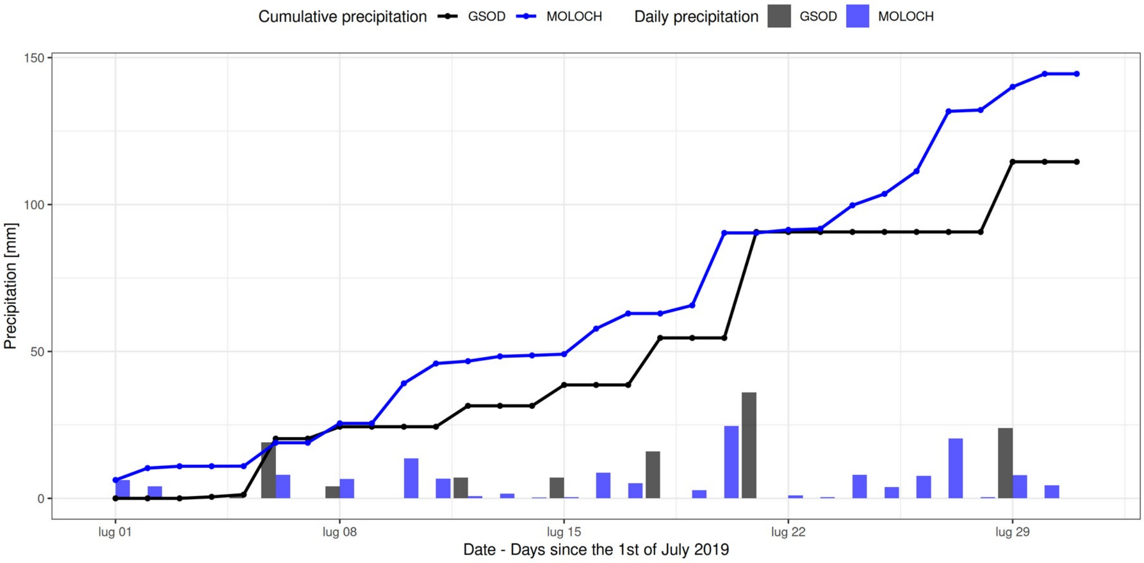
Figure 2: approximate location of the weather stations used for comparison aginst modelled MOLOCH data


Daily and cumulative precipitation - TILLABERY (Niger)
Station ID: 610360-99999 | lat = 14.2, lon = 1.45
Cumulative precipitation
GSOD
MOLOCH
Daily precipitation
GSOD
MOLOCH
Precipitation [mm]
0
25
50
75
100
lug 01
lug 08
lug 15
lug 22
lug 29
Date - Days since the 1st of July 2019

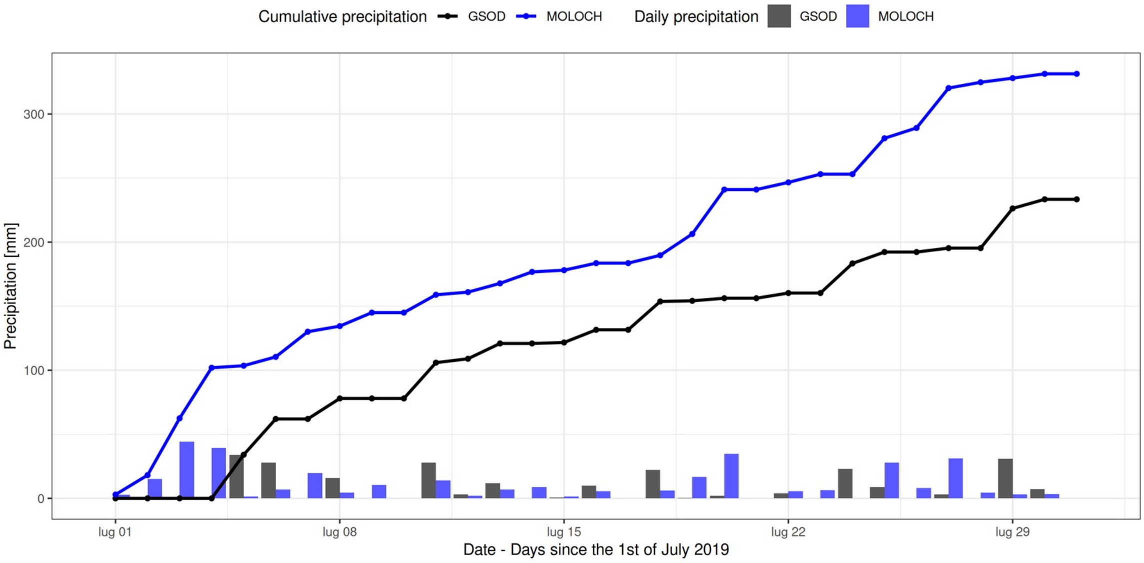

Daily and cumulative precipitation - DOSSO (Niger)
Station ID: 610530-99999 | lat = 13.033, lon = 3.3
Cumulative precipitation
GSOD
MOLOCH
Daily precipitation
GSOD
MOLOCH
Precipitation [mm]
0
50
100
150
200
lug 01
lug 08
lug 15
lug 22
lug 29
Date - Days since the 1st of July 2019


Daily and cumulative precipitation - MARADI (Niger)
Station ID: 610800-99999 | lat = 13.503, lon = 7.127
Cumulative precipitation
GSOD
MOLOCH
Daily precipitation
GSOD
MOLOCH
Precipitation [mm]
0
50
100
150
lug 01
lug 08
lug 15
lug 22
lug 29
Date - Days since the 1st of July 2019

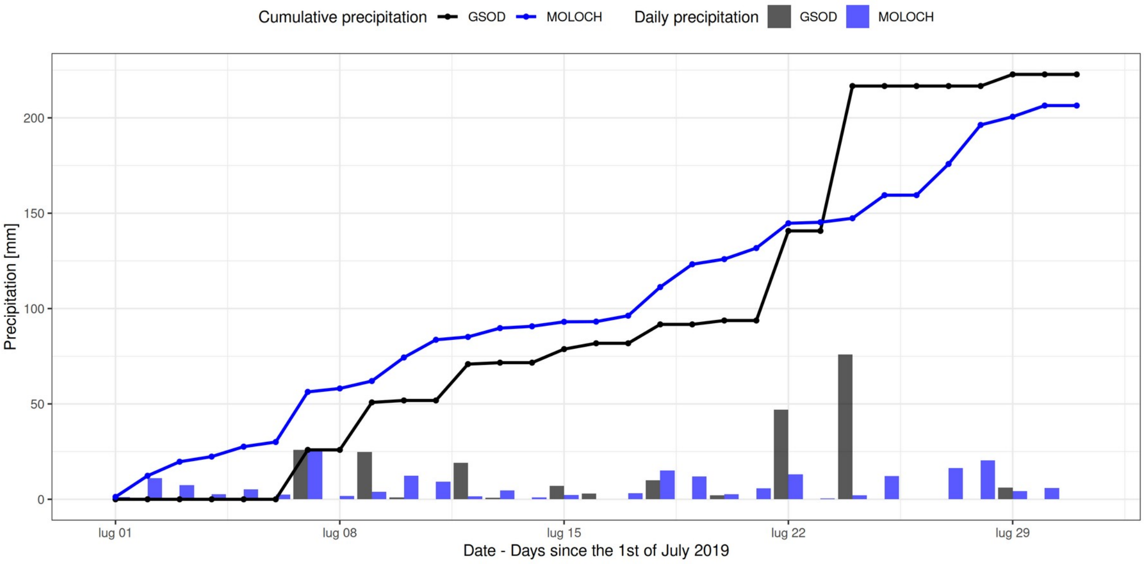
Daily and cumulative precipitation - ZINDER (Niger)
Station ID: 610900-99999 | lat = 13.779, lon = 8.984



Daily and cumulative precipitation - KITA (Mali)
Station ID: 612700-99999 | lat = 13.067, lon = -9.467



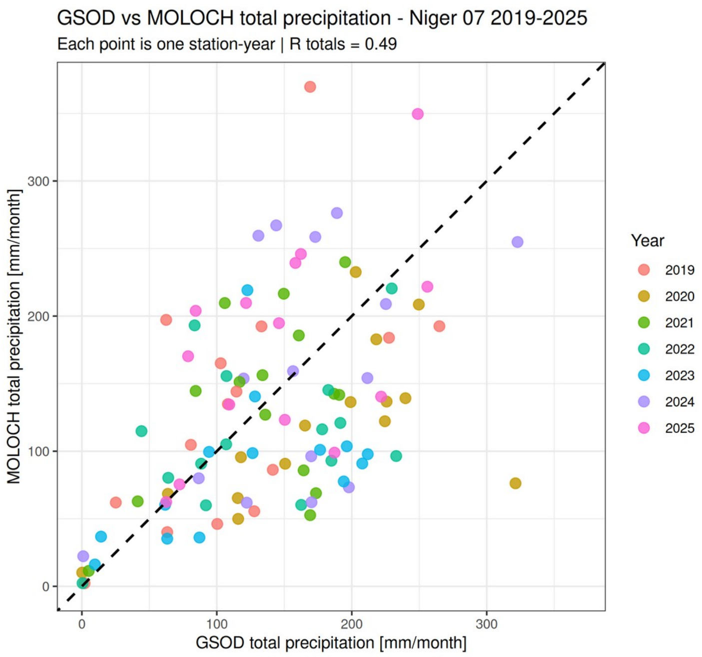
Daily and cumulative precipitation - SEGOU (Mali)
Station ID: 612720-99999 | lat = 13.4, lon = -6.15

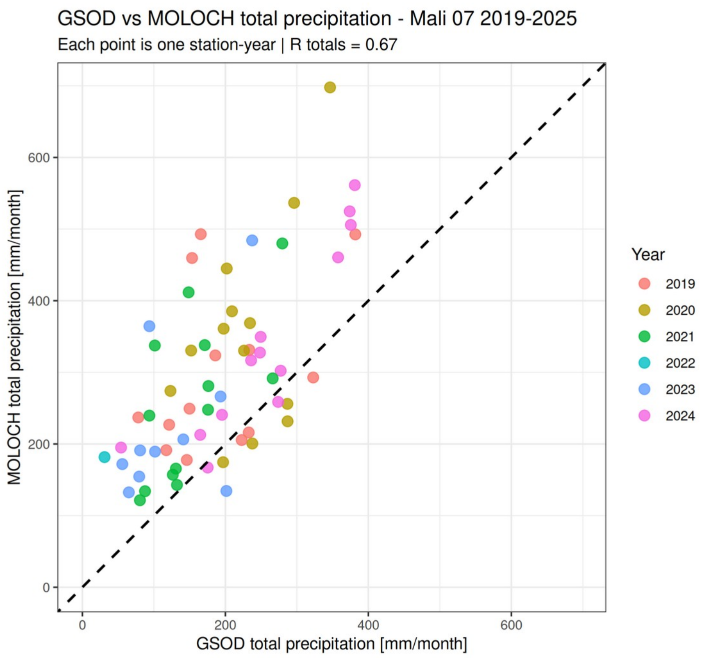


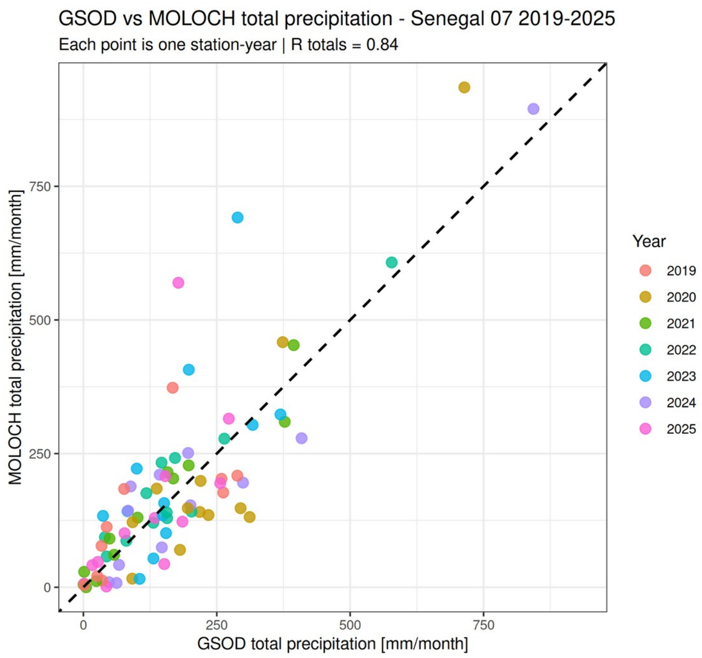
The sample plots above suggest that the MOLOCH data may overestimate the precipitation. To test this hypothesis, we calculated the daily root mean square error (RMSE) and the additive BIAS defined as the average difference between modelled and observed data. On average among the 7-year period 2019-2025 and all the rain-gauges shown in Figure 2, we obtained the following results:

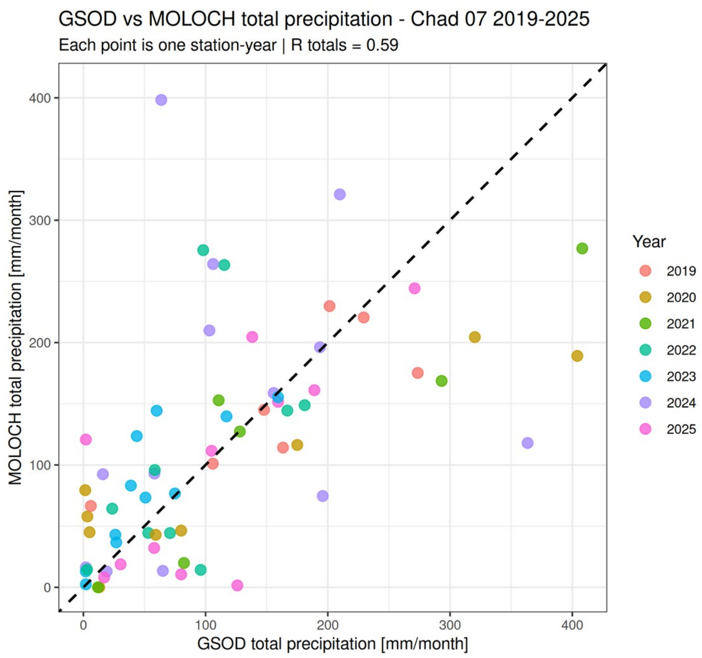
Country	RMSE (mm/day)	BIAS (mm/day)
Senegal	16.9	0.42
Mali	19.7	3.54
Niger	13.7	-0.32
Chad	14.7	0.82

This confirms that MOLOCH generally overestimates precipitation (BIAS is positive) with a RMSE that spans from 13.7 to 19.7 mm/day.
However, scatter plots demonstrate a good agreement between observed and modelled data for all the countries with Pearson correlation coefficients spanning from 0.49 to 0.84 (see the figures below)






GSOD vs MOLOCH total precipitation - Senegal 07 2019-2025
Each point is one station-year | R totals = 0.84
MOLOCH total precipitation [mm/month]
0
250
500
750
GSOD total precipitation [mm/month]
0
250
500
750
Year
2019
2020
2021
2022
2023
2024
2025


GSOD vs MOLOCH total precipitation - Chad 07 2019-2025
Each point is one station-year | R totals = 0.59
MOLOCH total precipitation [mm/month]
0
100
200
300
400
GSOD total precipitation [mm/month]
0
100
200
300
400
Year
2019
2020
2021
2022
2023
2024
2025