CCE vs District-Level APY Coherence & Quality Report

Crop Cutting Experiment Yield Validation Against DES District-Level Data

Chapter 1: CCE Dataset — Structure & Quality Assessment

Record Schema, Geographic/Crop Coverage, Missing Data, Outlier Detection

Chapter 2: CCE vs District APY — Static Coherence

Yield Deviation, Crop-wise & State-wise Breakdown, Agreement Classification

Chapter 3: Temporal, Seasonal & Sample-Size Analysis

Year-wise Trends, Kharif vs Rabi Test, Sample-Size Effect

Chapter 4: Spatial Multiplier Comparison

CCE vs APY District Performance Relative to National Baseline

Chapter 5: Conversion Factor Validation

Observed vs Expected Form-Difference Ratios (Rice, Cotton, Maize, Groundnut)

CCE Data Source:	GCES Crop Cutting Experiments (2023–2024)
APY Data Source:	DES District Data (2023-24 and 2024-25)
Matched Observations:	1,297 district-crop-season-year groups
Report Date:	February 2026

Contents

1 Executive Summary **2**

2 Chapter 1: CCE Dataset — Structure & Quality Assessment **3**
- 2.1 What the CCE Data Contains . . . 3
- 2.2 Record Counts and Temporal Coverage . . . 3
- 2.3 Geographic Coverage . . . 4
- 2.4 Crop Coverage . . . 5
- 2.5 Yield Distribution and Descriptive Statistics . . . 6
- 2.6 Missing Data Analysis . . . 6
- 2.7 Irrigation Status . . . 7
- 2.8 Crop × Season Yield Matrix . . . 8
- 2.9 Chapter 1 Summary . . . 8

3 Chapter 2: CCE vs District APY - Static Coherence Analysis **9**
- 3.1 Methodology . . . 9
  - 3.1.1 Notation and Definitions . . . 9
  - 3.1.2 Step 1: CCE Aggregation to District Level . . . 9
  - 3.1.3 Step 2: Deviation Metrics . . . 9
  - 3.1.4 Step 3: Agreement Classification . . . 10
  - 3.1.5 Step 4: Area-Weighted Aggregation . . . 10
  - 3.1.6 Step 5: Correlation Analysis . . . 10
- 3.2 Matching Funnel . . . 10
- 3.3 Overall Coherence Results . . . 11
- 3.4 Crop-wise Coherence . . . 12
- 3.5 State-wise Coherence . . . 14

4 Chapter 3: Temporal, Seasonal & Sample-Size Analysis **16**
- 4.1 Year-wise Comparison . . . 16
- 4.2 Season-wise Comparison . . . 16
  - 4.2.1 Methodology: Weighted Welch's t-test . . . 16
- 4.3 Sample Size Effect . . . 16

5 Chapter 4: Spatial Multiplier Comparison **18**
- 5.1 Methodology . . . 18
  - 5.1.1 National Base Yield . . . 18
  - 5.1.2 District Multipliers . . . 18
  - 5.1.3 Classification . . . 18
- 5.2 Results . . . 18

6 Chapter 5: Conversion Factor Validation **20**
- 6.1 Methodology . . . 20
- 6.2 Results . . . 20

7 Summary and Recommendations **23**

1 Executive Summary

This report presents a rigorous statistical comparison between **Crop Cutting Experiment (CCE)** plot-level yield measurements and **DES District-Level APY** (Area, Production, Yield) administrative data. By aggregating 119,372 CCE field experiments to district level and matching them against DES district records, we obtained **1,297 matched observations** across 17 states, 27 crops, and 2 fiscal years.

Key Findings:

- **Area-weighted mean deviation:** +23.5% (driven by Rice and Cotton form-differences)
- **50.9%** of cropped area shows agreement within ±10%; **64.8%** within ±20%
- Pearson correlation $r = 0.67$; Spearman $\rho = 0.76$ across all crop-district pairs
- **Kharif crops deviate significantly more** than Rabi (weighted Welch's $t = 7.33$, $p < 0.0001$)
- **Best-performing states:** Haryana (+8.8%), Rajasthan (+11.4%), Gujarat (+15.1%)

Important:

Crops with Known Form-Differences
The following crops are measured in a different physical form in CCE experiments versus APY reporting. This creates a systematic and expected deviation that is **not** a data quality issue:

- **Rice:** CCE measures *paddy* (unmilled); some states report *milled rice* in APY. Expected ratio ≈ 1.49 (milling recovery ∼67%)
- **Cotton:** CCE measures *kapas* (seed cotton); APY reports *lint* (fibre). Expected ratio ≈ 2.86 (ginning outturn ∼35%)
- **Maize:** Some CCEs weigh *cob*; APY reports *shelled grain*. Expected ratio ≈ 1.25 (shelling recovery ∼80%)
- **Groundnut:** Some CCEs weigh *in-shell pods*; some APY reports as *kernel*. Expected ratio ≈ 1.43 (extraction ∼70%)

2 Chapter 1: CCE Dataset — Structure & Quality Assessment

This chapter presents a detailed examination of the Crop Cutting Experiment (CCE) dataset sourced from the GCES (General Crop Estimation Survey) programme. The CCE data represents **plot-level field measurements** of crop yield conducted by trained enumerators at randomly selected farm plots across India.

2.1 What the CCE Data Contains

Each CCE record represents a single crop cutting experiment at a specific farm plot. Table 1 summarizes the key data fields.

Table 1: Key Fields in the CCE Dataset

Field	Unit / Values	Description
Year	2023, 2024	Calendar year of experiment
Season	Kharif / Rabi / Late Kharif	Cropping season
Crop Name	49 unique crops	e.g., Rice, Wheat, Maize, Chickpea
Yield	**Kg/Ha**	Measured dry yield from the experiment plot
Green Weight	Kg	Fresh (green) biomass weight at harvest
Irrigated	Yes / No	Whether the plot was irrigated
Location (GPS)	Lat/Lon (decimal degrees)	GPS coordinates of the experiment plot
State / District / Tehsil / Village	Names + LGD codes	Administrative hierarchy of the plot location

Note:

Yield Unit: CCE yield is recorded in **Kg/Ha** — the same unit used by DES district-level APY data. No yield unit conversion is required between the two datasets.

2.2 Record Counts and Temporal Coverage

Table 2: CCE Dataset — Record Summary

Metric	Value
Total raw records	173,947
Records with non-null yield	129,210 (74.3%)
Records with null/empty yield	44,737 (25.7%)
Records after QC ($10 \leq Y \leq 100{,}000$ Kg/Ha)	128,311
Outliers removed ($Y > 100{,}000$ or $Y < 10$)	899
Calendar year 2023	76,900 (44.2%)
Calendar year 2024	97,047 (55.8%)
Unique states	21
Unique districts	496
Unique tehsils	3,661
Unique villages	56,070
Unique crops	49

Table 3: CCE Records by Year and Season

Season	2023	2024	Total
Kharif	30,266	33,700	63,966
Rabi	32,921	31,197	64,118
Late Kharif	233	442	675
Zaid / Other	253	247	500
Empty	*13,227*	*31,461*	*44,688*
Total	**76,900**	**97,047**	**173,947**

2.3 Geographic Coverage

Table 4: CCE Records by State (Top 15 by record count)

Rank	State	Records	% of Total	Districts
1	Gujarat	31,635	18.2%	33
2	Odisha	26,931	15.5%	30
3	Uttar Pradesh	24,004	13.8%	75
4	Bihar	13,923	8.0%	38
5	Rajasthan	12,927	7.4%	33
6	Telangana	11,140	6.4%	33
7	Uttarakhand	7,506	4.3%	13
8	Punjab	6,240	3.6%	23
9	Haryana	6,072	3.5%	22
10	Chhattisgarh	5,676	3.3%	28
11	Tamil Nadu	5,552	3.2%	32
12	Jharkhand	4,327	2.5%	24
13	Maharashtra	4,217	2.4%	33
14	Kerala	3,883	2.2%	14
15	Assam	3,570	2.1%	33
	Others (6 states)	6,344	3.6%	—
	Total	**173,947**	**100%**	**496**

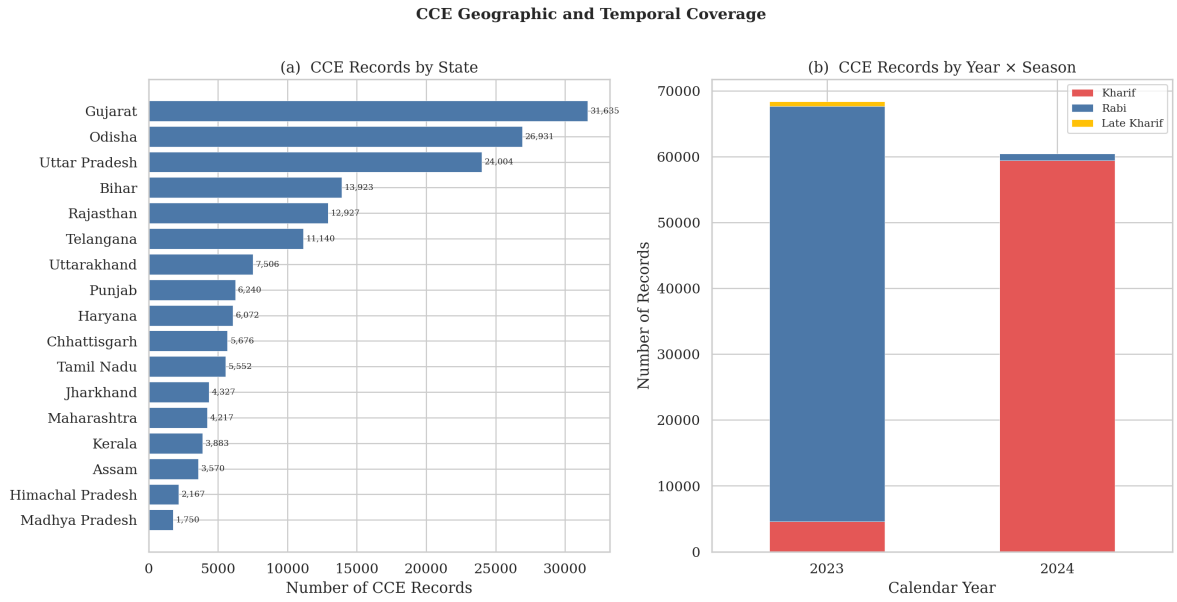

Figure 1: (a) CCE records by state. Gujarat, Odisha, and Uttar Pradesh together account for 47.5% of all experiments. (b) Year × season breakdown showing approximately equal Kharif–Rabi split.

2.4 Crop Coverage

Table 5: CCE Records by Crop (Top 15 of 49 unique crops)

Rank	CCE Crop Name	Records	% of Valid	APY Equivalent
1	Rice	29,251	22.6%	Rice
2	Wheat	26,990	20.9%	Wheat
3	Maize	8,113	6.3%	Maize
4	Chickpea	7,037	5.4%	Gram
5	Indian Mustard	6,356	4.9%	Rapeseed & Mustard
6	Tetraploid Cotton	6,136	4.7%	Cotton
7	Groundnut	4,575	3.5%	Groundnut
8	Pearl Millet	3,952	3.1%	Bajra
9	Urdbean	3,126	2.4%	Urad
10	Mungbean	3,116	2.4%	Moong
11	Soybean	2,714	2.1%	Soybean
12	Sesame	2,659	2.1%	Sesamum
13	Barley	2,359	1.8%	Barley
14	Potato	2,257	1.7%	— (no APY match)
15	Sorghum	1,980	1.5%	Jowar
	34 other crops	17,326	13.4%	—

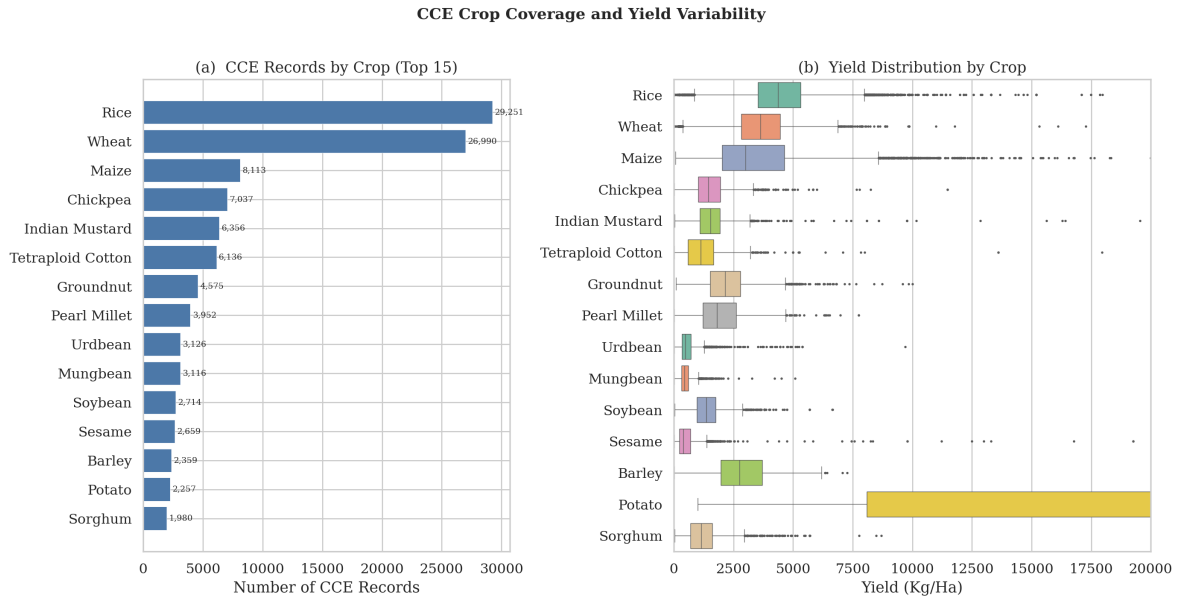

Figure 2: (a) CCE records by crop — Rice and Wheat dominate at 43.5% combined. (b) Yield distribution by crop: note the different yield ranges across crops (Sugarcane at ∼60,000 Kg/Ha vs pulses at ∼500–1,500 Kg/Ha).

2.5 Yield Distribution and Descriptive Statistics

Table 6: CCE Yield Descriptive Statistics (after QC filter: $10 \leq Y \leq 100{,}000$ Kg/Ha)

Statistic	**Value (Kg/Ha)**
Valid records	128,311
Mean	3,778
Median	2,576
Std. Deviation	7,044
Q_1 (25th percentile)	1,233
Q_3 (75th percentile)	4,245
IQR ($Q_3 - Q_1$)	3,012
Minimum	10
Maximum	100,000

2.6 Missing Data Analysis

Table 7: Missing / Empty Fields in CCE Dataset

Field	**Missing/Empty**	**%**	**Impact**
`green_yield_plot`	173,302	99.6%	Unusable — nearly all empty
`yield`	44,737	25.7%	Significant — 1 in 4 records
`crop_name`	44,737	25.7%	Significant — same records as yield
`season`	44,688	25.7%	Significant — same cohort
`district_name`	44,688	25.7%	Same cohort as above
`village_lgd_code`	235	0.1%	Negligible
`latitude / longitude`	0	0.0%	Complete

Important:

Correlated Missing Data
The 44,737 records missing yield are the *same* records missing crop name, season, and district name. These appear to be **incomplete form submissions** where the enumerator recorded GPS coordinates and the farmer name but did not complete the crop measurement fields. After removing these records, the remaining 129,210 records have $> 99.9\%$ completeness on all critical fields.

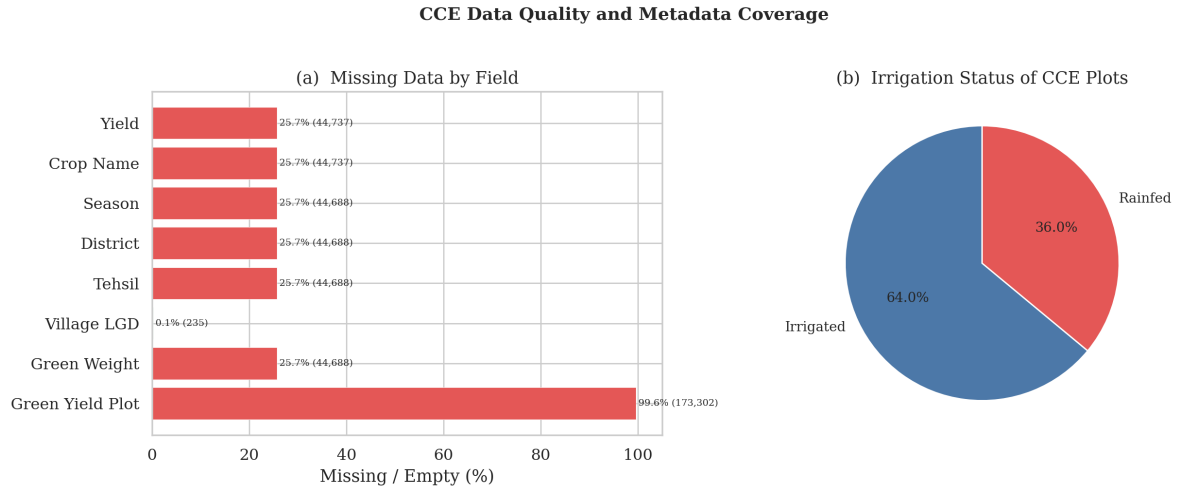

Figure 3: (a) Missing data by field: yield, crop name, season, and district all share the same 25.7% missing cohort. (b) Irrigation status: 64% of CCE plots are irrigated.

2.7 Irrigation Status

Table 8: CCE Plot Irrigation Breakdown

Status	Records	%
Irrigated	111,330	64.0%
Rainfed	62,617	36.0%

2.8 Crop × Season Yield Matrix

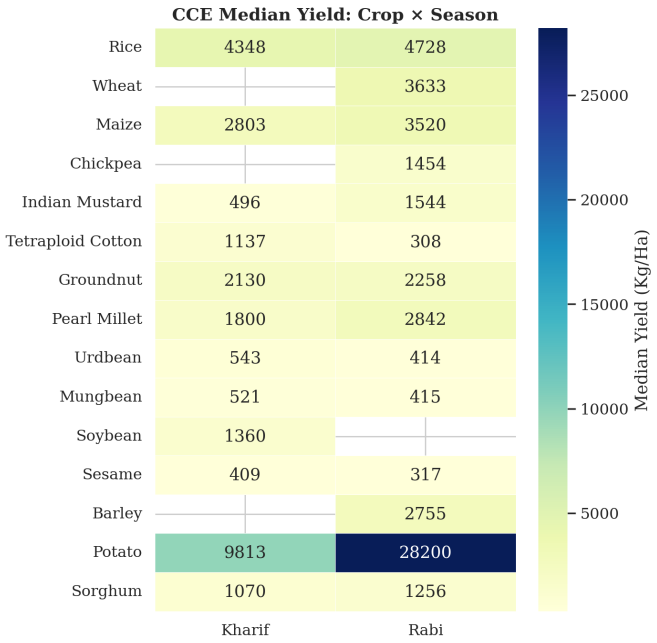

Figure 4: CCE median yield by crop and season. Rice is exclusively Kharif; Wheat exclusively Rabi. Crops like Maize, Groundnut, and Sorghum appear in both seasons with different yield profiles.

2.9 Chapter 1 Summary

CCE Dataset Assessment

- **Scale:** 173,947 raw records from 21 states, 496 districts, 56,070 villages, covering 49 crops
- **Usable records:** 128,311 after removing 25.7% incomplete submissions and 899 outliers
- **Yield unit:** Kg/Ha — directly comparable to DES district APY data
- **Temporal:** 2 calendar years (2023, 2024), mapping to fiscal 2023-24 and 2024-25
- **Strengths:** GPS-tagged, plot-level granularity, irrigation metadata, 49 crops
- **Weaknesses:** 25.7% incomplete submissions, `green_yield_plot` field unusable, no area or production data

3 Chapter 2: CCE vs District APY - Static Coherence Analysis

This chapter compares CCE district-aggregated yields against DES district-level APY yields. The goal is to quantify how well plot-level field measurements agree with administratively-reported district crop statistics.

3.1 Methodology

3.1.1 Notation and Definitions

Base Variables

- y_i = Yield of the i-th CCE experiment in a district-crop-season-year group (Kg/Ha)
- n = Number of valid CCE experiments in the group ($n \geq 5$)
- Y_{APY} = DES-reported district-level yield (Kg/Ha)
- A_d = DES-reported cultivated area for district d (Ha)
- D_s = Set of districts in state s; C_s = Set of crops grown in state s

3.1.2 Step 1: CCE Aggregation to District Level

For each matched group $(s, d, c, \text{season}, y)$, CCE plot-level yields are aggregated:

$$\boxed{\hat{Y}_{\text{CCE}} = \frac{1}{n}\sum_{i=1}^{n} y_i} \tag{1}$$

Additional descriptive statistics computed per group:

$$\tilde{Y}_{\text{CCE}} = \text{median}(y_1, y_2, \ldots, y_n) \tag{2}$$

$$\sigma_{\text{CCE}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{Y}_{\text{CCE}})^2} \tag{3}$$

$$\text{CV}_{\text{CCE}} = \frac{\sigma_{\text{CCE}}}{\hat{Y}_{\text{CCE}}} \times 100\% \tag{4}$$

3.1.3 Step 2: Deviation Metrics

$$\boxed{R = \frac{\hat{Y}_{\text{CCE}}}{Y_{\text{APY}}}} \quad \text{(Ratio)} \tag{5}$$

$$\boxed{D = \frac{\hat{Y}_{\text{CCE}} - Y_{\text{APY}}}{Y_{\text{APY}}} \times 100\%} \quad \text{(Percentage Deviation)} \tag{6}$$

Interpretation: $D > 0$ means CCE measures *higher* yield than APY reports; $D < 0$ means CCE measures *lower*.

3.1.4 Step 3: Agreement Classification

Table 9: Agreement Classification Thresholds

Category	Condition	Interpretation
EXCELLENT	$\|R-1\| < 0.10$	CCE and APY within ±10%
GOOD	$0.10 \leq \|R-1\| < 0.20$	Within ±20%
ACCEPTABLE	$0.20 \leq \|R-1\| < 0.30$	Within ±30%
EXPECTED	$0.8 f_c \leq R \leq 1.2 f_c$	Matches known conversion factor f_c
CCE_HIGHER	$R > 1.30$, $R \leq 2.0$	CCE systematically higher
CCE_LOWER	$R < 0.70$, $R \geq 0.5$	CCE systematically lower
LARGE_DIFF	$R > 2.0$ or $R < 0.5$	Major discrepancy

Here f_c is the **known conversion factor** for crops where CCE and APY measure different physical forms (see Chapter 5).

3.1.5 Step 4: Area-Weighted Aggregation

For state-level and crop-level summaries, we weight by the DES-reported district area:

$$\boxed{\bar{D}_s^{(w)} = \frac{\sum_{d \in D_s} (D_d \times A_d)}{\sum_{d \in D_s} A_d}} \tag{7}$$

This ensures high-producing districts dominate the aggregate rather than small districts with noisy CCE estimates.

3.1.6 Step 5: Correlation Analysis

For each crop, Pearson correlation and OLS regression are computed:

$$r_{\text{Pearson}} = \frac{\sum_d (\hat{Y}_d - \bar{\hat{Y}})(Y_d - \bar{Y})}{\sqrt{\sum_d (\hat{Y}_d - \bar{\hat{Y}})^2 \cdot \sum_d (Y_d - \bar{Y})^2}} \tag{8}$$

$$\hat{Y}_{\text{CCE}} = \alpha + \beta \cdot Y_{\text{APY}} + \varepsilon, \qquad R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} \tag{9}$$

3.2 Matching Funnel

Table 10: Data Matching Pipeline

Stage	Count
CCE records after QC	119,372
CCE district-crop-season-year groups ($n \geq 5$)	2,575
DES district records (2023-24 + 2024-25, excl. "Total" season)	33,011
Matched observations	**1,297**

3.3 Overall Coherence Results

Table 11: Overall CCE vs District APY Coherence (N = 1,297)

Metric	Value	Interpretation
Area-Weighted Mean Dev. ($\bar{D}^{(w)}$)	+23.5%	CCE systematically higher (Rice/Cotton dominate)
Area-Weighted Std. Dev. ($\sigma_D^{(w)}$)	46.5%	High variability (heterogeneous crop mix)
Median Deviation ($\tilde{D}$)	+5.6%	Most observations close to zero (unweighted, for reference)
Area-weighted proportion of cropped area within deviation bands:		
Within ±5%	24.8%	—
Within ±10%	50.9%	Half of cropped area
Within ±20%	64.8%	Nearly two-thirds of cropped area
Within ±30%	71.3%	About three-quarters of cropped area

All metrics except the median are **area-weighted**: each (district, crop, season, year) observation is weighted by its DES-reported crop area A_d, so that major production zones dominate the aggregate. The median is reported unweighted for reference.

Table 12: Agreement Category Distribution (area-weighted %)

Category	Count	%	Meaning
EXCELLENT	637	49.9%	$<\pm10\%$ difference
EXPECTED	178	18.9%	Matches known conversion factor
GOOD	172	13.0%	±10–20%
CCE_HIGHER	131	6.6%	CCE > APY by > 30%
ACCEPTABLE	90	5.7%	±20–30%
LARGE_DIFF	56	3.2%	Ratio > 2.0 or < 0.5
CCE_LOWER	33	2.6%	CCE < APY by > 30%
Total	**1,297**	**100%**	

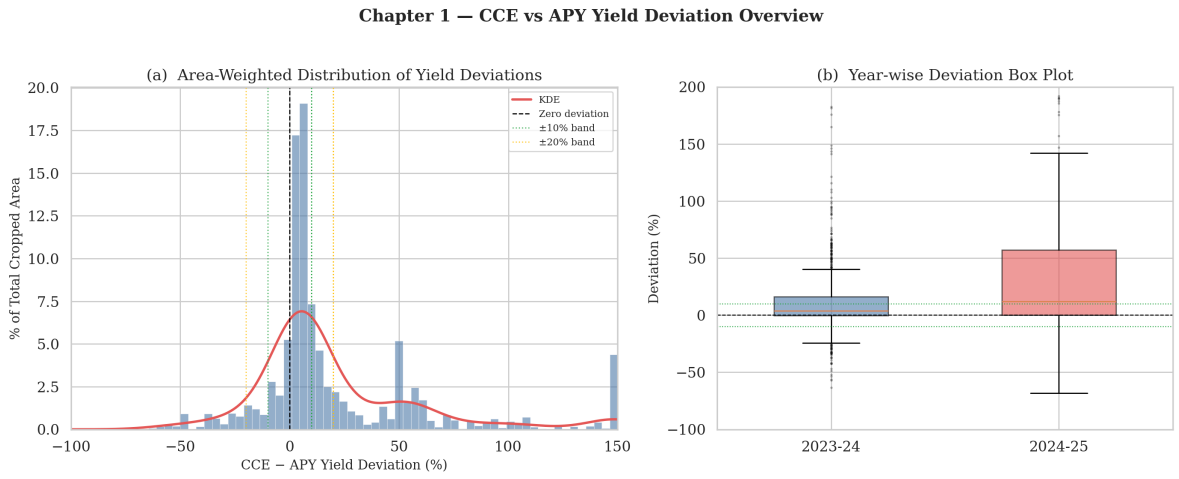

Figure 5: (a) Distribution of CCE–APY yield deviations with KDE overlay. Right-skew is driven by form-difference crops. (b) Box plot by fiscal year.

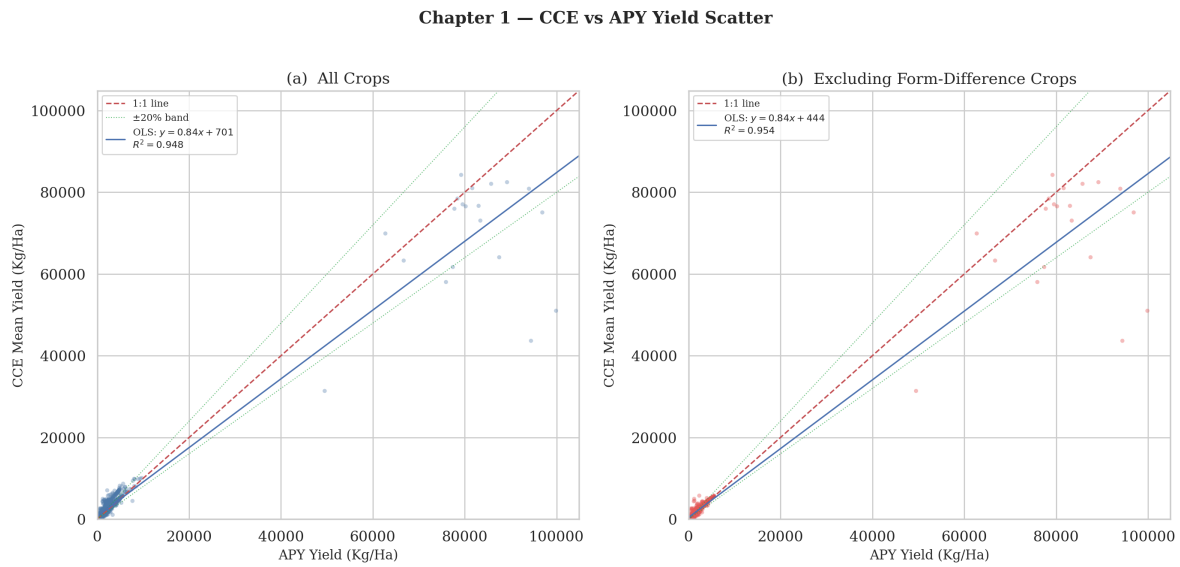

Figure 6: (a) All crops: strong linear relationship but systematic upward bias from Rice and Cotton. (b) Excluding form-difference crops: relationship tightens around 1:1 line with improved R^2.

3.4 Crop-wise Coherence

Table 13: Crop-wise CCE vs District APY Coherence (sorted by area). All metrics area-weighted. $\bar{D}^{(w)}$ = mean deviation, $P_{10}^{(w)}$ = % of cropped area within ±10%, r = Pearson. † = known form-difference.

Crop	N	$\bar{D}^{(w)}$	$\tilde{D}$	$P_{10}^{(w)}$	r	Assessment
Wheat	263	+6.1	+5.9	84.8%	0.94	Strong
Rice†	142	+64.0	+57.6	0.9%	0.92	Paddy vs milled
Mustard	172	+3.7	+0.9	79.1%	0.82	Good
Soybean	34	−4.3	+5.5	30.4%	0.64	Moderate
Cotton†	32	+142.2	+145.1	12.6%	0.42	Kapas vs lint
Maize†	100	+18.1	+8.7	38.4%	0.86	Cob vs shelled
Gram	93	+10.2	+2.2	53.9%	0.84	Good
Groundnut†	42	−1.2	−0.1	64.3%	0.93	Excellent
Tobacco	10	+0.2	+1.3	95.6%	0.95	Best match
Linseed	6	+0.4	+0.0	93.1%	1.00	Perfect
Sugarcane	20	−22.1	−7.6	15.3%	0.31	CCE lower
Tur	16	−30.0	−6.7	16.4%	0.31	CCE lower

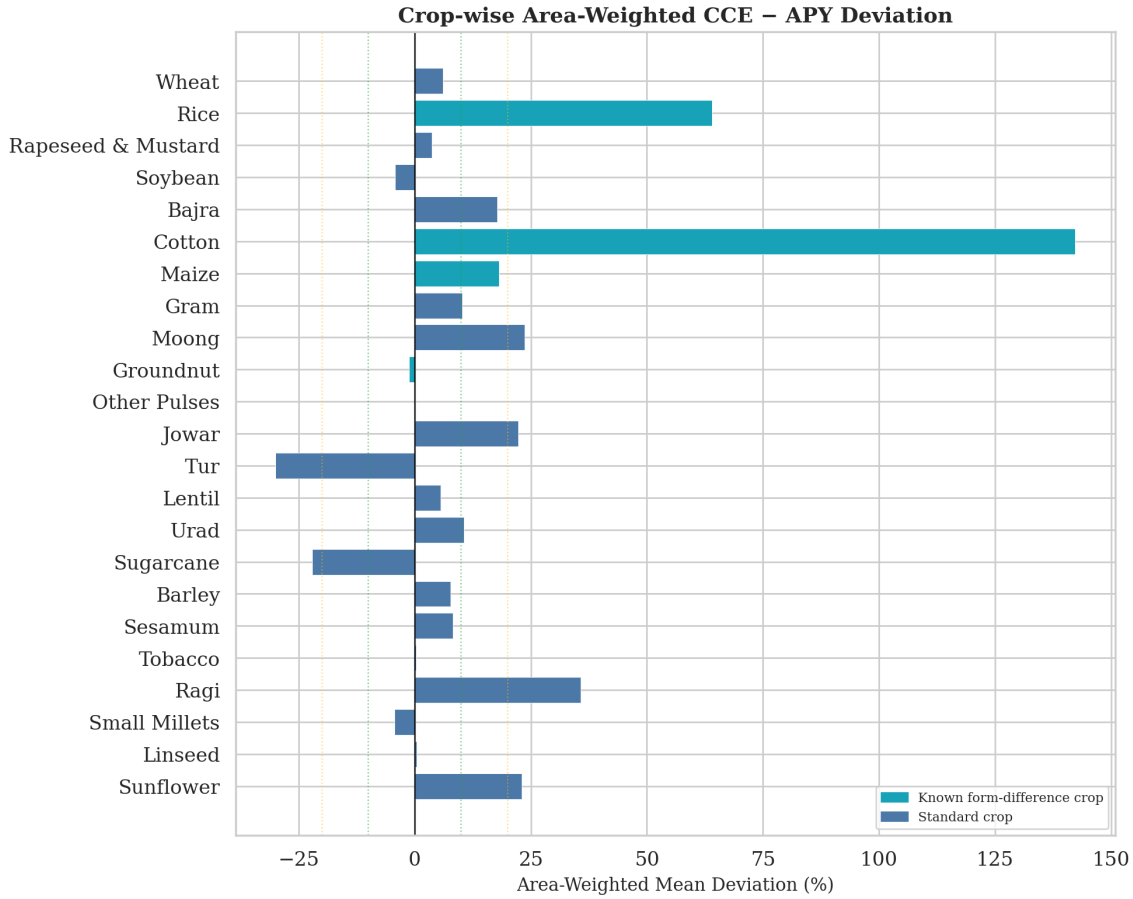

Figure 7: Crop-wise area-weighted mean deviation. Teal bars = crops with known form-differences; blue bars = standard crops. Dashed lines mark ±10% and ±20% bands.

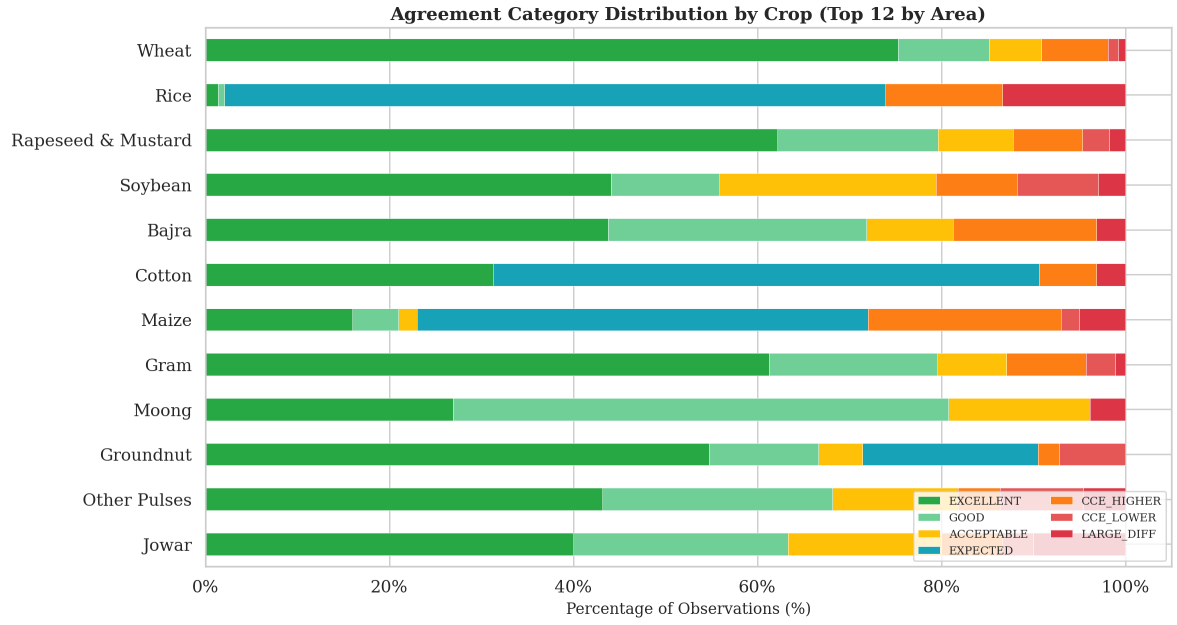

Figure 8: Agreement category distribution by crop (top 12 by area). Teal "EXPECTED" denotes deviations consistent with known form-differences.

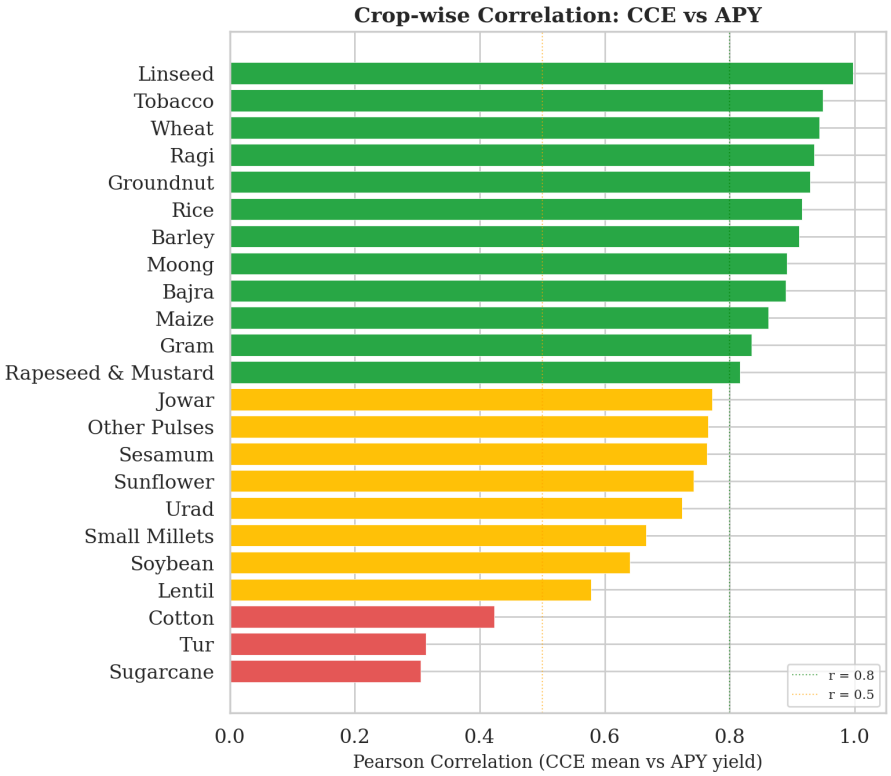

Figure 9: Pearson r between CCE mean and APY yield by crop. Green: $r \geq 0.8$; yellow: $0.5 \leq r < 0.8$; red: $r < 0.5$.

3.5 State-wise Coherence

Table 14: State-wise CCE vs District APY Coherence. All metrics area-weighted. $P_{10}^{(w)}$ = % of cropped area within ±10%.

State	N	Crops	Dist.	$\bar{D}^{(w)}$	$P_{10}^{(w)}$	$P_{20}^{(w)}$
Haryana	56	7	21	+8.8%	82.6%	87.1%
Rajasthan	253	17	40	+11.4%	60.7%	86.8%
Uttar Pradesh	178	11	69	+12.8%	73.7%	79.4%
Gujarat	95	10	30	+15.1%	74.1%	77.6%
Bihar	173	8	38	+17.9%	60.2%	77.5%
Punjab	111	13	23	+29.3%	51.0%	51.2%
Tamil Nadu	68	9	26	+41.1%	18.2%	32.2%
Maharashtra	106	16	29	+44.3%	12.6%	26.3%
Kerala	13	1	13	+58.8%	3.2%	3.2%
Jharkhand	40	4	17	+88.5%	9.5%	12.3%
Chhattisgarh	11	3	7	+179.1%	5.2%	9.5%

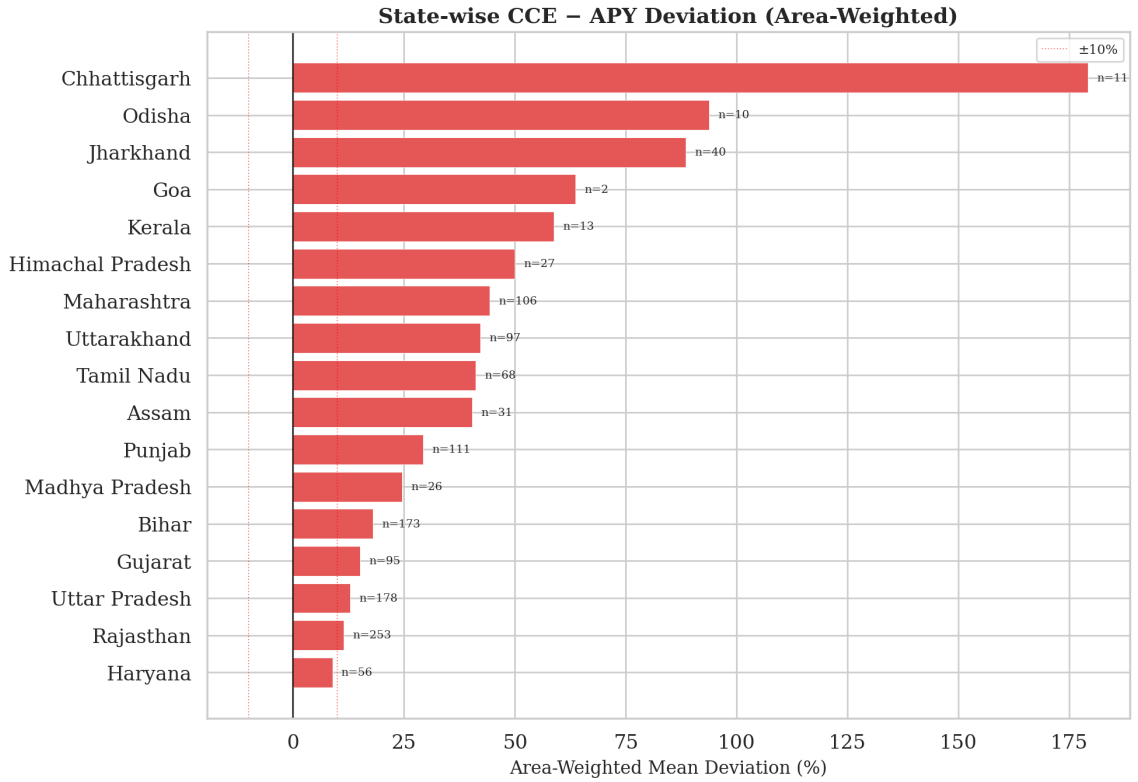

Figure 10: State-wise area-weighted mean deviation. All states show positive bias. States with large Rice share (Chhattisgarh, Jharkhand, Kerala) show the highest deviations due to the paddy-vs-milled form-difference.

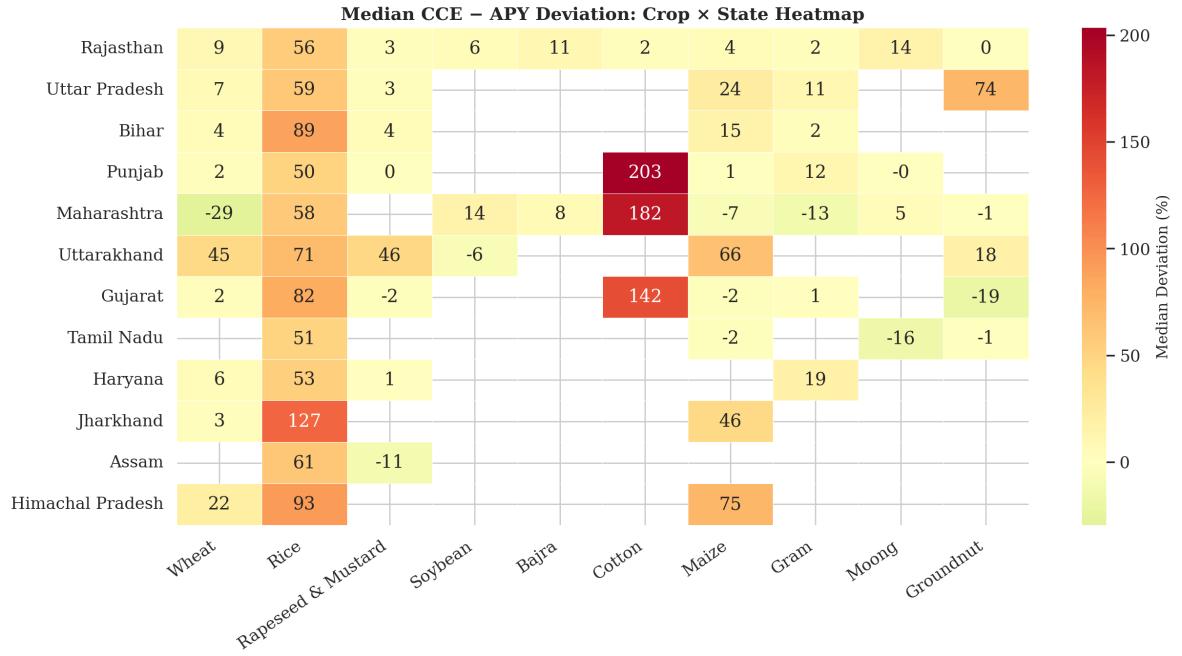

Figure 11: Median CCE–APY deviation: crop × state heatmap (top 10 crops, top 12 states). Red = CCE > APY; green = CCE < APY. The Rice row is uniformly red across all states.

4 Chapter 3: Temporal, Seasonal & Sample-Size Analysis

4.1 Year-wise Comparison

Table 15: Year-wise CCE vs District APY Deviation Summary (area-weighted)

Year	N	$\bar{D}^{(w)}$ (%)	$\tilde{D}$ (%)	$\sigma_D^{(w)}$ (%)	$P_{10}^{(w)}$
2023-24	889	+12.6	+4.0	29.7	69.3%
2024-25	408	+38.2	+12.1	59.5	25.8%

The 2024-25 data shows higher area-weighted deviations (+38.2% vs +12.6%). This is partly compositional: the 2024-25 DES dataset covers only 14 states, many of which are Rice-dominant, amplifying the paddy-vs-milled effect.

4.2 Season-wise Comparison

4.2.1 Methodology: Weighted Welch's t-test

To test whether Kharif and Rabi show systematically different CCE–APY deviations, we use a **weighted Welch's t-test** where each observation is weighted by its crop area A_d:

$$t = \frac{\bar{D}^{(w)}_{\text{Kharif}} - \bar{D}^{(w)}_{\text{Rabi}}}{\sqrt{\dfrac{(\sigma_K^{(w)})^2}{n_{\text{eff},K}} + \dfrac{(\sigma_R^{(w)})^2}{n_{\text{eff},R}}}} \tag{10}$$

where $n_{\text{eff}} = (\sum A_i)^2 \,/\, \sum A_i^2$ is the Kish effective sample size, which corrects for unequal weights.

Table 16: Season-wise Deviation and Hypothesis Test (area-weighted)

Season	N	$\bar{D}^{(w)}$ (%)	$\tilde{D}$ (%)	$\sigma_D^{(w)}$ (%)
Kharif	479	+41.0	+16.6	58.3
Rabi	818	+7.0	+3.6	21.0
Weighted Welch's t-test: $t = 7.329$, $p < 0.0001$				
Conclusion: Seasonal difference is **highly significant** at $\alpha = 0.05$.				

Note:

Why Kharif Deviates More

The three crops with known form-differences — Rice (paddy vs milled), Cotton (kapas vs lint), and Maize (cob vs shelled) — are all Kharif crops. They collectively account for 274 of the 479 Kharif observations (57%). Once these conversion-factor crops are excluded, the seasonal gap narrows substantially.

4.3 Sample Size Effect

To quantify how CCE accuracy improves with sample size, we bin observations by n (CCE experiments per group) and compute the area-weighted mean deviation per bin:

$$\bar{D}_{\text{bin}}^{(w)} = \frac{\sum_{j \in \text{bin}} D_j \cdot A_j}{\sum_{j \in \text{bin}} A_j} \tag{11}$$

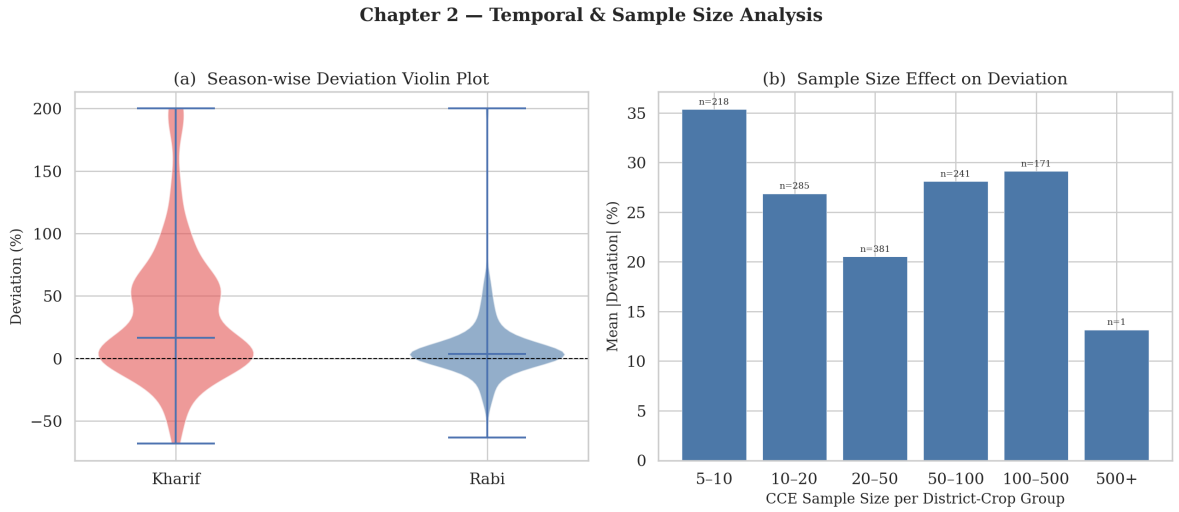

Figure 12: (a) Violin plot of deviation by season: Kharif shows wider spread and higher median. (b) Sample size effect: larger CCE groups produce more stable agreement with APY, consistent with the Central Limit Theorem.

5 Chapter 4: Spatial Multiplier Comparison

This chapter tests whether CCE and APY data **rank districts similarly** relative to a national baseline.

5.1 Methodology

5.1.1 National Base Yield

For each crop c, the national area-weighted base yield from DES district data:

$$\boxed{Y_{\text{base}}(c) = \frac{\sum_d \left(Y_d^{\text{APY}} \times A_d\right)}{\sum_d A_d}} \tag{12}$$

5.1.2 District Multipliers

$$\boxed{M_{\text{APY}}(d,c) = \frac{Y_{\text{APY}}(d,c)}{Y_{\text{base}}(c)}} \qquad M_{\text{CCE}}(d,c) = \frac{\hat{Y}_{\text{CCE}}(d,c)}{Y_{\text{base}}(c)} \tag{13}$$

$M = 1.5$ means 50% above national average; $M = 0.7$ means 30% below.

5.1.3 Classification

Table 17: Multiplier Performance Tiers

Category	Range	Meaning
EXCELLENT	$M \geq 1.5$	$\geq 50\%$ above national
ABOVE_AVG	$1.2 \leq M < 1.5$	20–50% above
NEAR_AVG	$0.8 \leq M < 1.2$	Within $\pm 20\%$
BELOW_AVG	$0.5 \leq M < 0.8$	20–50% below
POOR	$M < 0.5$	$> 50\%$ below

5.2 Results

Table 18: Spatial Multiplier Comparison Summary (N = 1,297)

Metric	Value
Pearson $r(M_{\text{CCE}}, M_{\text{APY}})$	0.666
Spearman $\rho(M_{\text{CCE}}, M_{\text{APY}})$	0.759
OLS slope β	1.078
OLS intercept α	0.093
OLS R^2	0.444
Mean $\Delta M = M_{\text{CCE}} - M_{\text{APY}}$	+0.169
Median ΔM	+0.054

Note:

Spearman $\rho = 0.759$ indicates **strong rank agreement** — districts that APY identifies as high-performing are also ranked highly by CCE. The OLS slope $\beta = 1.08 \approx 1$ confirms comparable multiplier scales. The positive mean $\Delta M = +0.17$ reflects the systematic CCE > APY bias from Chapter 2.

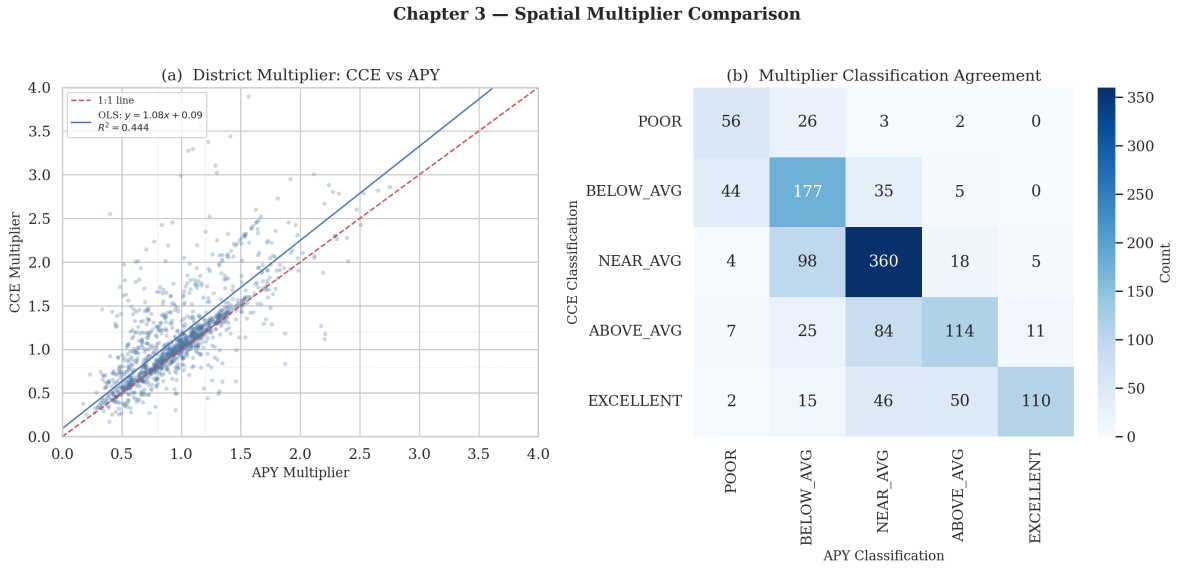

Figure 13: (a) CCE multiplier vs APY multiplier scatter with OLS line. (b) Classification confusion matrix: diagonal cells (agreement) are strongest; some APY BELOW_AVG districts get reclassified as NEAR_AVG by CCE.

Table 19: Performance Tier Distribution: CCE vs APY

Category	APY	CCE
EXCELLENT	126 (9.7%)	223 (17.2%)
ABOVE_AVG	189 (14.6%)	241 (18.6%)
NEAR_AVG	528 (40.7%)	485 (37.4%)
BELOW_AVG	341 (26.3%)	261 (20.1%)
POOR	113 (8.7%)	87 (6.7%)

CCE data tends to *upgrade* districts: more are classified EXCELLENT or ABOVE_AVG by CCE than by APY, consistent with the positive yield bias.

6 Chapter 5: Conversion Factor Validation

Several crops exhibit a known **physical form difference** between what CCE measures in the field and what APY reports.

6.1 Methodology

For each crop with a known factor f_c:

$$R_{\text{expected}} = f_c = \frac{\text{CCE form (e.g., paddy)}}{\text{APY form (e.g., milled rice)}} \tag{14}$$

$$R_{\text{observed}} = \text{median}\left(\frac{\hat{Y}_{\text{CCE},d}}{Y_{\text{APY},d}}\right) \quad \text{across all matched districts} \tag{15}$$

$$P_{\text{match}} = \frac{|\{d : 0.8 f_c \leq R_d \leq 1.2 f_c\}|}{N} \times 100\% \tag{16}$$

6.2 Results

Table 20: Conversion Factor Validation: Expected vs Observed CCE/APY Ratios

Crop	CCE Form	APY Form	f_c	N	$\tilde{R}$	$P_{\pm 20\%}$	Assessment
Rice	Paddy	Milled rice	1.49	142	1.58	71.8%	Good match
Cotton	Kapas	Lint	2.86	32	2.45	59.4%	Partial
Maize	Cob weight	Shelled	1.25	100	1.09	49.0%	Mixed
Groundnut	In-shell	Kernel	1.43	42	1.00	19.0%	Both report pods

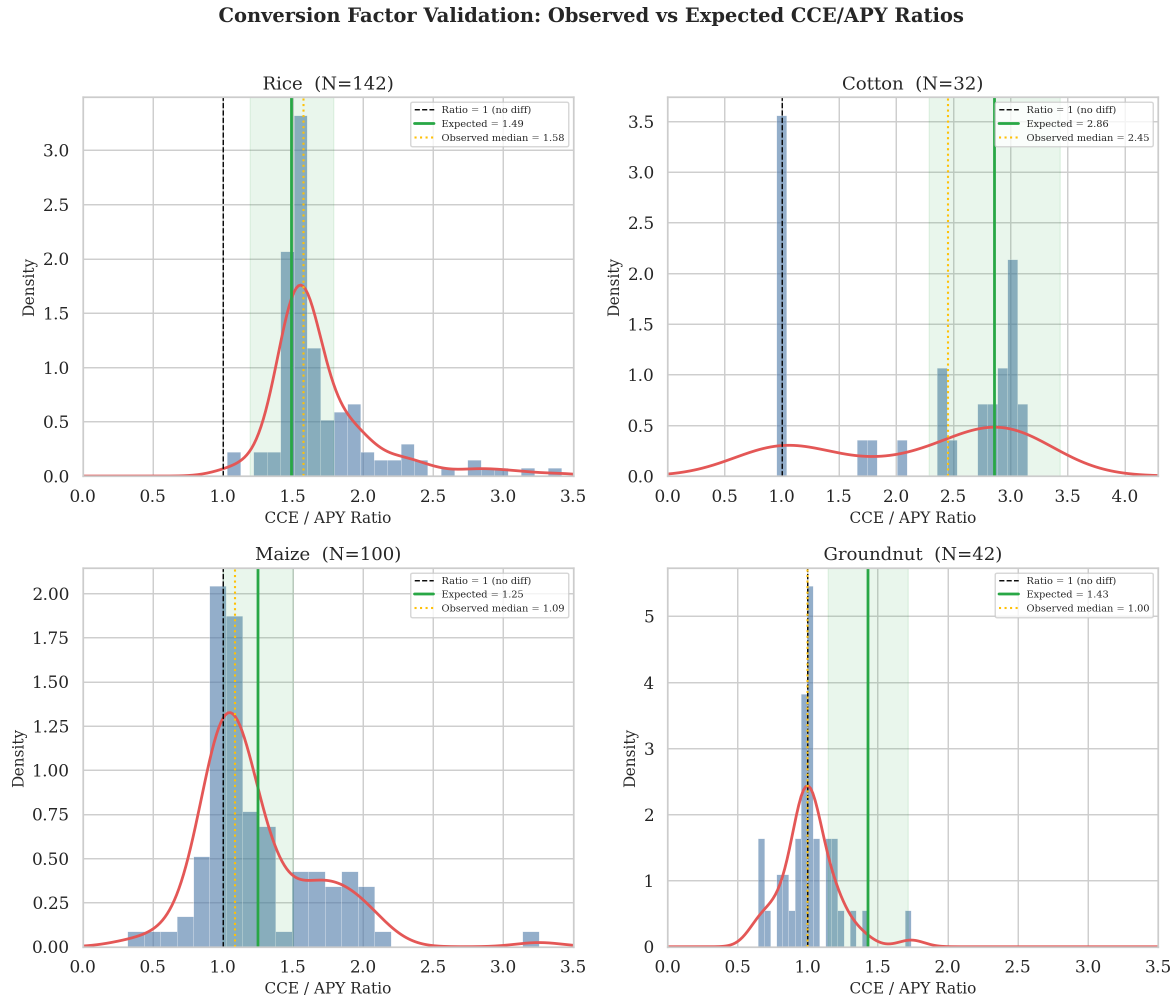

Figure 14: Observed CCE/APY ratio distributions for form-difference crops. Green line = expected f_c; yellow dashed = observed median; green shading = ±20% acceptance window; black dashed = 1.0 (no difference).

Finding:

Rice ($f_c = 1.49$, observed $\tilde{R} = 1.58$)
Observed median falls within 6% of expected, confirming APY reports **milled rice** while CCE measures **paddy**. 71.8% match rate validates the factor. The slight upward bias may reflect state-specific milling recovery variation (∼63–70%).

Finding:

Cotton ($f_c = 2.86$, observed $\tilde{R} = 2.45$)
Observed median is 14% below expected, suggesting effective ginning outturn of ∼41% rather than the assumed 35%. This could indicate improved cotton varieties or some APY data reporting kapas rather than lint. Only 59.4% match.

Finding:

Maize ($f_c = 1.25$, observed $\tilde{R} = 1.09$)
Observed median substantially lower than expected — most CCE experiments appear to report **shelled grain**, not cob weight. The cob-vs-shelled correction is not universally applicable.

Finding:

Groundnut ($f_c = 1.43$, observed $\tilde{R} = 1.00$)
Ratio of ∼1.0 confirms **both CCE and APY report in-shell pods**. The kernel conversion factor does not apply. For Groundnut, CCE and APY are directly comparable without correction.

7 Summary and Recommendations

CCE vs District APY Coherence: Overall Assessment

Agreement Metrics (N = 1,297, all area-weighted):

- 49.9% of cropped area EXCELLENT ($< \pm 10\%$), 63.0% EXCELLENT + GOOD ($< \pm 20\%$)
- 18.9% classified as EXPECTED (matching known form-difference factors)
- Pearson $r = 0.67$, Spearman $\rho = 0.76$ (moderate-to-strong)

Crops with Best Agreement ($\bar{D}^{(w)} < \pm 10\%$):

- Tobacco (+0.2%), Linseed (+0.4%), Groundnut (−1.2%), Rapeseed & Mustard (+3.7%)
- Small Millets (−4.4%), Soybean (−4.3%), Wheat (+6.1%)

Crops Requiring Conversion:

- **Rice:** Divide CCE by 1.49 (paddy → milled)
- **Cotton:** Divide CCE by 2.45 (observed kapas → lint)
- **Maize:** No universal correction (most CCE already reports shelled)
- **Groundnut:** No correction (both report in-shell pods)

States with Best Agreement ($\bar{D}^{(w)} < 15\%$):

- Haryana (+8.8%), Rajasthan (+11.4%), Uttar Pradesh (+12.8%)

Seasonal Pattern (area-weighted):

- Kharif: +41.0% (driven by Rice/Cotton form-differences)
- Rabi: +7.0% (close agreement)
- Difference is **statistically significant** (weighted Welch's $t = 7.33$, $p < 0.0001$)

Spatial Ranking:

- Spearman $\rho = 0.76$: CCE and APY **rank districts similarly**
- Classification tier agreement is strongest for NEAR_AVG districts

Important:

Recommendations for Using CCE Data Alongside District APY

1. For **Rice**: apply paddy-to-milled correction (÷1.49) before comparing with APY
2. For **Cotton**: apply kapas-to-lint correction (÷2.45, observed) before comparing
3. For **Rabi crops** (Wheat, Gram, Mustard, Lentil, Barley): CCE data can be used *directly* — strong agreement with APY ($\bar{D}^{(w)} < 10\%$ for most)
4. Require $n \geq 20$ CCE experiments per district-crop group for reliable estimates
5. Flag observations where $|D| > 50\%$ for manual review