Weather or Not: Leveraging in Situ Sensors to Improve Measurement of Weather, Climate Variability, and Their Links to Socioeconomic Outcomes in Low-Income Countries

Adriana Paolantonio, **Anna Josephson**, & Emanuele Clemente,

with Rodrigo Guerra Su, Jeffrey Michler, John Ilukor

Better Data for Better Jobs & Lives: Innovations in Survey Measurement in the Age of AI

Washington DC

8 December 2025

Motivation

- **Accurate agricultural production statistics depend on reliable weather data**, especially in rainfed systems.
  - Climate change intensifies intraseasonal extremes, making precise measurement and integration of weather conditions critical for reliable estimates of agricultural production.
- **Errors in weather measurement can distort agricultural statistics**, undermining evidence used to track progress towards Sustainable Development Goals.
- **Africa's weather observation system is sparce and below standard**, creating a data poor environment and motivating the of alternative data (e.g., low-cost weather stations, remotely sensed Earth Observation (RSEO) data).

Objectives

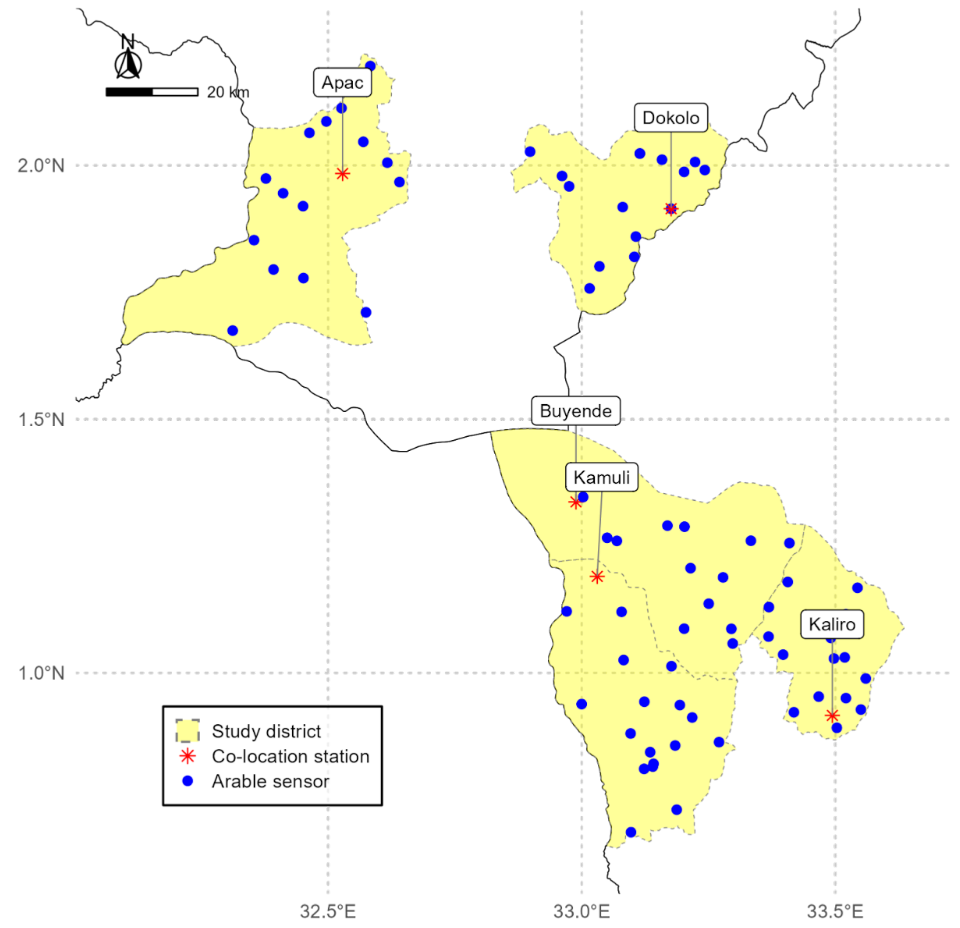
Objective 1: Investigate differences between weather data sources, including:

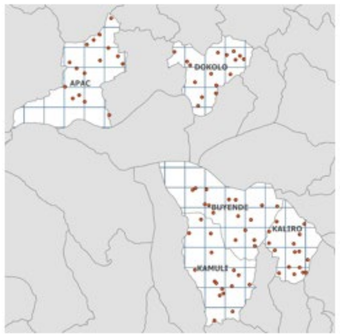
1. Professional automatic weather stations, operated by the Ugandan Department of Meteorology (DoM): ***ADCON/UNMA***
2. Low-cost in situ weather stations produced by Arable: ***Arables***
3. Remote sensing Earth Observation data sets: ***RSEO***

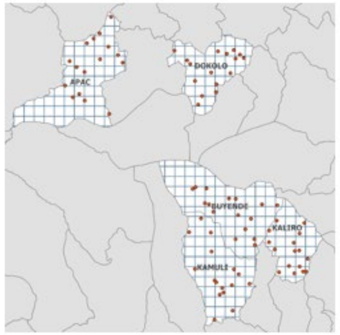
Objective 2: Evaluate the extent to which these differences impact the estimation of agricultural production for maize-producing households.

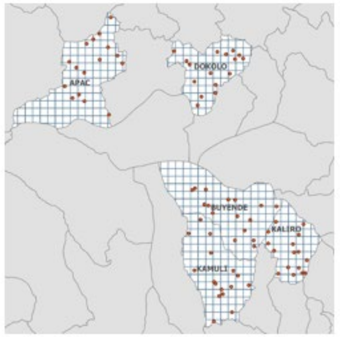
Data + Deployment

- 80 Arable Mark2 weather sensors deployed across five districts in Uganda
    - 5 co-located with professional ADCON AWSs operated by Uganda DoM
    - 75 installed on farms of designated model farmers in randomly selected survey rural EAs
- Arable data paired with ADCON data and RSEO datasets: ARC2, CHIRPS v2.0, TAMSAT
- Survey conducted in 3 rounds with 900 maize-producing HHs (1 maize plot randomly selected per HH)
    - Complete data available for 888 HHs/plots




ARC2 (~11 km)


CHIRPS-2.0 (~5 km)


TAMSAT (~3.75 km)

Analyses and Alignment with Objectives

🌧️ RAINFALL MEASUREMENT VALIDATION: Objective 1

- Sources:
  - ADCON professional AWSs (***benchmark***)
  - Arable Mark2 in situ weather sensors
  - RSEO datasets: ARC2, CHIRPS, TAMSAT
- Aggregation: daily / dekadal / monthly
- Level of analysis: co-location sites (4 sites, Kaliro excluded)
- Matching of RSEO data using GPS coordinates of Arable devices
- Timeframe: September 2023 – June 2024
- Methods: performance metrics, regression analysis

→

🌽 EFFECTS ON MAIZE YIELD: Objective 2

- Sources:
  - Arable Mark2 in situ weather sensors
  - RSEO datasets: ARC2, CHIRPS, TAMSAT
  - HH survey
- Aggregation: 20 derived rainfall indicators
- Level of analysis: plots (777 plots)
- Matching of weather data:
  - Arable: EA-assigned / plot-nearest / IDW
  - RSEO: using plot GPS coordinates
- Timeframe: August 2023 – January 2024
  - (1) FAO standard seasons and (2) individualized seasons based on planting and harvest dates
- Methods: regression analysis

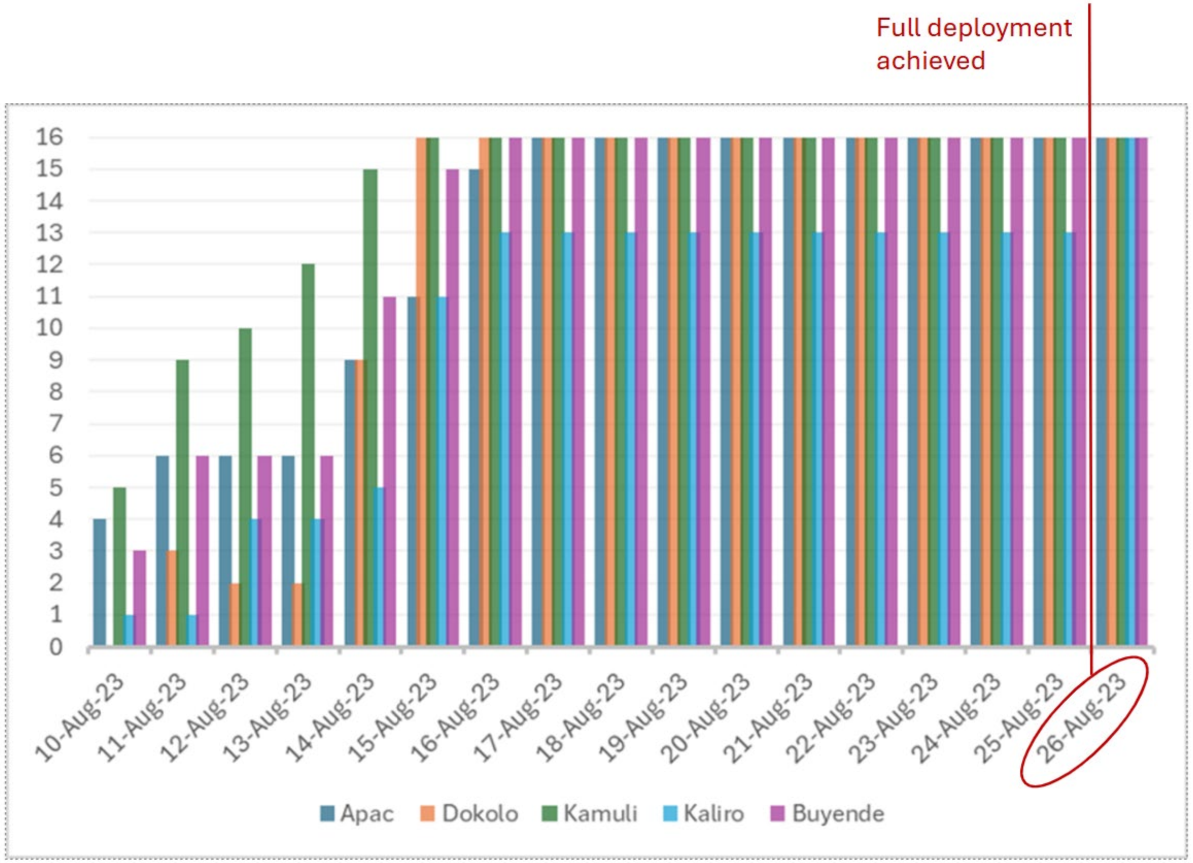
Household-reported Planting Completion Dates

Week finished planting	# HHs/plots	% HHs/plots
.	11	1.24
Jul, 1st week	23	2.59
Jul, 2nd week	30	3.38
Jul, 3rd week	21	2.36
Jul, 4th week	26	2.93
Aug, 1st week	125	14.08
Aug, 2nd week*	117	13.18
Aug, 3rd week	81	9.12
Aug, 4th week**	80	9.01
Sep, 1st week	174	19.59
Sep, 2nd week	114	12.84
Sep, 3rd week	38	4.28
Sep, 4th week	35	3.94
Oct, 1st week	9	1.01
Oct, 2nd week	4	0.45
Total	888	100

*Arable deployment started

**Arable deployment completed

- 47% finished planting *before* full Arable deployment (Aug 26 – 4th week)
  - 26% finished planting before deployment even began (Aug 10 – 2nd week)



Adaptations for Missing Data

1. Missing data before 26 August
    - Reduce the sample: only include HHs who finished planting after 26 August
        - "Small sample": 374 households (won't show these results, but they are the same)
    - Impute Arable rainfall data between 10 August and 26 August
        - Use k-Nearest Neighbor (kNN) and inverse distanced weighting (IDW)
    - Adjust planting dates: move 3 August planters to 10 August

2. Missing data due to Arable sensor malfunctions
    - If had at least 4 hours of data: linear interpolation
    - < 4 hours of data: use kNN and IDW imputation

Samples for Yield Analysis

Week finished planting	# HHs/plots	% HHs/plots		
.	11	1.24	Excluded	
Jul, 1st week	23	2.59		
Jul, 2nd week	30	3.38		
Jul, 3rd week	21	2.36		
Jul, 4th week	26	2.93		
Aug, 1st week	125	14.08	Shifted to Aug, 2nd week Arable rainfall data imputed	**Full sample** (N=777)
Aug, 2nd week*	117	13.18		
Aug, 3rd week	81	9.12		
Aug, 4th week**	80	9.01		
Sep, 1st week	174	19.59	None >> **Small sample** (N=374)	
Sep, 2nd week	114	12.84		
Sep, 3rd week	38	4.28		
Sep, 4th week	35	3.94		
Oct, 1st week	9	1.01		
Oct, 2nd week	4	0.45		
Total	888	100		

*Arable deployment started

**Arable deployment completed

Validation Methods

1 Continuous and categorical performance metrics

(Ahmed et al., 2024; Wodebo et al., 2025)

Metric class	Statistical metric	Formula	Perfect match
Continuous	Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(BM_t - Pr_t)^2}$	0
	Correlation Coefficient (R)	$R = \frac{\sum_{t=1}^{n}(BM_t - \overline{BM})(Pr_t - \overline{Pr})}{\sqrt{\sum_{t=1}^{n}(BM_t - \overline{BM})^2}\sqrt{\sum_{t=1}^{n}(Pr_t - \overline{Pr})^2}}$	1
	Coefficients of Variation ratio (CV)	$CV = \frac{\left(\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Pr_t - \overline{Pr})^2}\right) / \left(\frac{1}{n}\sum_{t=1}^{n} Pr_t\right)}{\left(\sqrt{\frac{1}{n}\sum_{t=1}^{n}(BM_t - \overline{BM})^2}\right) / \left(\frac{1}{n}\sum_{t=1}^{n} BM_t\right)}$	1
Categorical	Probability of detection (POD)	$POD = \frac{A}{A + C}$	1
	False alarm ratio (FAR)	$FAR = \frac{B}{A + B}$	0
	Frequency of bias index (FBI)	$FBI = \frac{A + B}{A + C}$	1
	Critical success index (CSI)	$CSI = \frac{A}{(A + C + B)}$	1

Note: in the formulas for continuous metrics t indexes the temporal resolution of the measurements that are compared (i.e., daily, dekadal, monthly). The number of observations (n) is 1,212 for daily measurements, 156 for dekadal measurements, and 40 for monthly measurements.

Contingency matrix for categorical performance metrics calculation

Evaluated product	Benchmark (ADCON)		
	Yes	No	Total
Yes	Hit (A)	False Alarm (B)	A+B
No	Miss (C)	Correct Negative (D)	C+D
Total	A+C	B+D	A+B+C+D

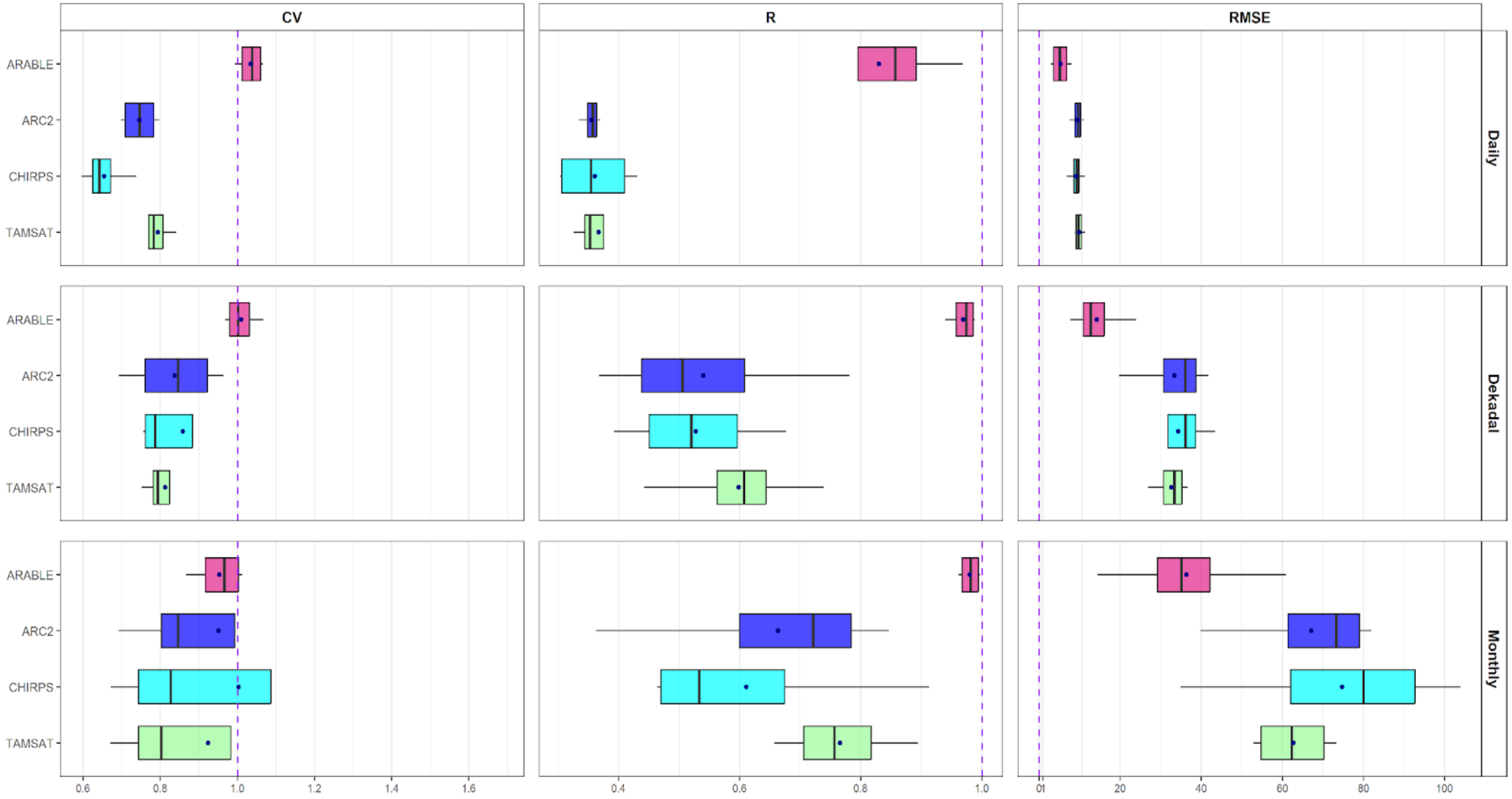
2 Regression (Carson and Yu, 2020) (won't show these)

$$BM_t = \alpha + \beta Pr_t + \varepsilon_t \qquad [EQ1]$$

where BM_t is rainfall measured by the benchmark measurement instruments (ADCON stations) at time t (daily / dekadal / monthly), and Pr_t is the corresponding rainfall estimate from the evaluated product (Arable, ARC2, CHIRPS, TAMSAT)

3 t-tests of differences in means between Arable and RSEO products for derived rainfall indicators across all sites (won't show these)

Validation Results (1/2)

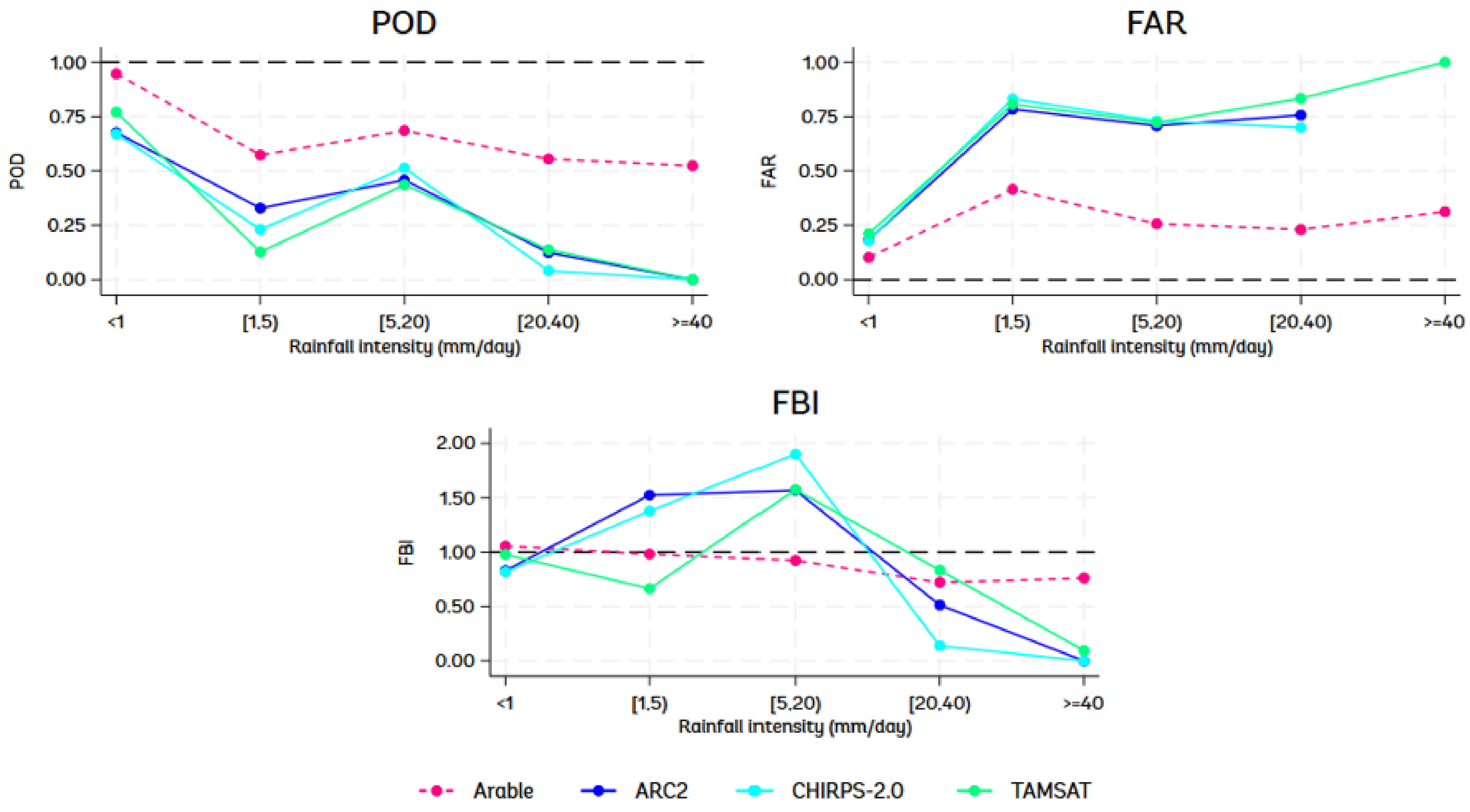


In each box-plot, the dashed purple line marks the ideal metric value, the box's ends correspond to the 25th and 75th percentiles of a product's performance across the co-location sites, the thick line inside indicates the median, and the dark blue dot is the mean.

Reference period: September 2023 – June 2024.

- Strong agreement of Arable with ADCON (benchmark) data across all time scales:
  - CV ratio remains very close to 1
  - R consistently high and close to 1, especially at the dekadal and monthly level
  - RMSE remains relatively low
- Arable outperforms RSEO products across all evaluation metrics and temporal scales
- Arable shows a more robust and stable performance, with mean and median values always very close, and often overlapping

Validation Results (2/2)



- Arable shows highest POD, lowest FAR values, and FBI ≈ 1across all intensity bins
- RSEO products struggle to detect heavy rainfall events (ARC2 and CHIRPS missed all violent events)
- FBI patterns show that RSEO tends to overreport light to moderate events and underreport heavy-violent events

Reference period: September 2023 – June 2024.

Yield Methods

- Following Deschênes and Greenstone (2007):

$$Y_{ih} = \alpha + \beta_j W_{ih}^{(j)} + \mu_{d(i,h)} + \varepsilon_{ih}$$

[EQ2: Model 'no inputs']

where Y_{ih} is maize yield (kilograms per hectare) on plot i for household h, $W_{ih}^{(j)}$ is rainfall-based indicator j with β_j associated coefficients and $\mu_{d(i,h)}$ is a district fixed effect for the district d where plot i/household h is located

- We also estimate an alternative specification that adds a vector of plot-household input variables (X_{ih}):

$$Y_{ih} = \alpha + \beta_j W_{ih}^{(j)} + X_{ih}' \pi + \mu_{d(i,h)} + \varepsilon_{ih}$$

[EQ3: Model 'inputs']

- Method akin to Michler at al. (2022) & Josephson et al. (2026)
- We estimate *EQ2* and *EQ3* one rainfall indicator at a time; we consider 20 rainfall indicators, each constructed using six weather data sources/methods: (i) Arable EA-assigned (ii) Arable plot-nearest, (iii) Arable IDW, and RSEO products—(iv) ARC2, (v) CHIRPS, and (vi) TAMSAT
- Both include soil traits and geo-variables.

Derived Rainfall Indicators

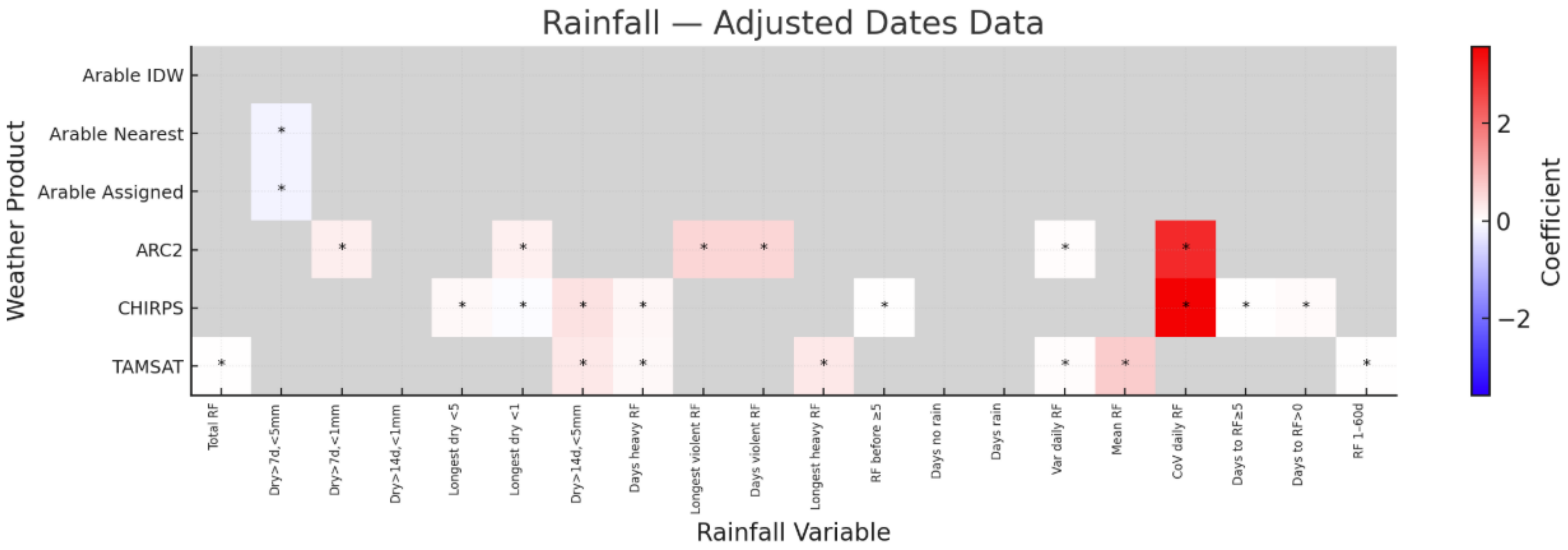
Indicator	Definition
Total seasonal rainfall	Cumulative daily rainfall over the rainy season*
Mean of daily rainfall	The first moment of the daily rainfall distribution over the rainy season
Number of days with rain	Total number of days with ≥ 1 mm of rainfall during the rainy season
Number of days without rain	Total number of days with < 1 mm of rainfall during the rainy season
Variance of daily rainfall	The second moment of the daily rainfall distribution over the rainy season
Coefficient of variation of daily rainfall	The ratio of the standard deviation to the mean of daily rainfall during the rainy season
Longest dry spell (1 mm)	Maximum number of consecutive days with < 1 mm of rainfall during the rainy season
Longest dry spell (5 mm)	Maximum number of consecutive days with < 5 mm of rainfall during the rainy season
Number of dry spells ≥ 7 days (< 1 mm)	Number of periods with at least 7 consecutive days of < 1 mm rainfall during the rainy season
Number of dry spells ≥ 7 days (< 5 mm)	Number of periods with at least 7 consecutive days of < 5 mm rainfall during the rainy season
Number of dry spells ≥ 14 days (< 1 mm)	Number of periods with at least 14 consecutive days of < 1 mm rainfall during the rainy season
Number of dry spells ≥ 14 days (< 5 mm)	Number of periods with at least 14 consecutive days of < 5 mm rainfall during the rainy season
Longest heavy rain spell [20,40 mm)	Maximum number of consecutive days with daily rainfall between 20 mm and < 40 mm during the rainy season
Longest violent rain spell (>= 40 mm)	Maximum number of consecutive days with daily rainfall ≥ 40 mm during the rainy season
Total number of days with heavy rain [20,40 mm)	Total number of days with daily rainfall between 20 mm and < 40 mm during the rainy season
Total number of days with violent rain (>= 40 mm)	Total number of days with daily rainfall ≥ 40 mm during the rainy season
Total rainfall in first 60 days after planting	Cumulative daily rainfall in the first 60 days following planting**
Days until first rain (> 0 mm)	Number of days from planting until the first day with rainfall > 0 mm
Days until first rain (≥ 5 mm)	Number of days from planting until the first day with rainfall ≥ 5 mm
Total rainfall before first rain (≥ 5 mm)	Cumulative daily rainfall until the first day with rainfall ≥ 5 mm

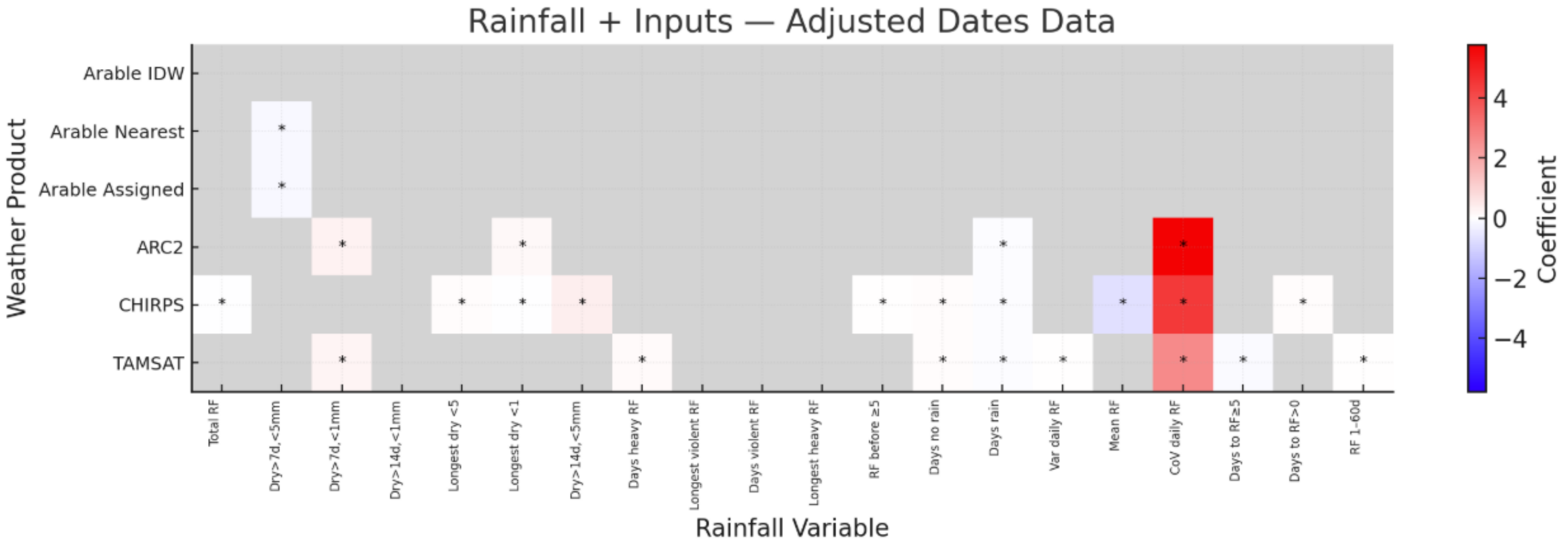
*The rainy season is defined as 26 August 2023–31 January 2024 for the small sample and 10 August 2023–31 January 2024 for the full sample.

**The reference point for the post-planting period is the household-reported planting completion date on the plot.

Yield Results (1/2)

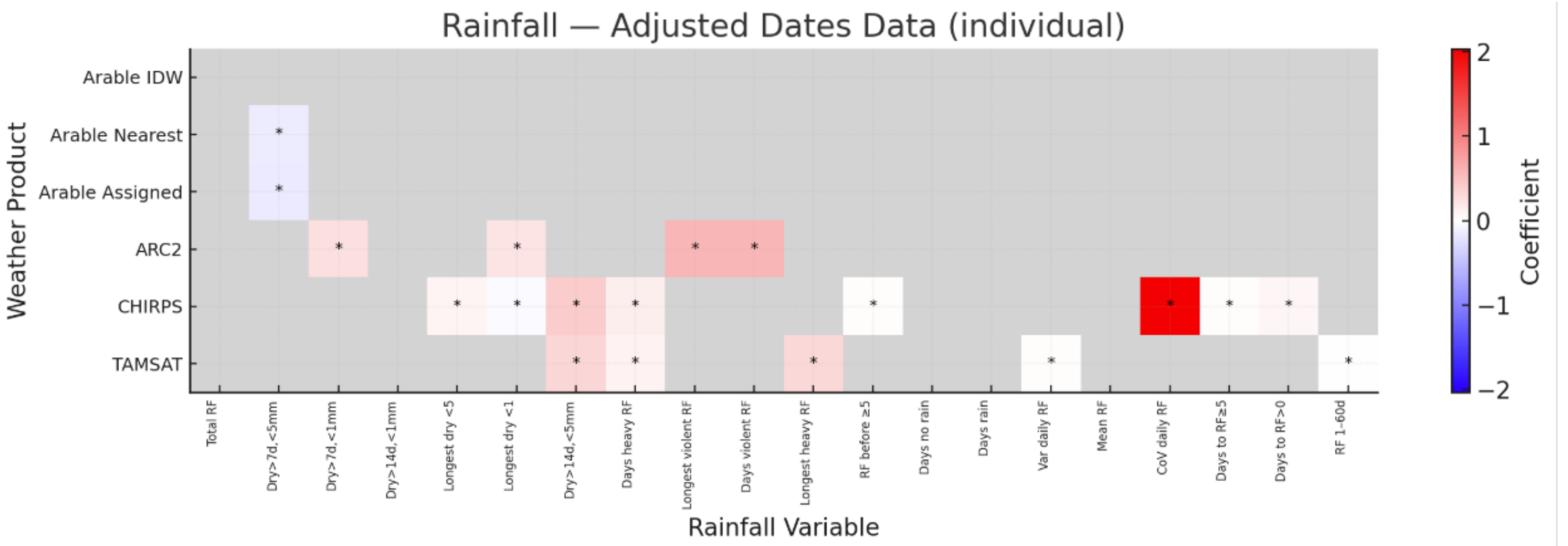
- Looking across all rainfall variables and products, lack of significance is evident for Arable-derived indicators
  - Possible explanation: uniformly wet 2023 second season reduced variability, muting detectable yield–rainfall effects for in-situ devices
  - But maybe we are missing effects, in aggregating to a local season
- More significant results (very close to 0) for RSEO products

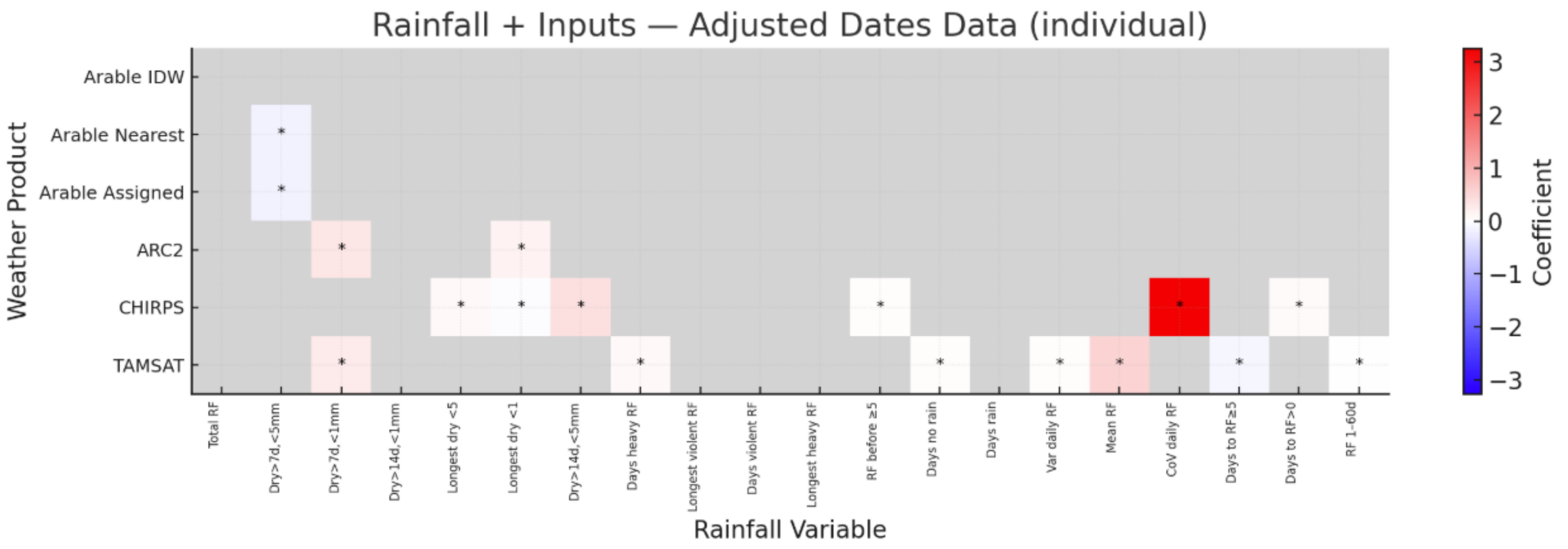




Yield Results (2/2)

- Turning to our findings for individual seasons:
  - Individual seasons look at planting date (with our adjustments) and reported harvest date.
- These results are very similar, with even less significance: it suggests that our RSEO results are brought closer (in terms of significance) to the Arables.
- Persistent positive result in Arables related to dry spells that is somewhat puzzling.





Limitations

- Timing: Single agricultural season
- Data: Limited set of rainfall indicators
- ADCON:
  - Reliance on ADCON for comparison, assumes uniform performance across all sites
  - Limited number of co-location sites
  - Short period of overlapping ADCON data

Conclusions

- **Rainfall measurement matters.** Comparing Arable sensors, professional stations, and remote sensing products reveals differences in rainfall estimates and their impact on maize yields
- **Different tools, different strengths.** Arable devices give consistent and precise (though not always significant) rainfall to yield links, while RSEO data sometimes show significance but with instability and wide errors due to issues like resolution and interpolation
- **Implications for research and policy.** Relying only on satellite data risks bias and misdirected policies; Arable sensors are best used as complementary ground-truthing tools rather than substitutes
- **Looking forward.** As climate risks rise, better local data are essential, but their usefulness depends on context, cost-effectiveness, and realistic expectations about how much better measurement improves decision-making

Extensions and Next Work

Changes to Current Version

- Yield regressions with multiple rainfall variables
- Interactions between rainfall indicators and planting dates
- Inclusion of different geovariables to specifications

Possible Extensions to Future Work

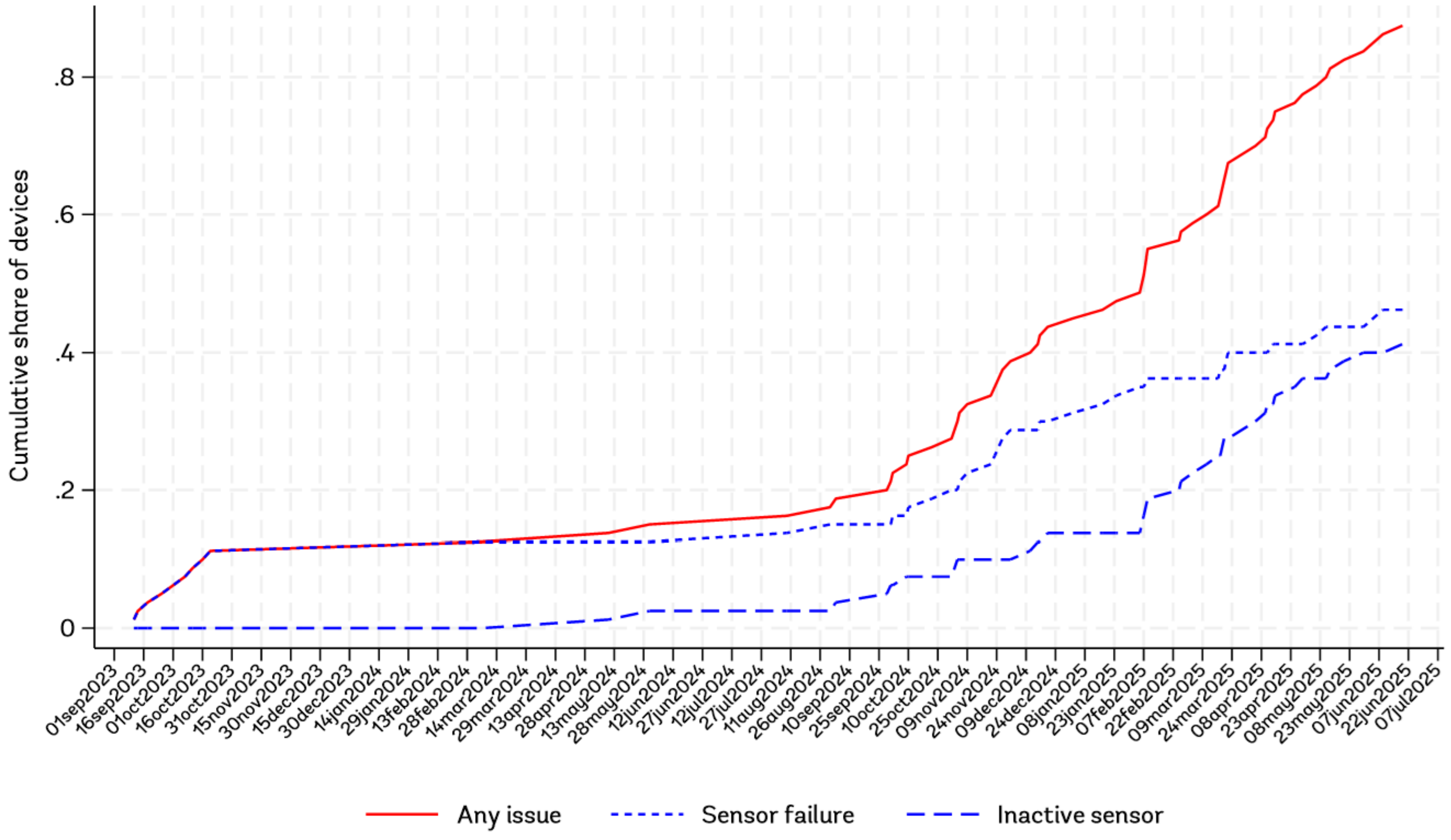
- Extend validation analysis to full Arable data collection period: August 2023 – July 2025 (will need to develop a protocol to handle rising sensor malfunctions)
- Bring in temperature data
- Cost-benefit/Cost-effectiveness analysis, combining with Arable performance data
- Parameter adjustments for RSEO products, based on ground-truth
- Optimizing weather sensors placement for enhanced groundtruthing of RSEO products

Questions? Comments?

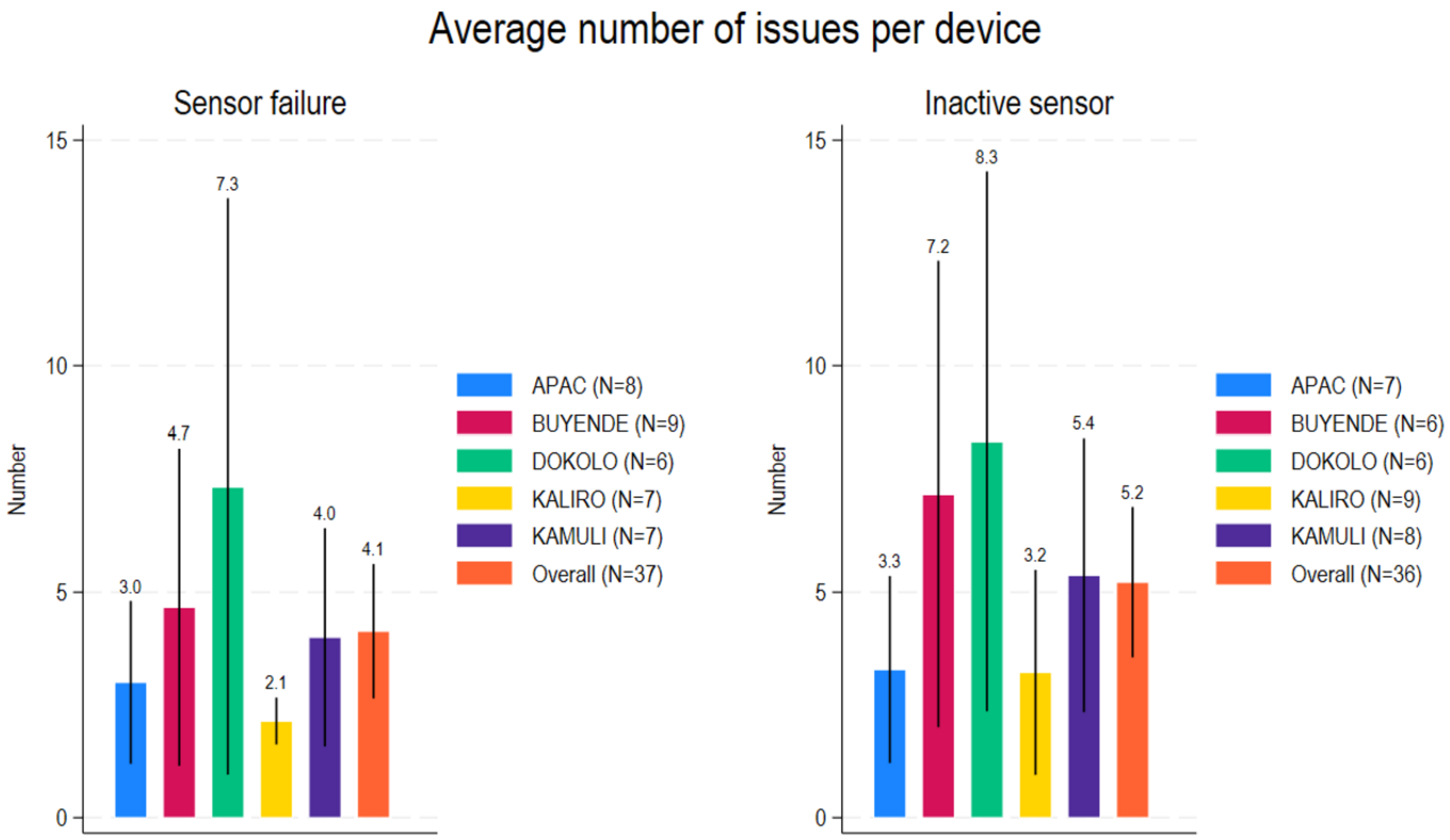
aljosephson@arizona.edu

apaolantonio@worldbank.org

Extras: Arable Mark2 Performance for the period September 2023 – June 2025

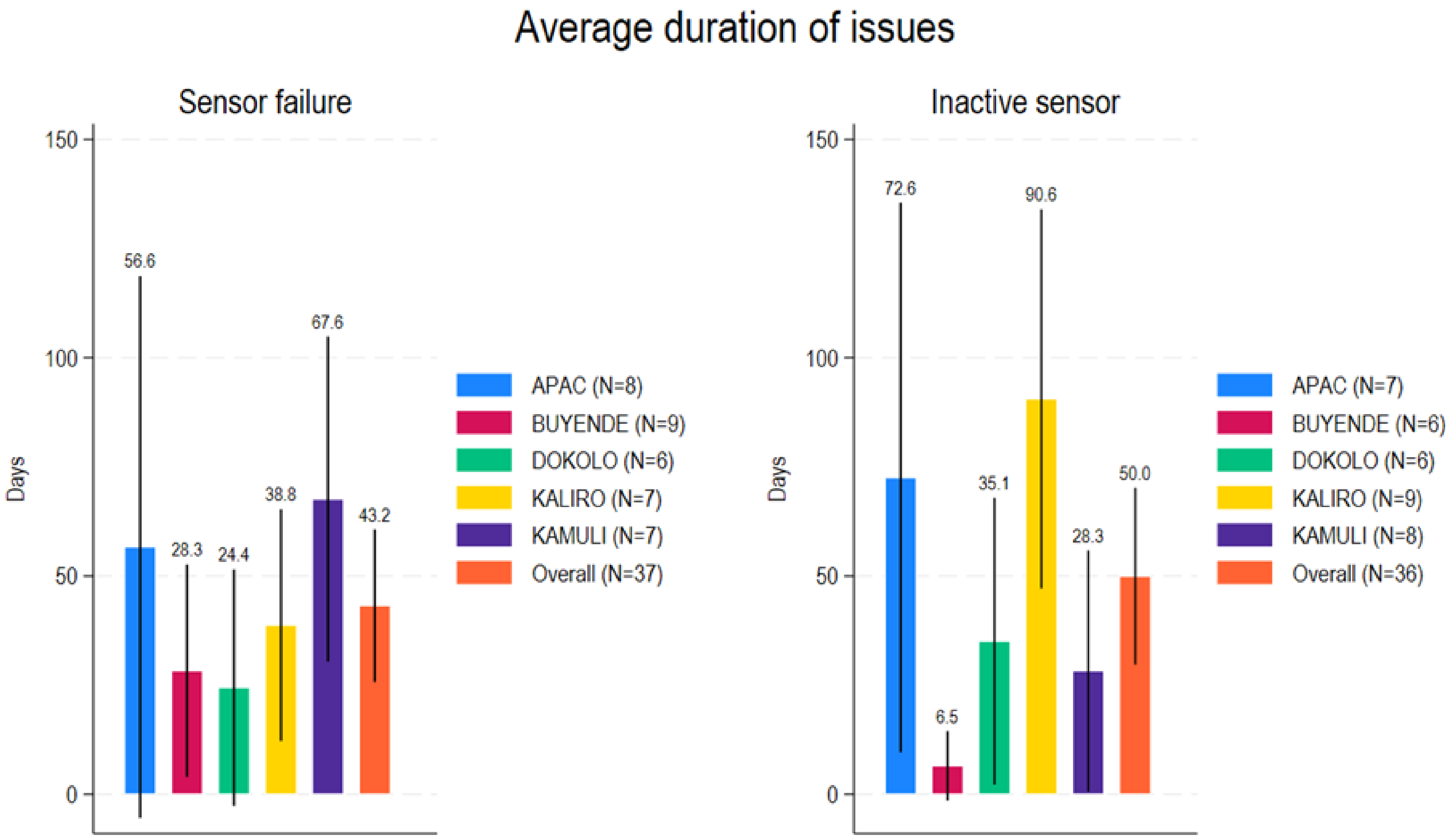


- Overall failures (red line) increased steadily from ~10% to >85%, with sharp acceleration in late 2024
- Sensor failures (blue dotted line) emerged within weeks of installation (rainfall measurement unaffected)
- Inactive sensors (blue dashed line) became a serious issue about one year after deployment



- Sensor failures: range from 2.1 (Kaliro) to 7.3 (Dokolo); overall mean: 4.1
- Inactive sensor incidents: range from 3.2 (Kaliro) to 8.3 (Dokolo); overall mean: 5.2
- Intermediate levels observed in Apac, Buyende, and Kamuli

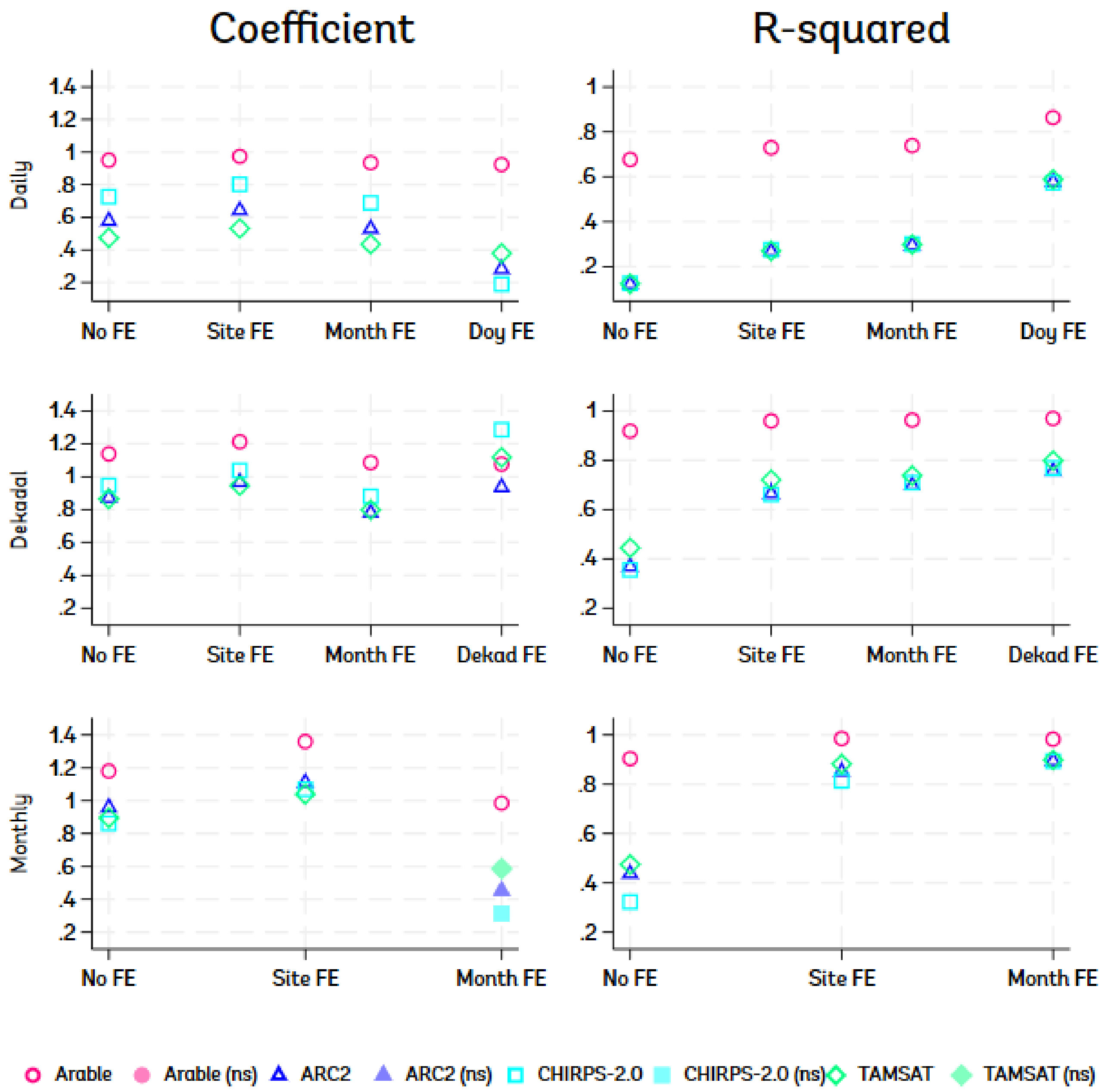
Average duration of issues



- Sensor failure duration: longest in Kamuli (67.6 days), followed by Apac (56.6), Kaliro (38.8), Buyende (28.3), and Dokolo (24.4); overall mean: 43.2 days
- Inactive sensor events: longest in Kaliro (90.6 days) and Apac (72.6), shortest in Buyende (6.5); overall mean: 50.0 days

Validation Results (3/3)

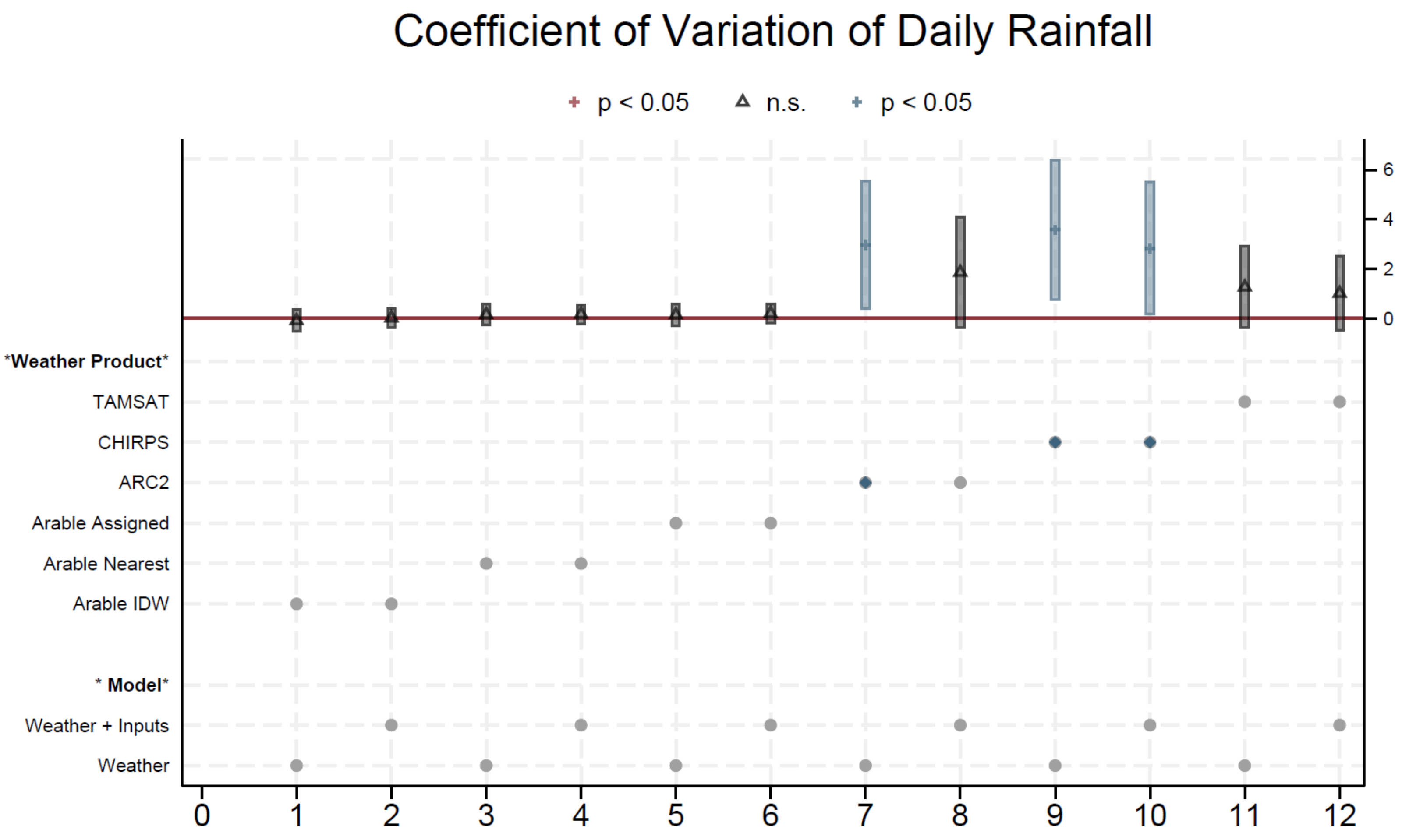
- Arable: Regression coefficients near 1 with high R^2 across all scales; inclusion of site or time FE <u>makes little difference</u>, suggesting Arable already captures spatial/temporal variation
- RSEO products: At daily scale, coefficients <1 and low R^2 (<0.5); site FE reduce location bias, and time FE raise R^2 but increase downward bias. Performance improves notably at dekadal scale
- Monthly scale: RSEO shows high R^2 when controlling for site or time FE but coefficients fall below 1 and lose significance with month FE (suggests greater underestimnation of monthly totals but limited sample size at this scale cautions interpretation)



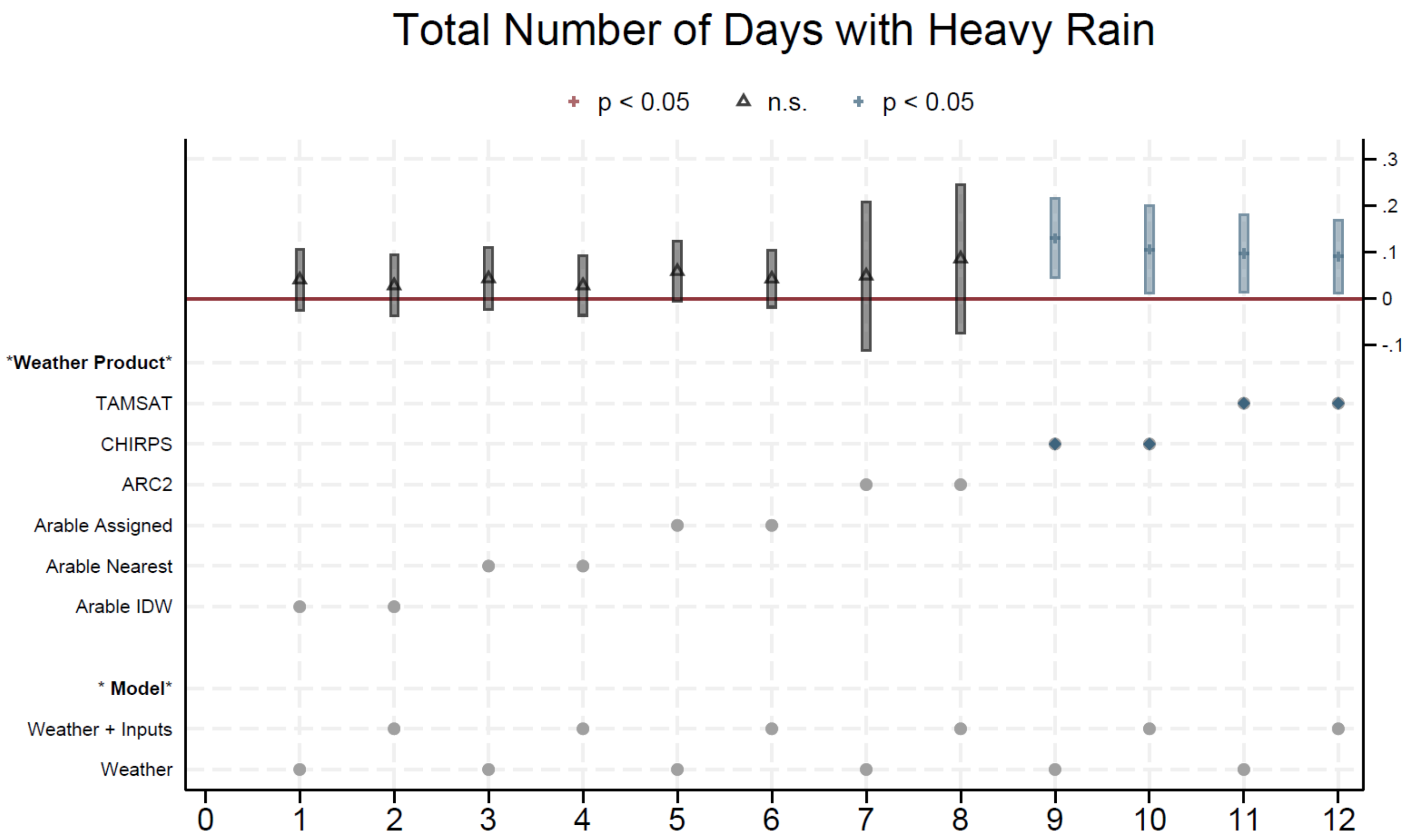
Reference period: September 2023 – June 2024. N = 1,212 for daily; 156 for dekadal; 40 for monthly.

Yield Results

- Three Arable products: smaller coefficients & more precisely estimated, but all insignificant
- Some variation in remotely sensed products:
  - CHIRPS both significant; one TAMSAT significant
  - All larger confidence intervals



Yield Results



- *Same results!*
- Three Arable products: smaller coefficients & more precisely estimated, but all insignificant
  - Confidence intervals larger than CoV
- Some variation in remotely sensed products:
  - CHIRPS and TAMSAT both significant
  - All larger confidence intervals