UGP 3 - ECO498A

Mentor- Dr. Anna David Thottappilly

District-Level Determinants of Crop Yields in India: The Role of Weather, Irrigation, and Fertilizer Use

1. Introduction

India is the second-largest producer of several key crops, including rice, wheat, sugarcane, cotton, and groundnuts [1]. Understanding the determinants of crop productivity in such a major agricultural economy requires integrating both climatic and management variables. In this study, we analyze district-level yield data alongside weather variables—temperature and precipitation—and management inputs such as irrigation and fertilizer use. Irrigation data are especially important for interpreting fertilizer consumption patterns, as the effectiveness of applied nutrients is closely linked to water availability.

Given the distinct agricultural seasons in India, we define the Rabi season (November to March) for wheat, sugarcane, and groundnut, and the Kharif season (June to October) for rice and cotton. This seasonal classification informs the temporal alignment of weather variables with crop-specific growing periods

Crop	Season	Sowing Time	Harvesting Time
Rice	**Kharif**	June – July	September – October
Wheat	**Rabi**	October – December	March – April
Sugarcane	**Annual/Perennial (mainly Rabi)**	October – March (can vary)	10–18 months later (varies by state)
Cotton	**Kharif**	April – May (in some states, June–July)	October – January
Groundnut	**Both Kharif & Rabi** (mostly **Kharif** in rainfed areas, **Rabi** in irrigated areas)	Kharif: June – July Rabi: October – November	Kharif: September – October Rabi: February – March

Crop wise Production details is given in below table

Crop	1st State (≈%)	2nd State (≈%)	3rd State (≈%)	4th State (≈%)	5th State (≈%)
Rice	West Bengal (~13.7%)	Uttar Pradesh (~12.8%)	Punjab (~10.0%)	Andhra Pradesh (~8%)	Telangana (~7%)
Wheat	Uttar Pradesh (~31.8%)	Madhya Pradesh (~22.6%)	Punjab (~17.7%)	Haryana (~11.2%)	Rajasthan (~8.6%)
Sugarcane	Uttar Pradesh (~46.9%)	Maharashtra (~26.6%)	Karnataka (~7.5%)	Tamil Nadu (~4.4%)	Bihar (~3.7%)
Cotton	Gujarat (~27.9%)	Maharashtra (~24.7%)	Telangana (~15.6%)	Rajasthan (~8.1%)	Karnataka (~6.3%)
Groundnuts	Gujarat (~42%)	Rajasthan (~17%)	Tamil Nadu (~11%)	Andhra Pradesh (~9%)	Karnataka (~6%)

We have taken below states for our analysis

State	Appears for Crop(s)
Uttar Pradesh	Rice, Wheat, Sugarcane
West Bengal	Rice
Punjab	Rice, Wheat
Andhra Pradesh	Rice, Groundnuts
Telangana	Rice, Cotton
Madhya Pradesh	Wheat

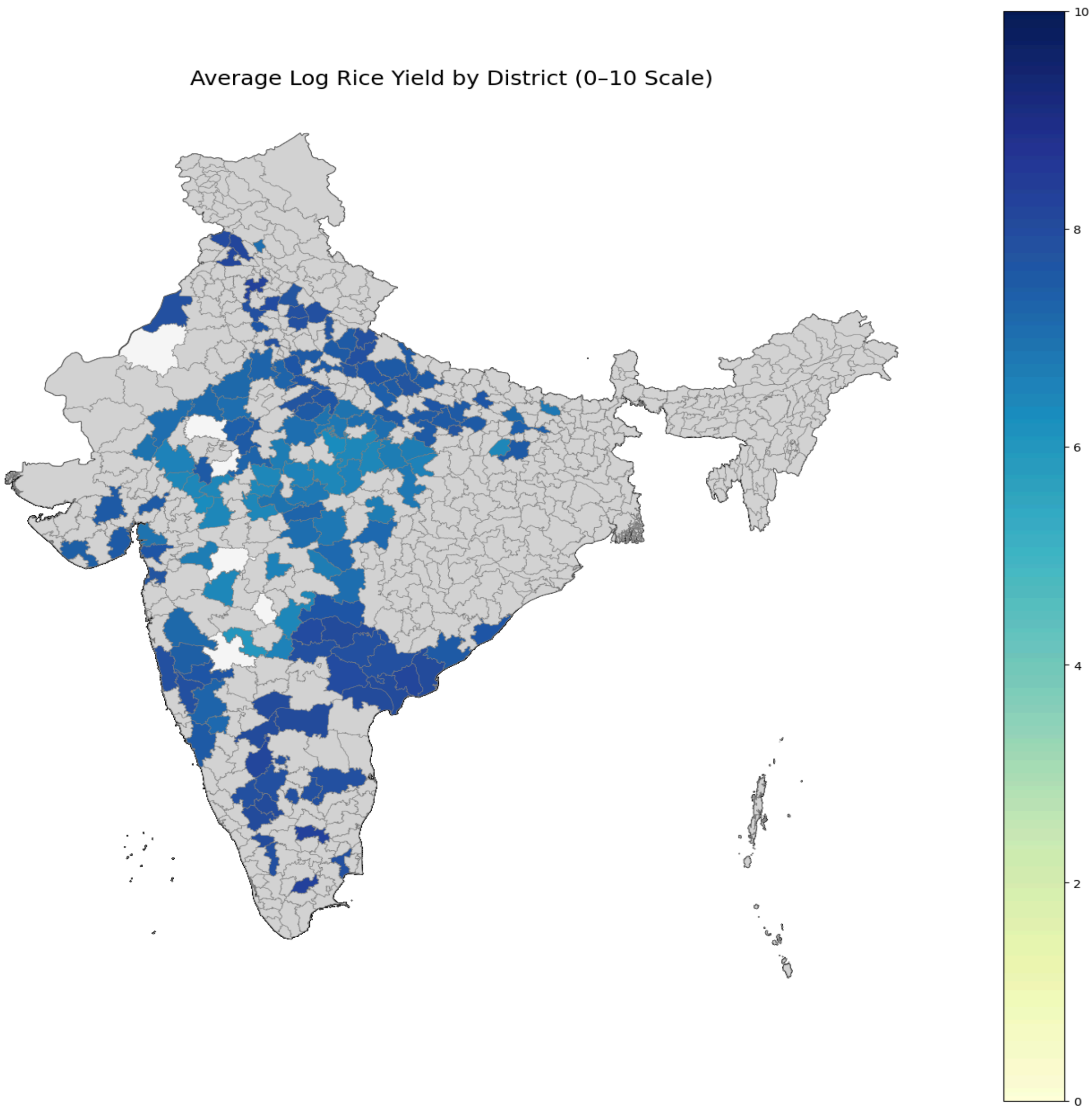
Haryana	Wheat, Cotton
Rajasthan	Wheat, Cotton, Groundnuts
Maharashtra	Sugarcane, Cotton
Karnataka	Sugarcane, Cotton, Groundnuts
Tamil Nadu	Sugarcane, Groundnuts
Bihar	Sugarcane
Gujarat	Cotton, Groundnuts

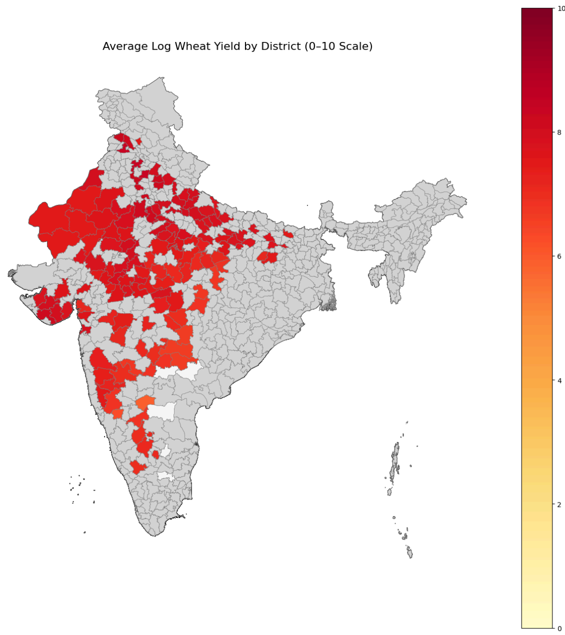
2. Data

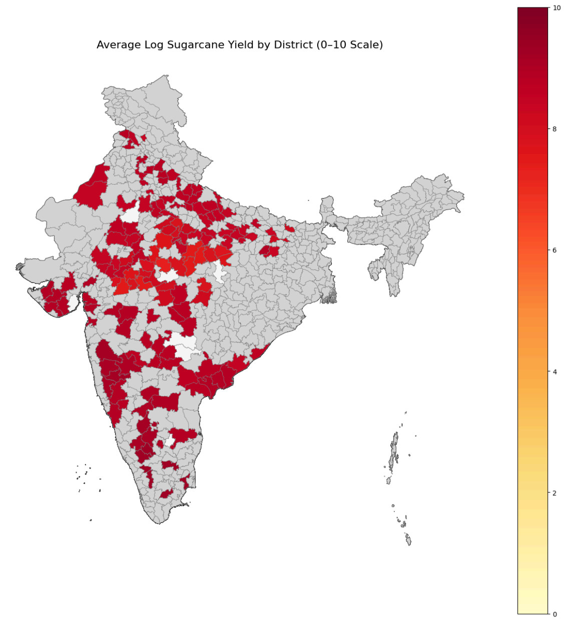
Crop	Total Observations	Total States	Total Districts	Time Range
Rice	3,969	12	148	1990 - 2017
Wheat	4,688	12	142	1984 - 2017
Sugarcane	4,758	12	150	1984 - 2017
Cotton	3,247	11	106	1984 - 2017
Groundnut	4,657	12	145	1984 - 2017

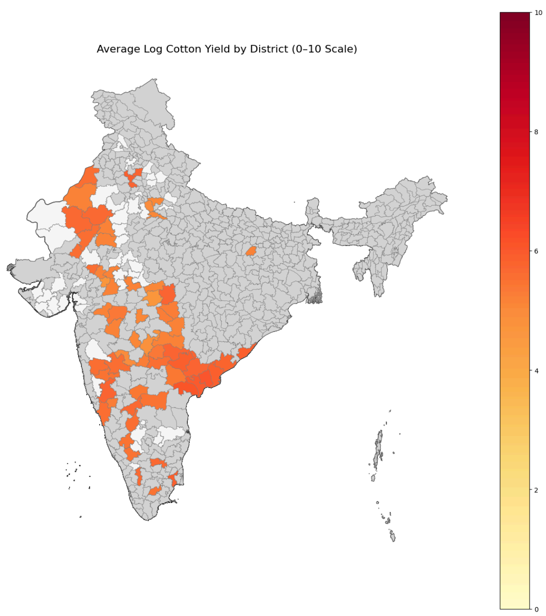
2.1 Yield Data

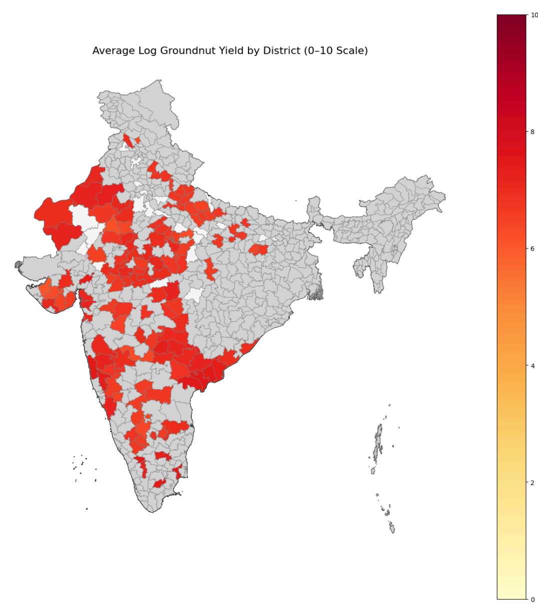
Our dependent variables are district-level log yields (log kg/ha) for five staple crops in India: rice, wheat, sugarcane, groundnut, and cotton. The yield data were sourced from ICRISAT and span multiple decades—specifically from 1984 to 2017 for most crops, and from 1990 for rice. Summary statistics of average log yields show meaningful variation across crops and districts. For instance, sugarcane exhibited the highest median log yield (8.63), while cotton showed the lowest (5.49). To ensure the robustness of our analysis, districts that displayed constant yield values across all available years—indicative of missing or uninformative reporting—were flagged and excluded from the baseline regression estimation.


Average Log Rice Yield by District (0–10 Scale)
10
8
6
4
2
0


Average Log Wheat Yield by District (0–10 Scale)
10
8
6
4
2
0

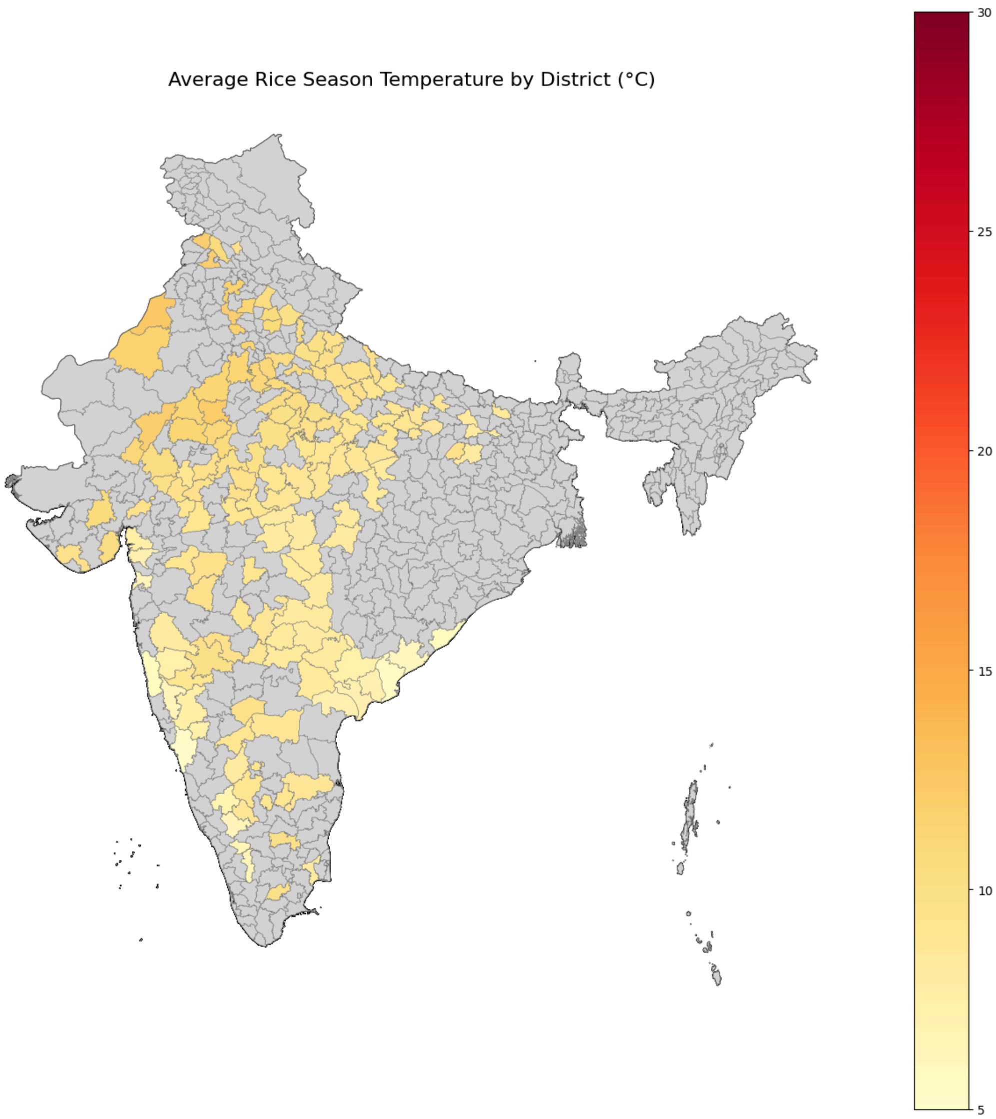

Average Log Sugarcane Yield by District (0–10 Scale)
10
8
6
4
2
0

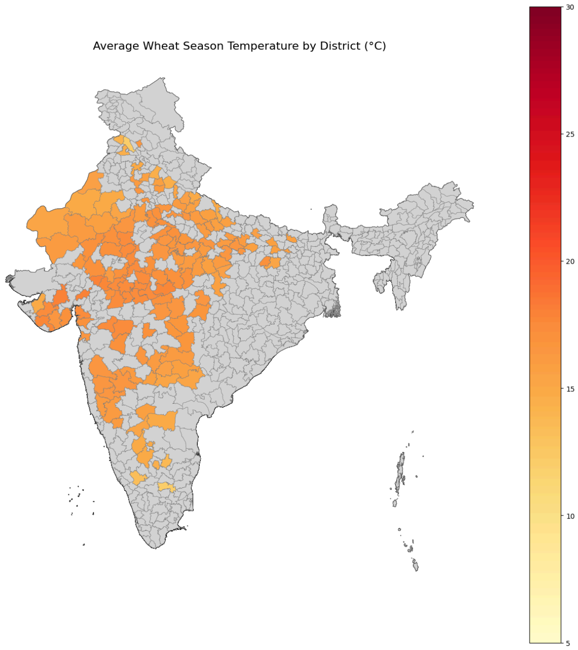

Average Log Cotton Yield by District (0–10 Scale)
10
8
6
4
2
0

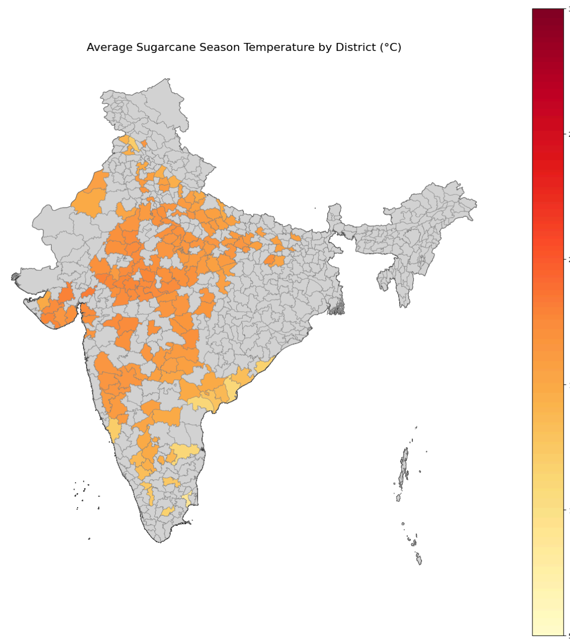

Average Log Groundnut Yield by District (0–10 Scale)
10
8
6
4
2
0

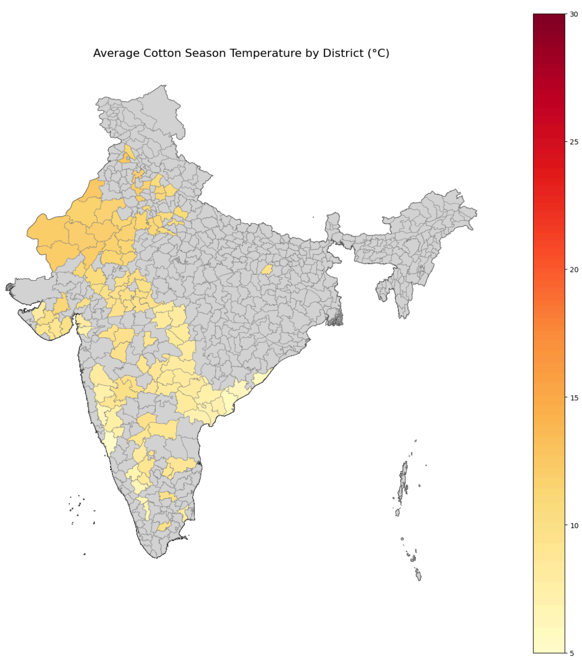
2.2 Weather Data

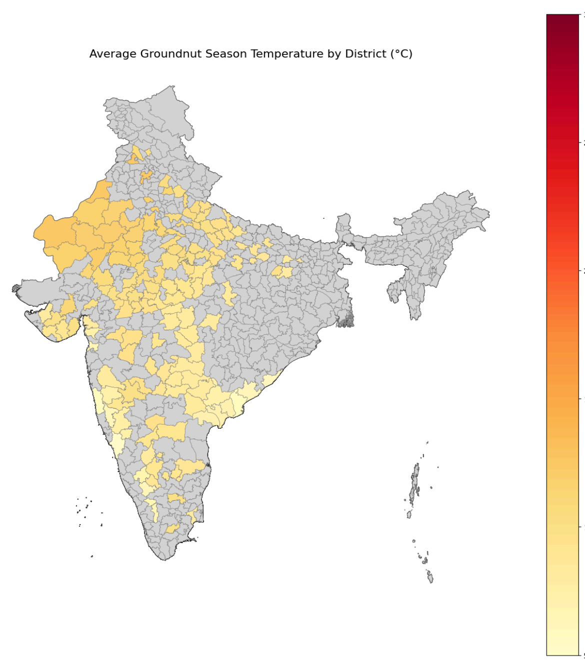
The primary weather variable used in this study is air temperature (T2M), defined as the average air (dry bulb) temperature measured at 2 meters above the Earth's surface. This data was sourced from the NASA POWER Data Access Viewer (DAV), provided at a 0.5° × 0.5° spatial resolution. To generate district-level temperature estimates, we obtained administrative boundary shapefiles from the GADM database (version 3.6). These shapefiles were then used to spatially aggregate the gridded temperature data, allowing us to compute district-wise average temperatures across years corresponding to our crop yield panel. This approach enables consistent integration of climate information with agricultural performance at the sub-national level.











In addition to temperature, we incorporate precipitation as a key climatic variable in our analysis. Daily precipitation data were obtained from the NASA POWER Data Access Viewer (DAV), available at a spatial resolution of 0.5° × 0.5°. Using the same administrative boundary shapefiles from the GADM database (version 3.6), we aggregated the gridded precipitation data to generate district-level average values. This spatial processing ensures alignment between climatic inputs and the geographical units used in our crop yield panel, enabling robust assessment of weather impacts on agricultural productivity.

2.3 Irrigation Data

Irrigation data for this study were obtained from the ICRISAT database, providing information on the extent of irrigated area for each crop at the district level. These data, reported in consistent units across years, were merged with the corresponding crop yield panel to capture the role of water availability through managed irrigation. Since irrigation is a key determinant of yield resilience under variable climatic conditions, its inclusion allows for a more comprehensive understanding of the interaction between weather variables and agricultural performance across districts and time.

3. Methodology

✅ General Panel Regression Framework

$$y_{it} = f(w_{it}) + \gamma_1 t + \gamma_2 t^2 + c_i + \epsilon_{it}$$

Where:

- $y_{it} = \log(\text{YIELD})$ for crop in district i, year t
- $f(w_{it})$ is the **weather function**, which includes weather, irrigation, and fertilizer variables
- γ_1, γ_2: coefficients for time trend
- c_i: district fixed effects
- ϵ_{it}: error term

📘 Model 1: Linear Weather Function

$$f(w_{it}) = \alpha_1 h_{it} + \beta_1 p_{it} + \delta_1 irr_{it} + \delta_2 fert_{it}$$

Where:

- h_{it}: average temperature (Rabi or Kharif depending on crop)
- p_{it}: average precipitation (Rabi or Kharif depending on crop)
- irr_{it}: irrigated area (in 1000 ha) for the crop in district i, year t
- $fert_{it}$: fertilizer consumption (tons per ha) for the season in district i, year t

📗 Model 2: Quadratic Weather Function

$$f(w_{it}) = \alpha_1 h_{it} + \alpha_2 h_{it}^2 + \beta_1 p_{it} + \beta_2 p_{it}^2 + \delta_1 irr_{it} + \delta_2 irr_{it}^2 + \delta_3 fert_{it} + \delta_4 fert_{it}^2$$

This version allows for nonlinearities in:

- Temperature and precipitation
- Fertilizer and irrigation

4. Regression Results

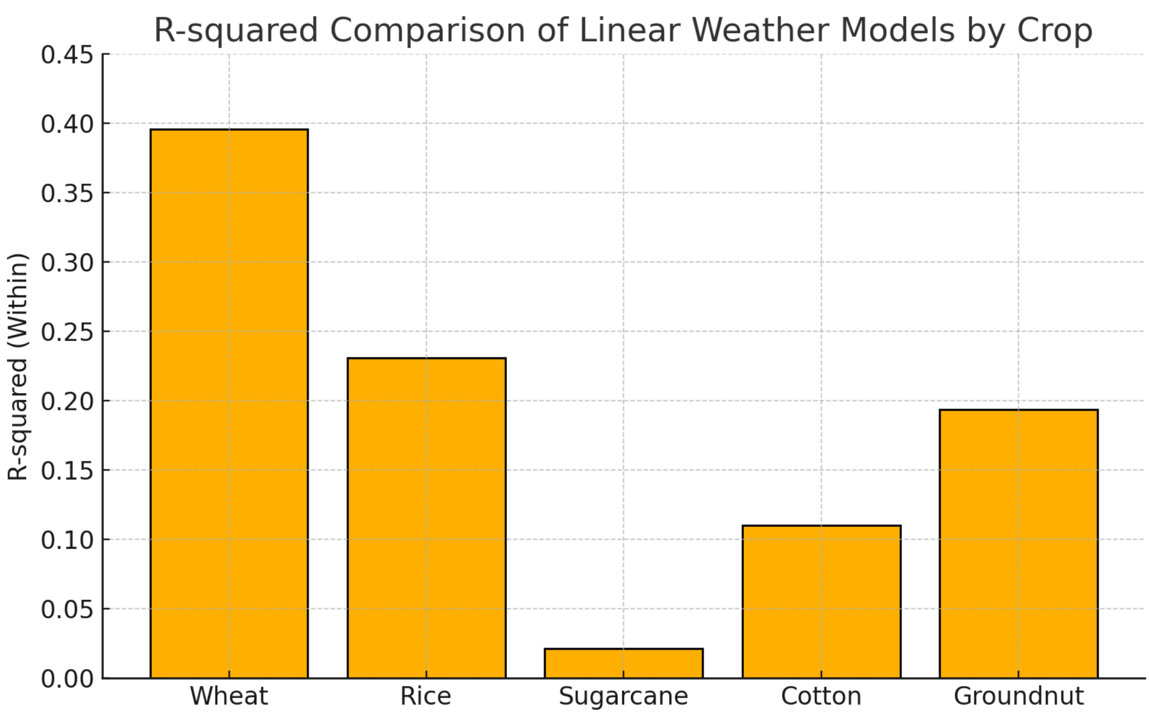
4.1 Model 1 : Linear Model with Average Temperature

Table 1: Model 1: Linear Model with Average Temperature

	Wheat	Rice	Sugarcane	Groundnut	Cotton
Avg. Temperature	-0.0209***	-0.1319***	-0.0860***	-0.2169***	-0.2285***
	(0.0071)	(0.0212)	(0.0143)	(0.0343)	(0.0444)
Precipitation	-0.0357***	0.0171	-0.0038	0.0034	0.0112
	(0.0046)	(0.0138)	(0.0054)	(0.0220)	(0.0268)
Irrigation	0.1906***	0.0648**	0.0274**	0.0563**	0.0945***
	(0.0237)	(0.0320)	(0.0115)	(0.0222)	(0.0242)
Fertilizer	-0.0058**	-0.0099	-0.0030	-0.0047***	0.1324**
	(0.0027)	(0.0251)	(0.0048)	(0.0008)	(0.0611)
R-squared	0.395	0.231	0.022	0.194	0.110
Observations	4673	3656	4748	4614	3098

Notes: Table reports regression coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

📊 R-squared Comparison



Interpretation by Crop

🌾 WHEAT

- **R-squared (within):** 0.395 → Model explains ~40% of variation within districts over time.
- **Temp:** Negative and significant (−0.021) — higher temperatures reduce yield.
- **Precip:** Also negative (−0.036)- possibly reflects excessive or poorly timed winter rainfall.
- **Irrigation:** Strongly positive (0.191) — effective input.
- **Fertilizer:** Slightly negative (−0.006), significant — may reflect overuse or saturation.
- **Trend (`t`):** Positive — wheat yields improved over time.

🌾 RICE

- **R-squared (within):** 0.231 → Moderate fit.
- **Temp:** Strongly negative (−0.132) — high Kharif heat harms rice, expected.
- **Precip:** Not significant — rainfall alone is less important due to irrigation.
- **Irrigation:** Positive and significant (0.065) — aligns with rice's water needs.
- **Fertilizer:** Not significant.
- **Trend (`t`):** Negative — possibly due to soil fatigue, water stress, or pest pressure.

🌱 SUGARCANE

- **R-squared (within):** 0.0215 → **Very poor fit**
- **Temp:** Negative and significant (−0.086) — excessive heat reduces cane biomass.
- **Precip, Fertilizer:** Not significant.
- **Irrigation:** Positive (0.027), significant but small — logical but limited effect in this model.
- **Trend:** Insignificant.

🌿 COTTON

- **R-squared (within):** 0.110 → Low to moderate fit.

- **Temp:** Strongly negative (−0.229) - high temp stresses cotton, especially during flowering.
- **Precip:** Not significant — rain timing matters more than totals.
- **Irrigation:** Positive (0.095) — supports cotton under thermal stress.
- **Fertilizer:** Positive and significant (0.132) — expected.

🥜 GROUNDNUT

- **R-squared (within):** 0.194 → Reasonable.
- **Temp:** Strongly negative (−0.217) — high temps reduce oilseed yield.
- **Precip:** Not significant.
- **Irrigation:** Positive (0.056), significant — helps mitigate moisture stress.
- **Fertilizer:** Negative (−0.005), significant — could indicate overuse or lack of response.

General Insights

- **Temperature** has a **consistently negative** effect across all crops — expected under India's warming trends.
- **Irrigation** is the most consistently beneficial input.
- **Fertilizer effects** are mixed — possibly due to:
  - Aggregated data masking timing effects
  - Threshold or overuse in some district
- **Rainfall totals** are mostly insignificant — rainfall **timing and intensity** would explain more.

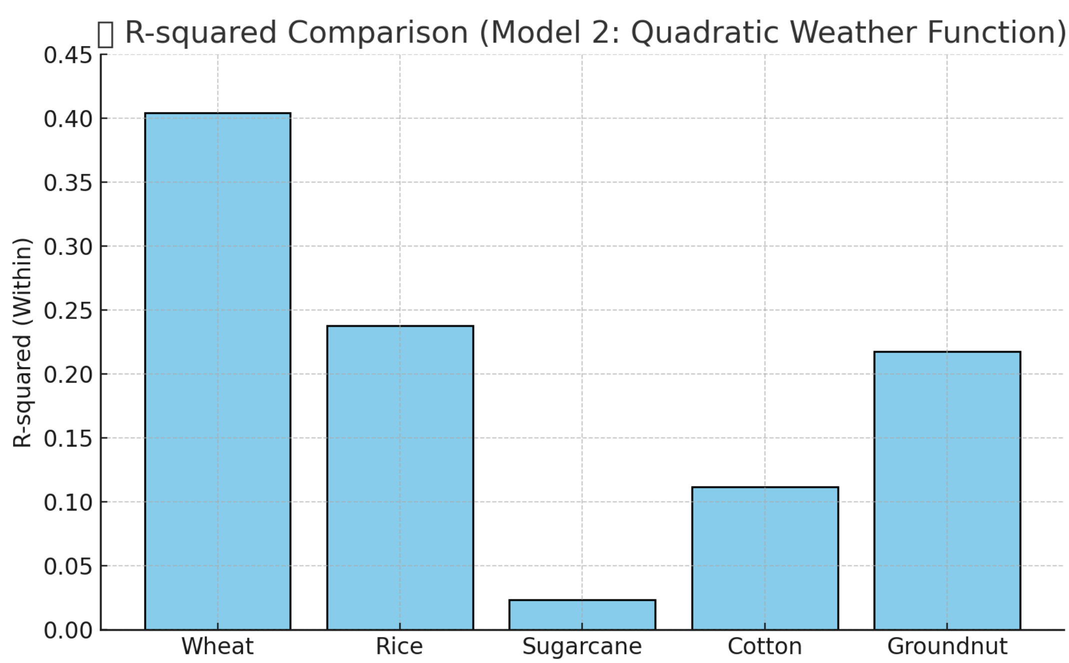
4.2 Model 2: Quadratic Model with Weather Function

Table 2: Model 2: Quadratic Weather Function

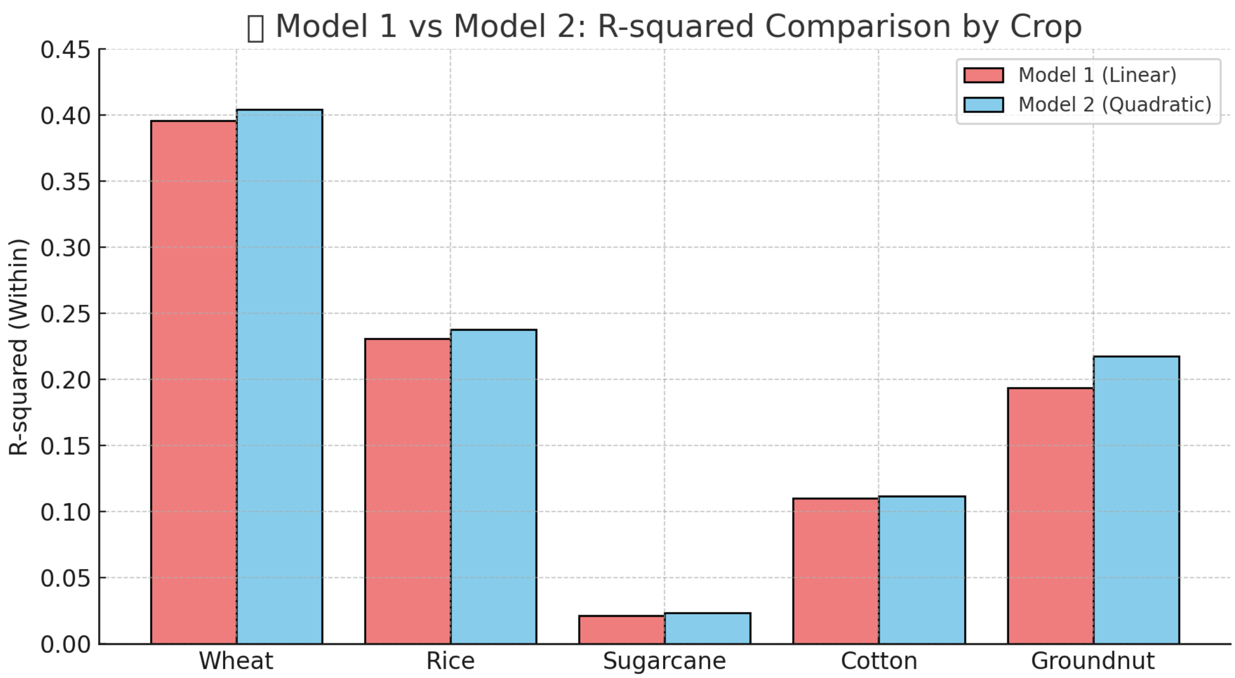
	Wheat	**Rice**	**Sugarcane**	**Groundnut**	**Cotton**
Temperature	-0.0165**	-0.1080***	-0.0837***	-0.0683**	-0.2548***
	(0.0073)	(0.0201)	(0.0139)	(0.0306)	(0.0459)
Temperature^2	-0.0048	-0.0054	-0.0147**	-0.0371	0.0317
	(0.0048)	(0.0110)	(0.0064)	(0.0234)	(0.0214)
Precipitation	-0.0287***	0.0369**	-0.0016	0.1459***	0.0064
	(0.0058)	(0.0174)	(0.0075)	(0.0511)	(0.0408)
Precipitation^2	-0.0033	-0.0050	0.0010	-0.0387**	-0.0079
	(0.0031)	(0.0037)	(0.0011)	(0.0185)	(0.0108)
Irrigation	0.2378***	0.0798*	0.0283	0.1118***	0.1037**
	(0.0265)	(0.0462)	(0.0210)	(0.0264)	(0.0497)
Irrigation^2	-0.0393***	-0.0102	0.0006	-0.0032***	-0.0016
	(0.0108)	(0.0174)	(0.0040)	(0.0009)	(0.0047)
Fertilizer	0.0336*	0.0508	0.0352	-0.0629***	0.2224
	(0.0197)	(0.0324)	(0.0315)	(0.0063)	(0.1460)
Fertilizer^2	-0.0009**	-0.0225***	-0.0013	0.0012***	-0.0149
	(0.0004)	(0.0084)	(0.0010)	(0.0001)	(0.0152)
R-squared	0.404	0.238	0.024	0.218	0.112
Observations	4673	3656	4748	4614	3098

Notes: Table reports regression coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

📊 R-squared Comparison



📊 R-squared Comparison Between Model 1 and Model 2.



Interpretation by Crop

🌾WHEAT

ΔR^2: +0.009 → Small but real improvement.

Significant nonlinear terms:

- `irrigation_sq`: Negative — diminishing returns to irrigation.
- `fertilizer_sq`: Negative — over-application may reduce yield.

Conclusion: Quadratic terms slightly improve fit; effects are plausible.

🌾 RICE

ΔR^2: +0.006 → Slight gain.

`temp`: still very negative and significant.

`preci:` becomes significant → quadratic model captures its delayed/threshold effects.

`fertilizer_sq`: Negative and significant — over-fertilization harmful.

🌱 SUGARCANE

ΔR^2: +0.002 → Marginal change.

`temp_sq`: Negative and significant — heat stress accelerates at higher temps.

All other quadratic terms: Not significant.

`t`, `t_sq`: remain weak.

🌿 COTTON

ΔR^2: +0.002 → Minimal improvement.

`temp`: Very strong negative impact.

`temp_sq`: weakly positive (marginally curved response).

Most other quadratic terms are **insignificant**.

🥜 GROUNDNUT

ΔR^2: +0.024 → Largest R^2 improvement among crops.

`precip` and `precip_sq`: classic quadratic curve — moderate rain good, excess bad.

`fertilizer`: sharp negative; but `fert_sq` positive — **U-shaped** → overcorrection possible.

`irrigation_sq`: Negative and significant — irrigation has limits.

Conclusion: Quadratic terms significantly improve understanding of nonlinear climate response.

🎯 Overall Insights

Where Quadratic Model Helps:

Groundnut: Significant gain in R^2 and more intuitive curvature.

Wheat & Rice: Small but logical improvements.

5 Conclusions

- **Temperature consistently reduces yields**, especially in rice, wheat, and cotton.
- **Irrigation** shows strong positive effects, particularly for wheat and rice.
- **Fertilizer** exhibits mixed or diminishing effects, with potential overuse in some regions.
- **Quadratic models** improve R^2 slightly, more noticeably for **groundnut**.
- **Sugarcane and cotton** models highlight the limitation of weather-only models — non-climatic factors are essential for accurate yield estimation.

6 References

1. https://www.fao.org/india/fao-in-india/india-at-a-glance/en/

2. Schlenker, Wolfram, and David B. Lobell. 2010. "Robust Negative Impacts of Climate Change on African Agriculture." *Environmental Research Letters* 5 (1): 014010. https://doi.org/10.1088/1748-9326/5/1/014010.

3. NASA POWER (Prediction Of Worldwide Energy Resources), NASA Langley Research Center (2023). *POWER Data Access Viewer (Version 2.1)* https://power.larc.nasa.gov/data-access-viewer/

4. ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), ICRISAT District-Level Data Portal. https://data.icrisat.org

5. GADM – Global Administrative Areas, GADM database of Global Administrative Areas, version 3.6 (2018). https://gadm.org

6. Chat GPT :- Chat shares from GPT + account of ANU PAL.
A. Panel Data Models Explanation:-
https://chatgpt.com/share/68773906-2c14-8012-8cfc-871cf99d7413

B. Excel Data Summary:-
https://chatgpt.com/share/68773a97-7a8c-8012-b636-0eb18fa51ce3

C. Panel Data Summary:-
https://chatgpt.com/share/68773abc-890c-8012-af31-33a2882b8977

D. Crop Yield Map Python:-
https://chatgpt.com/share/68773b0b-fb74-8012-bd2b-478cd89ddbf5

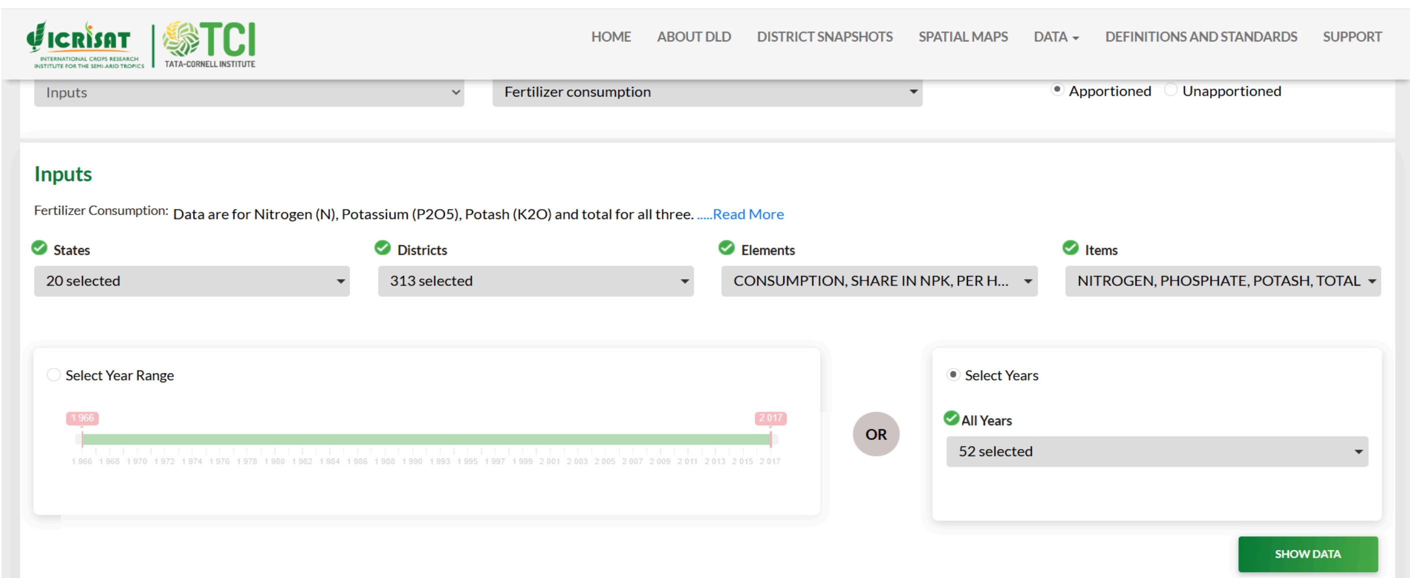
E. India Precipitation Data Download:-
https://chatgpt.com/share/68773b52-4fac-8012-b25f-c8b0556b72ed

F.

7 Appendix

Data Collection

Fertilizer consumption data



Few rows of Data

Data

VISUALIZATION | EXPORT AS CSV

Show 10 entries

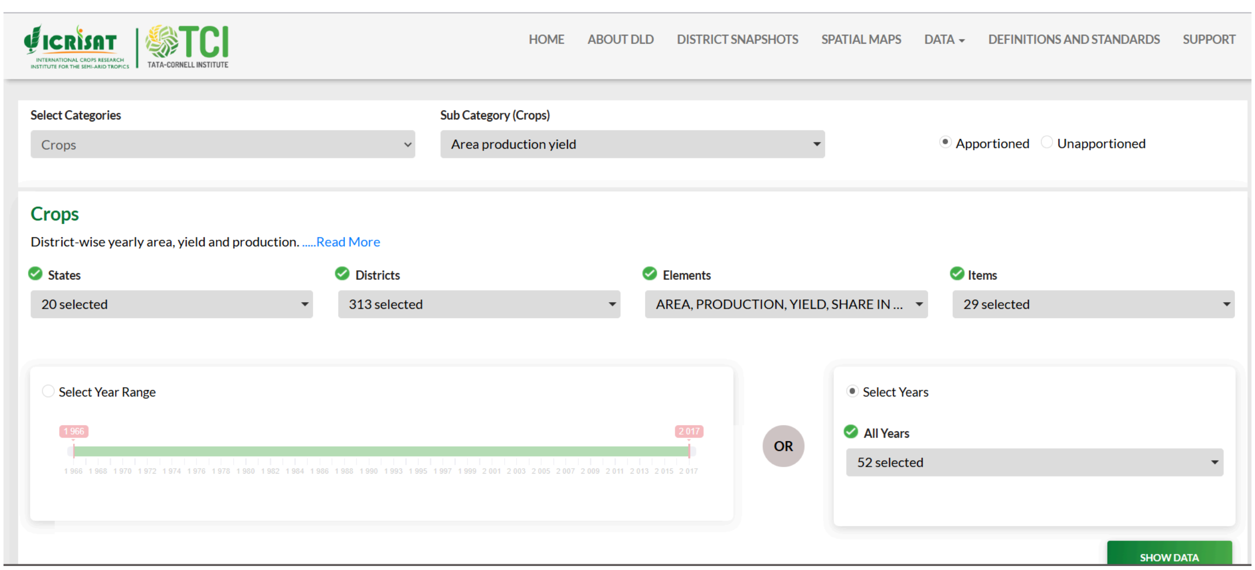
Search:

Dist Code	Year	State Code	State Name	Dist Name	NITROGEN CONSUMPTION (tons)	NITROGEN SHARE IN NPK (Percent)	NITROGEN PER HA OF NCA (Kg per ha)	NITROGE
1	1966	14	Chhattisgarh	Durg	1375.0	80.74	1.38	1.27
1	1967	14	Chhattisgarh	Durg	1516.0	78.79	1.5	1.17
1	1968	14	Chhattisgarh	Durg	3042.0	85.84	2.95	2.27
1	1969	14	Chhattisgarh	Durg	4131.0	82.39	3.99	3.08
1	1970	14	Chhattisgarh	Durg	4594.0	67.92	4.44	3.32
1	1971	14	Chhattisgarh	Durg	7152.0	66.9	6.87	5.11
1	1972	14	Chhattisgarh	Durg	8345.0	64.58	8.01	6.09
1	1973	14	Chhattisgarh	Durg	4679.0	64.68	4.47	3.31
1	1974	14	Chhattisgarh	Durg	2704.0	65.5	2.6	2.32
1	1975	14	Chhattisgarh	Durg	4673.0	81.8	4.39	3.38

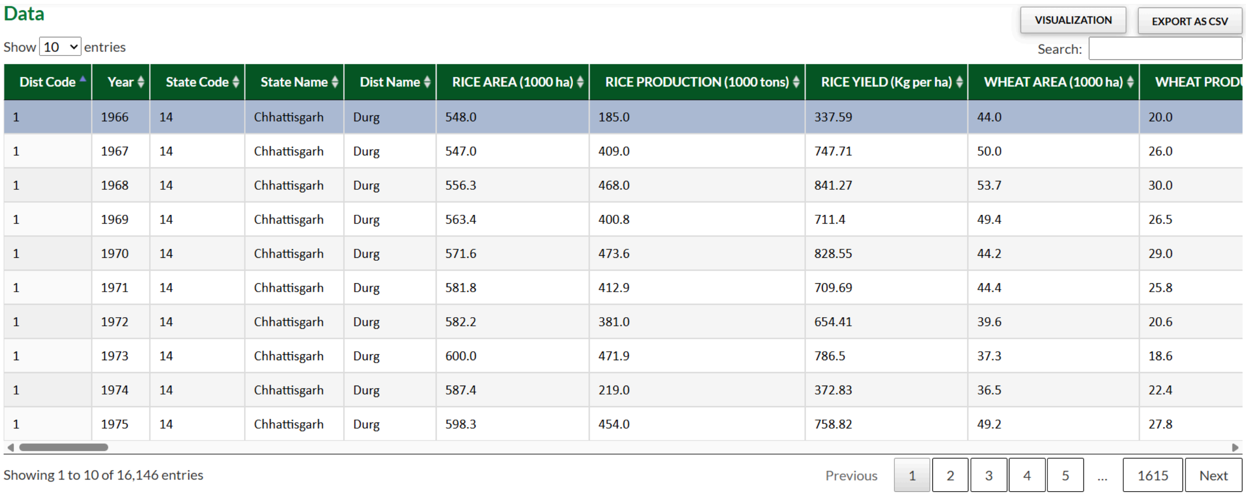
Showing 1 to 10 of 16,047 entries

Previous 1 2 3 4 5 ... 1605 Next

Crop Yield Data



Few rows of Data



Data

Show 10 entries

VISUALIZATION EXPORT AS CSV

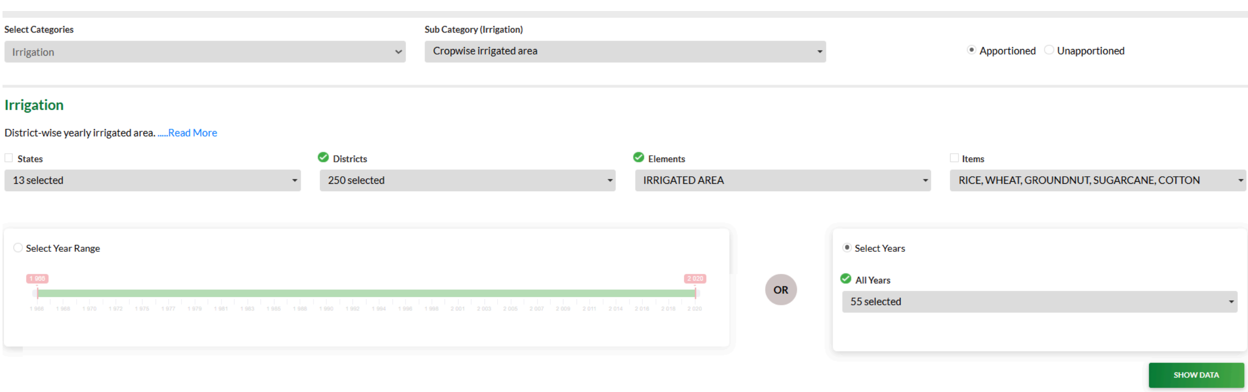
Search:

Dist Code	Year	State Code	State Name	Dist Name	RICE AREA (1000 ha)	RICE PRODUCTION (1000 tons)	RICE YIELD (Kg per ha)	WHEAT AREA (1000 ha)	WHEAT PROD
1	1966	14	Chhattisgarh	Durg	548.0	185.0	337.59	44.0	20.0
1	1967	14	Chhattisgarh	Durg	547.0	409.0	747.71	50.0	26.0
1	1968	14	Chhattisgarh	Durg	556.3	468.0	841.27	53.7	30.0
1	1969	14	Chhattisgarh	Durg	563.4	400.8	711.4	49.4	26.5
1	1970	14	Chhattisgarh	Durg	571.6	473.6	828.55	44.2	29.0
1	1971	14	Chhattisgarh	Durg	581.8	412.9	709.69	44.4	25.8
1	1972	14	Chhattisgarh	Durg	582.2	381.0	654.41	39.6	20.6
1	1973	14	Chhattisgarh	Durg	600.0	471.9	786.5	37.3	18.6
1	1974	14	Chhattisgarh	Durg	587.4	219.0	372.83	36.5	22.4
1	1975	14	Chhattisgarh	Durg	598.3	454.0	758.82	49.2	27.8

Showing 1 to 10 of 16,146 entries

Previous 1 2 3 4 5 … 1615 Next

Irrigation Data - First 10 row



Data

VISUALIZATION EXPORT AS CSV

Show 10 entries

Search:

Dist Code	Year	State Code	State Name	Dist Name	RICE IRRIGATED AREA (1000 ha)	WHEAT IRRIGATED AREA (1000 ha)	GROUNDNUT IRRIGATED AREA (1000 ha)	SUGARCANE IRRIGATED AREA (1000 ha)	COTTON IRRIGATED AREA (1000 h
7	1966	6	Madhya Pradesh	Jabalpur	8.2	1.2	0.0	0.2	0.0
7	1967	6	Madhya Pradesh	Jabalpur	8.7	1.5	0.0	0.3	0.0
7	1968	6	Madhya Pradesh	Jabalpur	10.7	3.8	0.0	0.3	0.0
7	1969	6	Madhya Pradesh	Jabalpur	12.7	5.9	0.0	0.4	0.0
7	1970	6	Madhya Pradesh	Jabalpur	11.6	8.3	0.0	0.4	0.0
7	1971	6	Madhya Pradesh	Jabalpur	12.5	13.2	0.0	0.4	0.0
7	1972	6	Madhya Pradesh	Jabalpur	11.4	12.9	0.0	0.4	0.0
7	1973	6	Madhya Pradesh	Jabalpur	5.1	16.2	0.0	0.4	0.0
7	1974	6	Madhya Pradesh	Jabalpur	10.1	13.9	0.0	0.4	0.0
7	1975	6	Madhya Pradesh	Jabalpur	9.5	22.9	0.0	0.4	0.0

Showing 1 to 10 of 12,935 entries

Previous 1 2 3 4 5 ... 1294 Next