# Online Resource 1: Supplementary review methods and constructed example

Rainfall timing and forecast verification for agricultural climate services in West Africa: a critical review

Journal: Theoretical and Applied Climatology

Authors: Segun Ezekiel Akorede; Opeyemi Samuel Ibitoye; Chioma Suzan Dibani; Chinyere Salome Ofordu

Corresponding author affiliation: International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences,

Corresponding author email: segun.akorede@st.futminna.edu.ng

Online Resource 1 documents the focused review approach, the evidence supporting the rainfall-climatology and verification synthesis, the application reporting record, and complete specifications of the constructed example. Online Resource 2 supplies the numerical data and code. All rainfall arrays in the example are deliberately specified; they are not observations or hindcasts.

## Appendix S1 Review approach and literature update

### S1.1 Question and scope

The review asks under what climatological and forecast-verification conditions a rainfall skill estimate can support an agricultural application. It compares rainfall targets, seasonal windows, event frequencies, forecast coverage, spatial support, climatological benchmarks and rainfall references. West Africa supplies the principal evidence context. Agricultural information-use and value studies identify the additional conditions for an outcome claim; they are not treated as evidence of rainfall skill. The synthesis does not estimate pooled effects or the prevalence of verification methods.

The initial synthesis covered material identified through 24 August 2026, using targeted searches of Google Scholar and publisher platforms, citation tracing, and official World Meteorological Organization, AGRHYMET, and regional climate-centre material. Complete query histories, screening counts, and a source-level exclusion log are unavailable. The concepts below describe the reported retrieval topics rather than a fully reproducible systematic-search record.

## S1.2 Search concepts

The search concepts combined the following terms:

(rainfall OR precipitation) AND (forecast OR prediction) AND (verification OR skill OR value) AND (agriculture OR farming).

(onset OR “dry spell” OR “heavy rainfall” OR cessation) AND forecast AND (“West Africa” OR Sahel OR a specified country).

(“climate service” OR “weather information”) AND (farmer OR agriculture) AND (use OR decision OR governance OR evaluation).

(“forecast value” OR “cost-loss” OR “willingness to pay” OR utility) AND agriculture AND Africa.

(CHIRPS OR TAMSAT OR gauge OR satellite) AND rainfall AND verification AND Africa.

CHIRPS denotes Climate Hazards Group Infrared Precipitation with Stations (Funk et al. 2015); TAMSAT denotes Tropical Applications of Meteorology using Satellite data and ground-based observations (Maidment et al. 2017). These names identify rainfall-reference products, whose version and support must be recorded for an application.

These combinations describe the search concepts rather than exact platform-specific query syntax. The available search documentation does not support an assessment of retrieval completeness.

## S1.3 Eligibility and interpretation

The synthesis includes methodological sources on verification and decision value; studies of rainfall seasonality, onset and dry-spell indices that define the target or calendar; rainfall forecast and hindcast evaluations relevant to West Africa; and agricultural service studies that delimit information use and outcome claims. Cross-sectoral sources supply established methodological principles. Climate-impact studies are used only where they directly illuminate rainfall-index or reference-data uncertainty, not as evidence of operational forecast performance.

Sources were interpreted by target, calendar definition, forecast horizon and support, reference, benchmark and supported inference. Agricultural sources were additionally classified by action, response constraints and type of value evidence. The notes in Appendix S2 state these roles

without implying that each paper supplies all fields. Climatological characterisation, hindcast scores, qualitative evidence, simulations and observational associations remain distinct.

No preregistered protocol, duplicate screening or formal risk-of-bias assessment was used. The review cannot support exhaustive coverage, pooled effect sizes, regional estimates of welfare return or estimates of how often practitioners use particular methods. The selected West African literature also gives uneven coverage of countries, rainfall regimes and production systems.

### S1.4 Targeted literature update

The targeted update on 5 October 2026 checked recent user-centred and operational-verification frameworks (Emerton et al. 2026; Pagano et al. 2024, 2025). It also added directly relevant rainfall and climate-data studies: Gbangou et al. (2020) on Ghanaian rainfall and dry-spell prediction, Romanovska et al. (2023) on West African climate-input uncertainty, and Mendes et al. (2025) on rainy-season timing. Publisher records and full-text interpretations were checked for these additions.

The update supplemented the focused synthesis without repeating the initial retrieval and screening process. It therefore does not establish exhaustive literature coverage through October 2026 or change the review's design to a systematic review.

## Appendix S2 Evidence supporting the climatological and verification synthesis

### S2.1 Verification theory and related evaluation frameworks

Murphy (1993) establishes that forecast quality and value are different dimensions. Jolliffe and Stephenson (2012), Wilks (2019), Hersbach (2000), and Gneiting and Raftery (2007) provide the basis for format-matched diagnostics and proper scores. These sources support metric selection and interpretation; they do not demonstrate agricultural outcomes from ARFEF.

Crochemore et al. (2024) connect verification with seasonal service evaluation, while Emerton et al. (2026) already address decision context, thresholds, historical event evaluation, and communication. The Agricultural Rainfall Forecast Evaluation Framework (ARFEF) specifies rainfall calendar conditioning, future-window onset rules, matching support, reference sensitivity, and evidence boundaries within this established literature. It does not claim that earlier frameworks omit user timing or that its reporting sequence has been operationally validated. Pagano et al. (2024, 2025) supply the operational planning and archive context.

## S2.2 Rainfall regimes and forecast targets

Nicholson (2013), Dunning et al. (2016) and Sultan et al. (2005) establish rainfall-regime and crop-timing context. Dunning et al. distinguish seasonal-cycle boundaries from a planting-condition threshold. Mendes et al. (2025) describe agronomically defined rainy-season timing and its spatial progression; these dates are conditional on the study definition. Vellinga et al. (2013), Gbangou et al. (2019), and Kumi et al. (2020) evaluate different regional or local onset targets. Their dates and skill estimates should not be treated as interchangeable.

These studies support recording the rainfall regime, eligible season or sowing window, evaluation years, event frequency, and calendar conditioning of the benchmark. Gbangou et al. (2020) compare seasonal rainfall and dry-spell indices in Ghana and find index-, station- and lead-dependent performance. This is empirical evidence for those evaluated indices, not a forecast of every post-sowing dry spell or a test of ARFEF. Main-manuscript Table 1 summarises the target distinctions.

Rodrigues et al. (2014) and Siegmund et al. (2015) address seasonal rainfall forecast performance and configuration dependence. Bliefernicht et al. (2019) evaluate regional outlook quality and potential value. Vogel et al. (2021) establish the relevance of statistical rainfall-occurrence comparators in northern tropical Africa. These studies supply product-performance evidence within their own designs, not empirical tests of the ARFEF sequence.

Sivakumar (1992) analyses dry spells relative to agricultural timing. The study defines onset after 1 May using at least 20 mm accumulated over three consecutive days and no dry spell exceeding seven days within the next 30 days. This literature example is not a universal planting standard. The numerical construction in Appendix S4 explicitly uses non-overlapping initial and subsequent windows; it is not a reimplementation of the study's archive or all its agronomic conventions.

## S2.3 References and observational support

Funk et al. (2015) describe CHIRPS and Maidment et al. (2017) describe TAMSAT; product provenance does not establish universal local-event accuracy. Wang et al. (2023) illustrate rainfall-reference and aggregation dependence in forecast verification. Romanovska et al. (2023) identify reference and agroclimatological-index uncertainty in West African climate-input

assessment, rather than operational forecast skill. These sources support checking reference support, provenance, coverage, and event labels; they do not remove observational uncertainty.

### S2.4 Information use and feasible responses

Ingram et al. (2002) and Hansen et al. (2011) identify opportunities and constraints for agricultural use of seasonal forecasts. Roncoli et al. (2002, 2009), Roncoli (2006), and Roudier et al. (2014) provide evidence on local knowledge, dissemination, and management response. Lemos et al. (2012) address usability, and Vedeld et al. (2020) address service governance. These sources inform actors, information baselines, and delivery conditions. They do not convert contextual information into a quantitative benchmark unless its forecast or decision rule can be reconstructed.

### S2.5 Types of value evidence

Katz and Murphy (1997), Richardson (2000), Palmer (2002), and Meza et al. (2008) support conditional decision-value analysis. Sultan et al. (2010) and Roudier et al. (2012, 2016) provide ex-ante or model-based agricultural examples in Senegal and Niger. Their outcomes depend on declared model and behavioural assumptions.

Amegnaglo et al. (2017) supply stated-preference evidence from Benin. Sanfo et al. (2022) examine productivity outcomes associated with customised services in Burkina Faso and Ghana. These are different evidence types; neither willingness to pay nor an unqualified participation-outcome association should be relabelled as the causal value of a rainfall forecast. Tall et al. (2018) and Vaughan et al. (2019) support keeping service use, modelled benefit and attributed outcomes separate. Full bibliographic details and DOI links are provided in the main manuscript.

## Appendix S3 Reporting record for agricultural rainfall verification

### S3.1 Agricultural application and calendar

Record crop, management operation, location, rainfall regime, eligible crop-calendar window and action deadline. Identify feasible alternatives, decision authority and resource constraints affecting the operation. State whether the action is one-off, staged or updated and what remains adjustable at issuance. Record the forecast producer and agricultural intermediary where delivery or translation affects the application claim.

### S3.2 Predictand

Record the rainfall variable or index, units, accumulation window, dry-day threshold, duration rule, eligible search dates and event frequency. Document agronomic or calendar evidence for the target. Distinguish seasonal accumulation, regional onset, local onset, candidate-window events and dry-spell duration or frequency. Define absent events, incomplete windows and missing values. Prespecify the primary definition and justified alternatives; a rainfall condition is not observed crop establishment.

### S3.3 Lead and support

Record issue, receipt and readiness-to-act times, the deadline, valid period and update frequency. Distinguish nominal lead, forecast horizon and actionable margin. Check that the product covers the entire future window defining the event. Identify forecast resolution, reference support, advisory area, and action extent. Record accumulation conventions, aggregation, interpolation, calibration, downscaling, and product versions, preserving the forecast as issued.

For future-window onset labels, specify which days were observed at issuance and which were forecast. Verify using subsequent reference data without allowing them to inform the historical trigger. A product unable to supply the remaining event window requires a different justified target or a validated translation model. Its horizon should not be silently extended using verifying observations.

### S3.4 Verification attributes

State the forecast format and diagnostics. For continuous amounts, report magnitude error and bias. For binary events, supply hits, misses, false alarms and correct negatives, event count, trigger count and the decision threshold. Define every rate or ratio, including how zero denominators are handled. For probabilities, report a proper score and diagnostics of calibration or other attributes used to interpret it. Separate the event’s physical threshold from the probability threshold triggering action.

For each score, identify cases, years, issue dates, lead and calendar conditioning. Disclose any restriction of onset-date errors to cases with onset in both forecast and reference, and report absent or censored onsets separately. Compare products on common targets and cases. Report event counts and missingness in each seasonal or regime stratum. Communication assessment is separate from probability calibration.

### S3.5 Benchmarks and uncertainty

Specify benchmark target, support, calendar conditioning, training years, information cut-off, and fitting rule. A climatological benchmark must forecast the same event or quantity in the same eligible window as the product. For onset, record the search dates and no-onset treatment; for changing event frequency, state whether fitting is fixed or updated using only information available before the evaluated issuance. Report absolute and benchmark scores with paired common-case differences. Separately specify the action baseline representing usual practice.

Record reference name, version, quality control, support, daily accumulation convention, and gaps. Recompute event labels for each justified reference and identify common available cases. Record shared inputs and calibration overlap. Assess sensitivity to thresholds, leads, and supports without tuning on test outcomes. Describe dependence-preserving uncertainty procedures, and event counts per stratum. Separate fully observed non-events from outcomes censored by incomplete future windows.

### S3.6 Boundaries of agricultural evidence

State whether the analysis estimates conditional decision loss, simulated agricultural outcomes, preference evidence, observed management response or attributed outcome effects. Specify costs, losses, utility, information costs and units. Describe the counterfactual and assumptions about receipt, interpretation and adoption. For observed effects, explain selection, confounding, spillovers and accompanying interventions and the design used for attribution.

Identify which claim is supported: rainfall prediction, potential loss reduction, simulated impact, preference, observed response or attributed outcome. Where feasible actions or consequences differ, report the relevant user groups. An undocumented condition restricts the claim; neither an incomplete record nor a completed checklist is an overall service-performance score.

## Appendix S4 Constructed numerical illustration

### S4.1 Purpose and status

The example demonstrates that aggregate amount error and one planting-event classification can rank the same deterministic rainfall trajectories differently. It also isolates initial-threshold and designed reference dependence. All values are constructed; there are no stations, observed seasons, forecast issue times, farmer responses or empirical uncertainty estimates. Every

trajectory has the same subsequent wet-day timing and a five-day maximum dry spell. The example therefore tests threshold classification, not the forecasting of dry-spell failure or its occurrence time.

### S4.2 Complete rainfall specification

Each case contains 33 relative days. For the primary constructed reference, the day-1 amounts in cases 1-8 are 21, 21, 21, 21, 19, 19, 19 and 19 mm. Days 2-3 are dry. Rainfall is 6 mm on days 4, 10, 16, 22, and 28, and zero on every other day.

Forecast A's day-1 amounts are 19, 19, 21, 21, 21, 21, 19, and 19 mm. Its other daily amounts match the primary reference. Forecast B's day-1 amounts are 24, 24, 24, 24, 16, 16, 16, and 16 mm. Its rainfall on days 4, 10, 16, 22, and 28 is 9 mm in every case, with zero on the remaining days.

The second constructed reference has the same daily arrays as A. Relative to the primary reference, only day 1 in cases 1-2 and 5-6 changes by minus 2 mm and plus 2 mm, respectively. This perturbation deliberately changes event labels close to the threshold. It is not an independently observed product or a calibrated error model.

### S4.3 Event and action definitions

Equation (6) is evaluated for the candidate window beginning on relative day 1 only. The initial window comprises days 1-3; the subsequent window comprises days 4-33. The initial rainfall threshold is 20 mm; a dry day has rainfall below 1 mm, and the maximum allowed subsequent dry spell is seven days. Every trajectory has a longest subsequent dry spell of five days, so event membership in the primary analysis depends on the initial threshold. The illustration does not evaluate onset-date error, a seasonal onset search, or dry-spell prediction skill.

The action rule plants when the deterministic trajectory satisfies this event and waits otherwise. Equation (11) assigns loss 2 to a false planting signal and loss 1 to a missed opportunity, with zero loss assigned to the two correct-action categories. The units are arbitrary. This deliberately restricted loss function omits planting costs common to successful actions, income, replanting dynamics and resource constraints. The decision baseline always waits. It is not a technical rainfall forecast or an observed usual-practice baseline.

### S4.4 Primary calculations

There are 264 case-days. A has four errors of 2 mm and no other daily errors, giving a squared-error sum of 16 and root mean square error (RMSE) 0.246182982 mm. B has six errors of 3 mm in every case, giving a squared-error sum of 432 and RMSE 1.279204298 mm, using Equation (10).

The primary reference contains four events and four non-events. A yields two hits, two misses, two false alarms, and two correct negatives. B yields four hits, no misses or false alarms, and four correct negatives. Their event-error fractions are 0.50 and 0.00. Under the declared loss function, their mean losses are 0.75 and 0.00, respectively. The always-wait baseline has a loss of 0.50. Loss reduction relative to that baseline is minus 0.25 for A and 0.50 for B.

These deterministic forecasts do not supply probabilities for an empirical reliability analysis. No confidence intervals or hypothesis tests are reported because the construction is not sampled from a population. Exact calculations establish consistency within the example, not uncertainty about operational performance.

### S4.5 Threshold and reference checks

The reproduction script evaluates initial thresholds of 18, 19, 20, 21 and 22 mm for both constructed references. All other event settings remain fixed. Under the primary reference at 18 or 19 mm, all eight cases qualify; A has no event error, and B misses four. At 20 or 21 mm, four cases qualify; A has an event-error fraction of 0.50, and B has 0.00. At 22 mm, none qualify; A has no event error, and B gives four false alarms.

Under the second reference at 20 mm, cases 3-6 qualify. A has four hits and four correct negatives. B has two hits, two misses, two false alarms, and two correct negatives. Their event-error fractions are 0.00 and 0.50, respectively, and their mean losses are 0.00 and 0.75. The always-wait baseline remains 0.50. Changing the reference therefore reverses the event ranking in this deliberately designed boundary case.

### S4.6 Reproduction files

ESM_2.zip (Online Resource 2) supplies onset_illustration.py, constructed_rainfall.csv, illustration_results.csv, rebuild_figures.py, and README.txt. The numerical calculation requires Python 3.9 or later and no external packages. Running python onset_illustration.py

regenerates the CSV files and checks the reported results. The figure script additionally requires Matplotlib and Pillow and renders the figures from the declared specifications and results.

The rainfall file records 264 case-days for the two constructed references and forecasts A and B. The results file contains 20 comparisons of A and B across two references and five initial rainfall thresholds; all other event settings remain fixed.