Exploratory (Spatial) Data Analysis for Rice Yields in the Indo-Gangetic Plains – 2018-19

- **Sumedha Shukla (PhD21402)**
- **Utkarsh Arora (2020143)**

Data Description

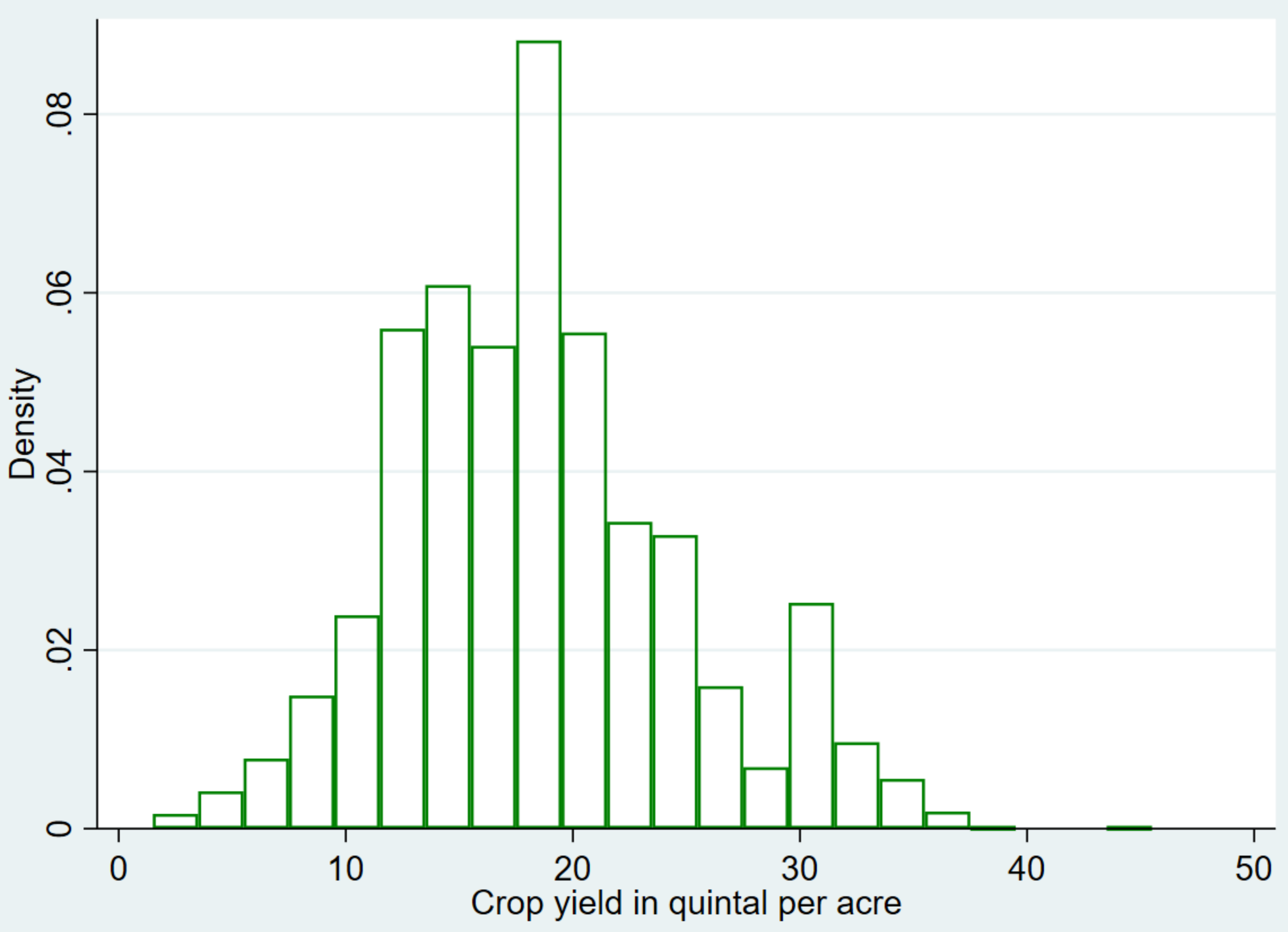
- Plot-level data on rice production practices
- Pertains to the 2018 Kharif season.
- Covers eight Indian states: Andhra Pradesh, **Bihar**, Chhattisgarh, **Haryana**, Odisha, **Punjab, Uttar Pradesh** and **West Bengal**.
- All plots are geo-coded with an accuracy of <10 metres.
- Provided by the International Maize and Wheat Improvement Center (CIMMYT).
- Detailed information on sampling design can be found in (Ajay et al. 2022)
- Similar data for wheat is also available (Ajay et al. 2021)
- **Our focus:** Rice Production

Presentation Outline:

- In the first part, we attempt to understand the sources/ drivers of spatial heterogeneity in yields by exploring the differences in yields by soil texture; perceived soil quality; social class; seed variety; land ownership; season; residue burning activity; and soil health card information.
- In the second part, we formalize the decision regarding spatial stationarity of yields and examine the (dis)similarity in input application decisions (i.e., one of the primary determinants of yields) across the (potentially) stationary zones.
- Finally, using the results in parts one and two, we calculate an experimental variogram for yields and further, model the yield variogram incorporating the (previously examined) drivers of yield in our de-trending model in order to comment on the spatial dependence in yields for any value of the lag.

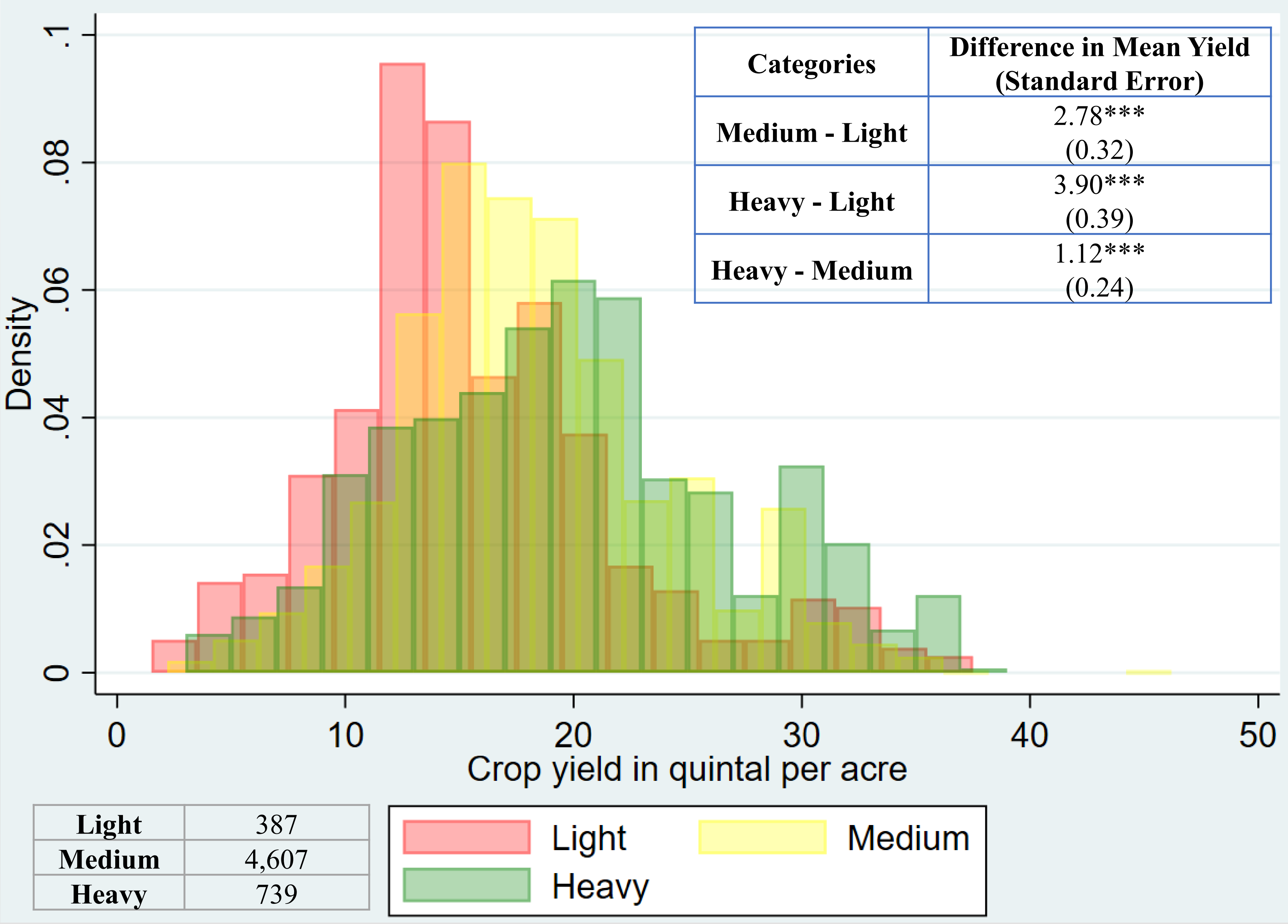
Note: Developing a spatial regression model for yields would also require an exploration of spatial patterns (clustering or heterogeneity) in the (potential) drivers of yield to guide the eventual choice of a model. Since we do not present the spatial regression results in this report, the maps for these drivers have been moved to the appendix.

N (Plots)	Mean	SD	Minimum	P25	P50	P75	P90	Maximum
5733	18.41	6.21	1.5	14.3	18	21.81	27.2	45.05



Crop Yield Density (Raw)

Crop Yield by Soil Texture

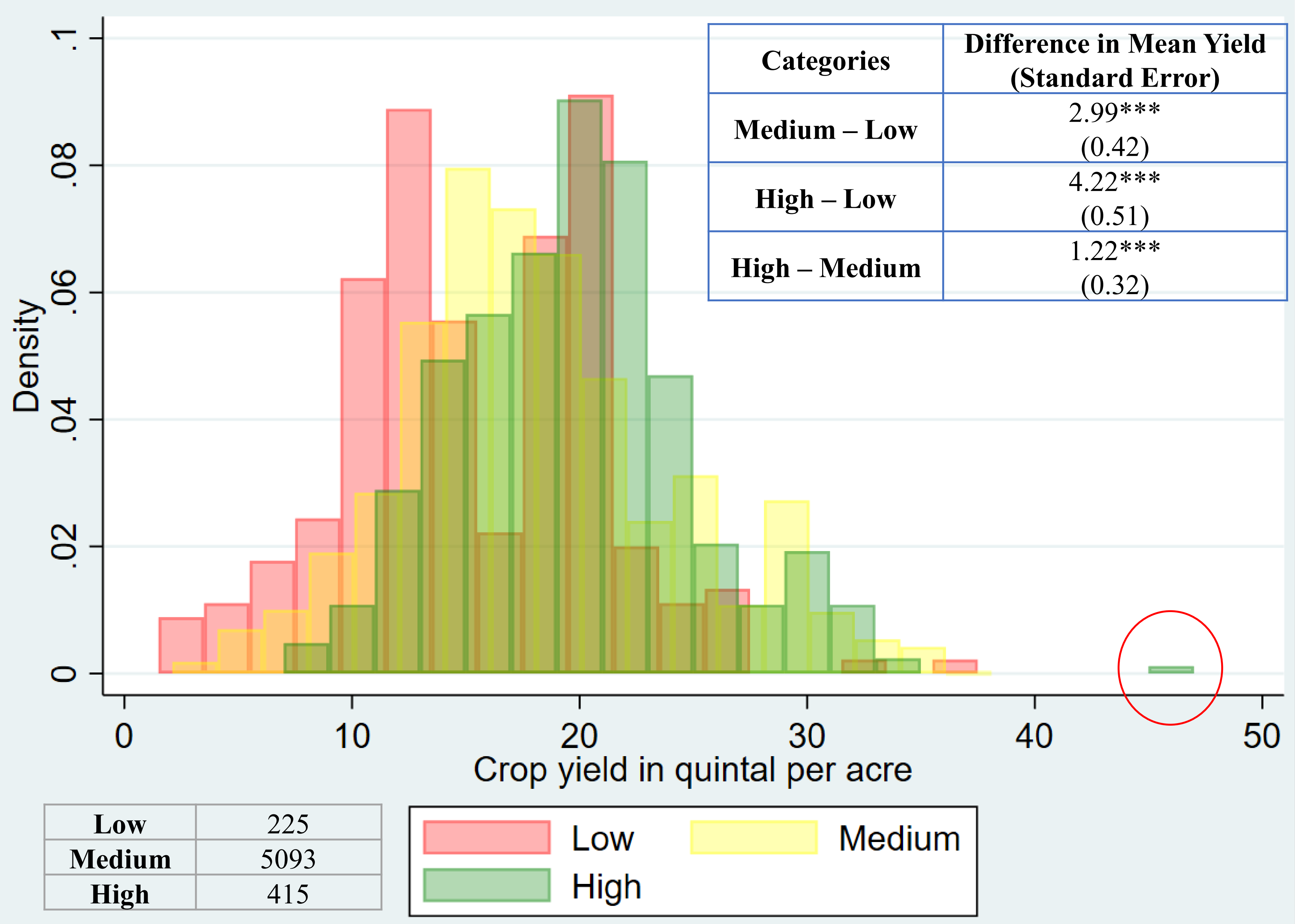


Categories	Difference in Mean Yield (Standard Error)
Medium - Light	2.78*** (0.32)
Heavy - Light	3.90*** (0.39)
Heavy - Medium	1.12*** (0.24)

Light	387
Medium	4,607
Heavy	739

Clearly, plots with lighter soils have lower yields as compared to plots with heavier soils.

Crop Yield by Perceived Soil Quality



Categories	Difference in Mean Yield (Standard Error)
Medium – Low	2.99*** (0.42)
High – Low	4.22*** (0.51)
High – Medium	1.22*** (0.32)

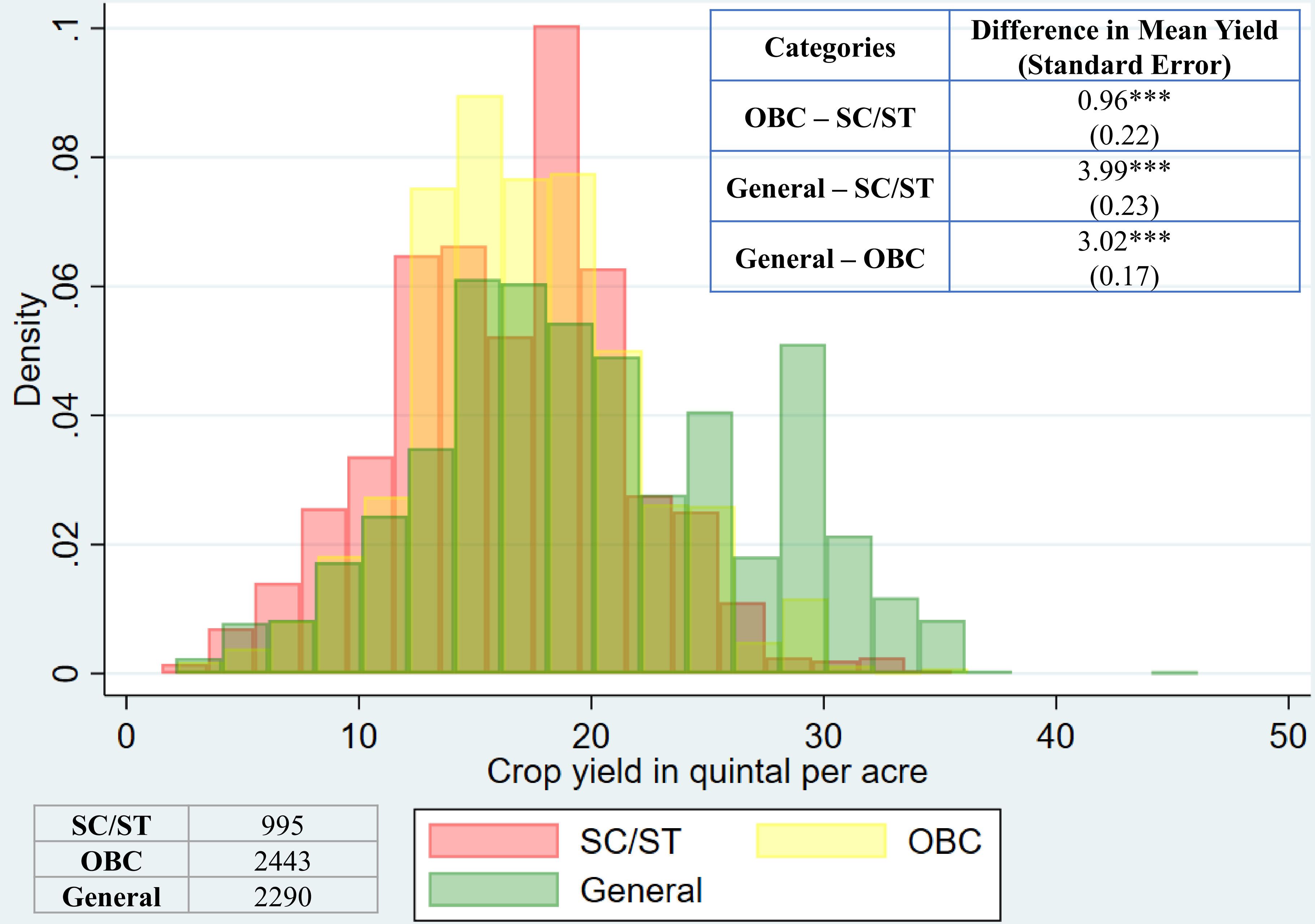
Low	225
Medium	5093
High	415

Clearly, plots with lower (perceived) soil quality have lower yields as compared to plots with better (perceived) soils.

Further, perceived soil quality does not display a clear relationship with yields.

- For example, the plots with lower quality soils exhibit a bi-modal yield distribution with higher yield values coinciding with high quality soils as well.

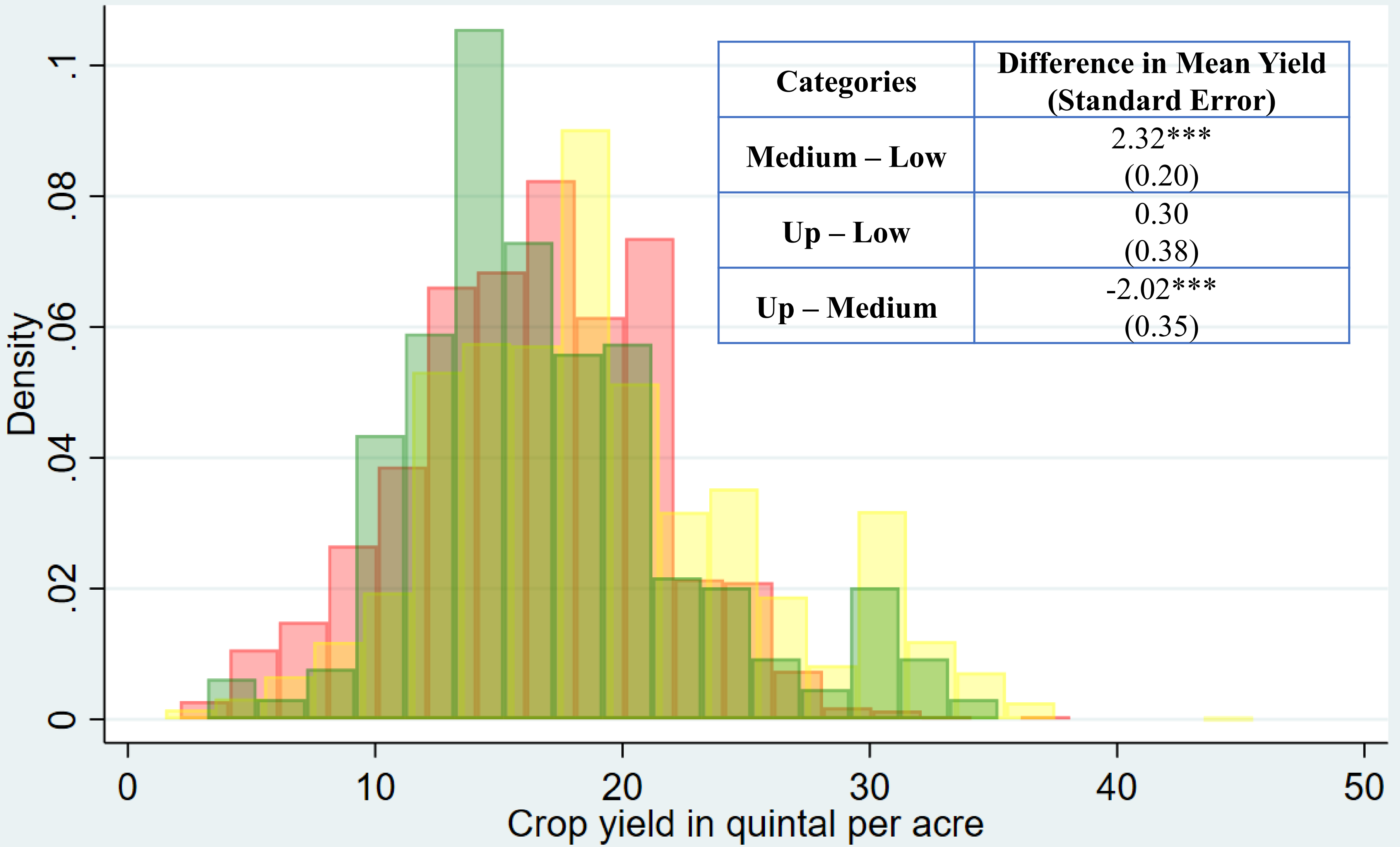
Crop Yield by Social Categories



Categories	Difference in Mean Yield (Standard Error)
OBC – SC/ST	0.96*** (0.22)
General – SC/ST	3.99*** (0.23)
General – OBC	3.02*** (0.17)

SC/ST	995
OBC	2443
General	2290

Crop Yield by Soil Drainage Class



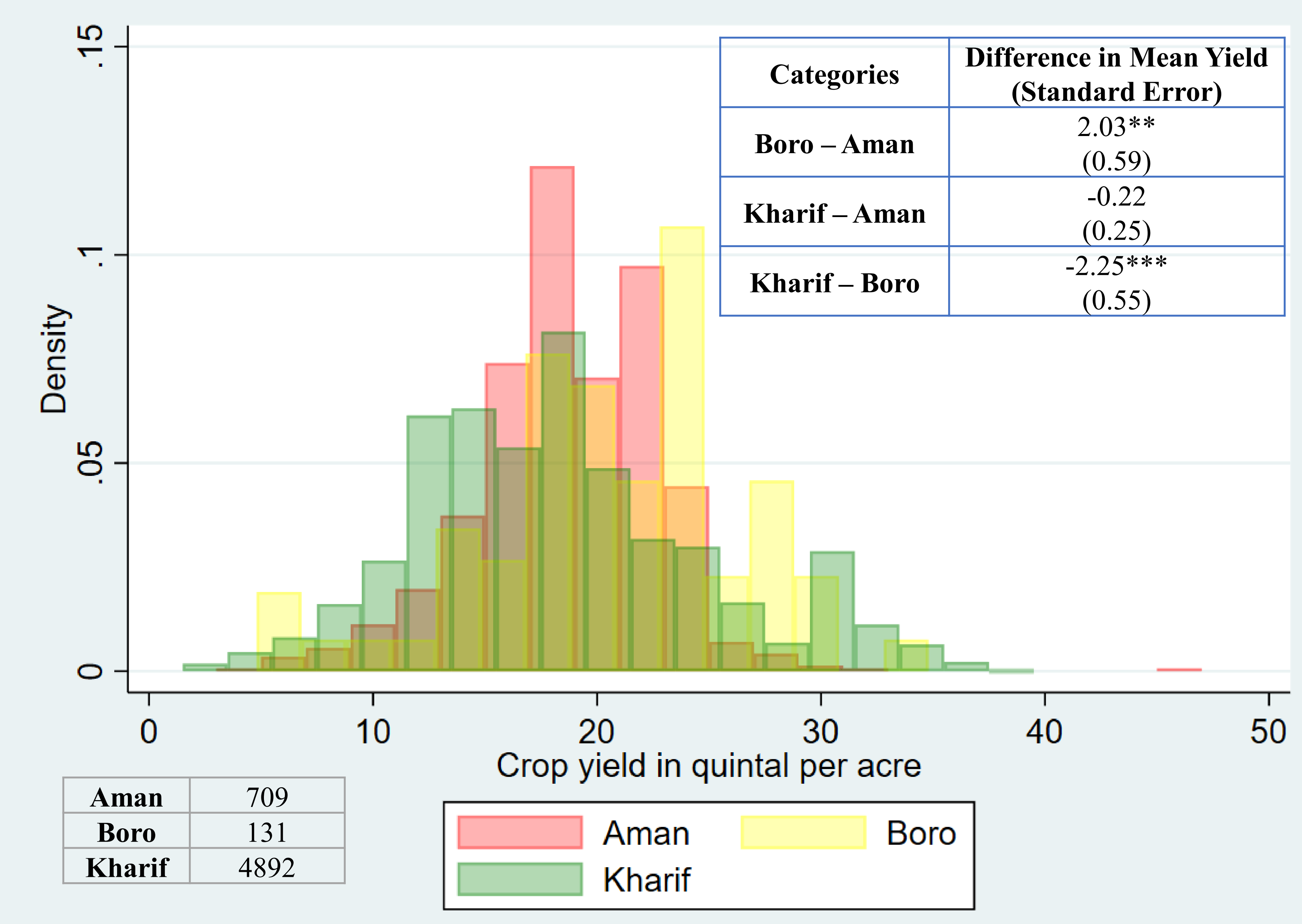
Categories	Difference in Mean Yield (Standard Error)
Medium – Low	2.32*** (0.20)
Up – Low	0.30 (0.38)
Up – Medium	-2.02*** (0.35)

Low	1188
Medium	4223
Up	322

Upland soils are generally "well-drained" and do not have as much surface water accumulation – which is important for the rice crop.

At the same time, up-land rice varieties are (largely) more drought tolerant as compared to low-land varieties.

Crop Yield by Season (Aman and Boro occur only in West Bengal)

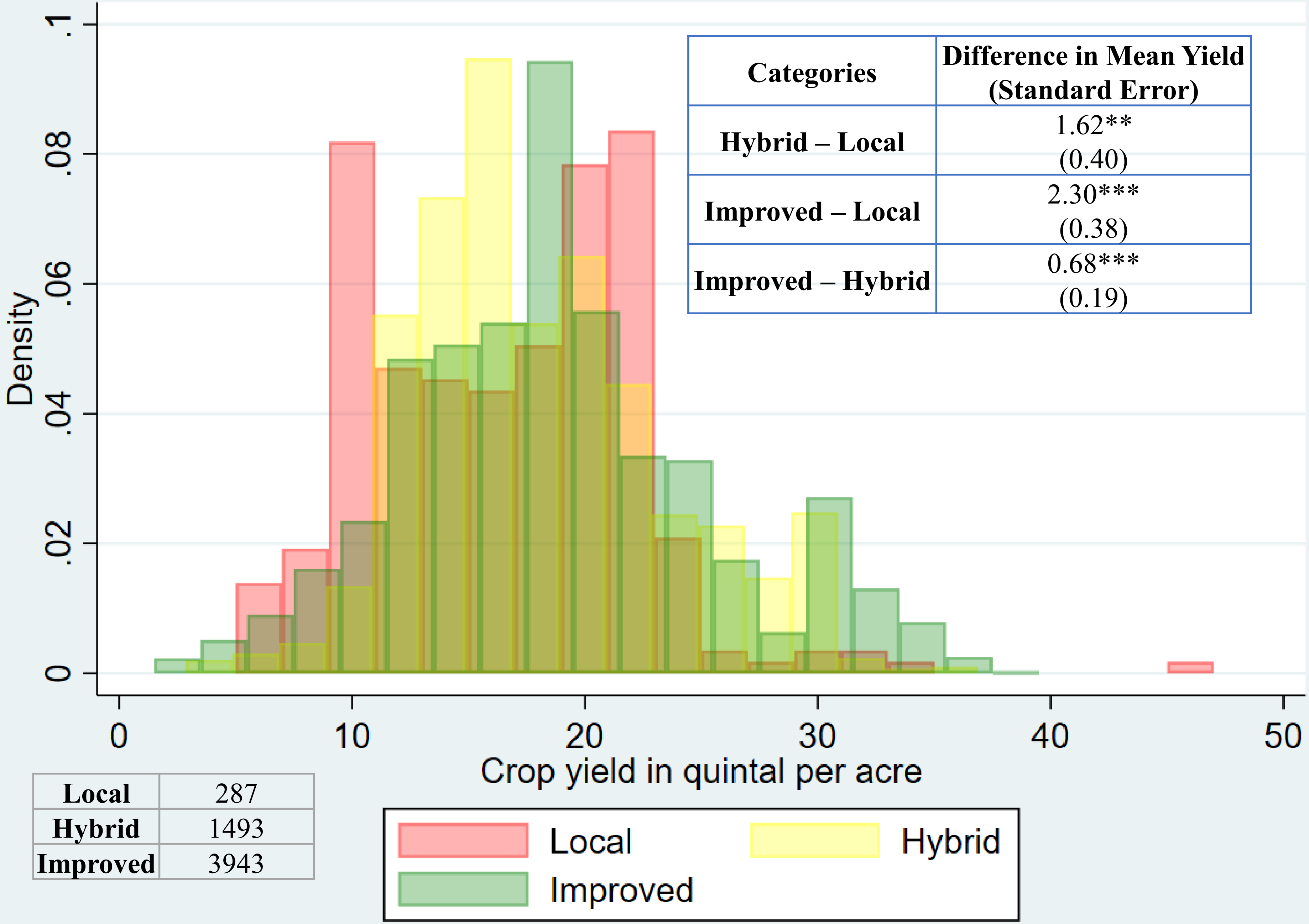


Categories	**Difference in Mean Yield (Standard Error)**
Boro – Aman	2.03** (0.59)
Kharif – Aman	-0.22 (0.25)
Kharif – Boro	-2.25*** (0.55)

Aman	709
Boro	131
Kharif	4892

Kharif and Boro yields are different on-average.
But, this difference could also be due to the spatial difference whereby Boro records correspond to West Bengal.

Crop Yield by Variety

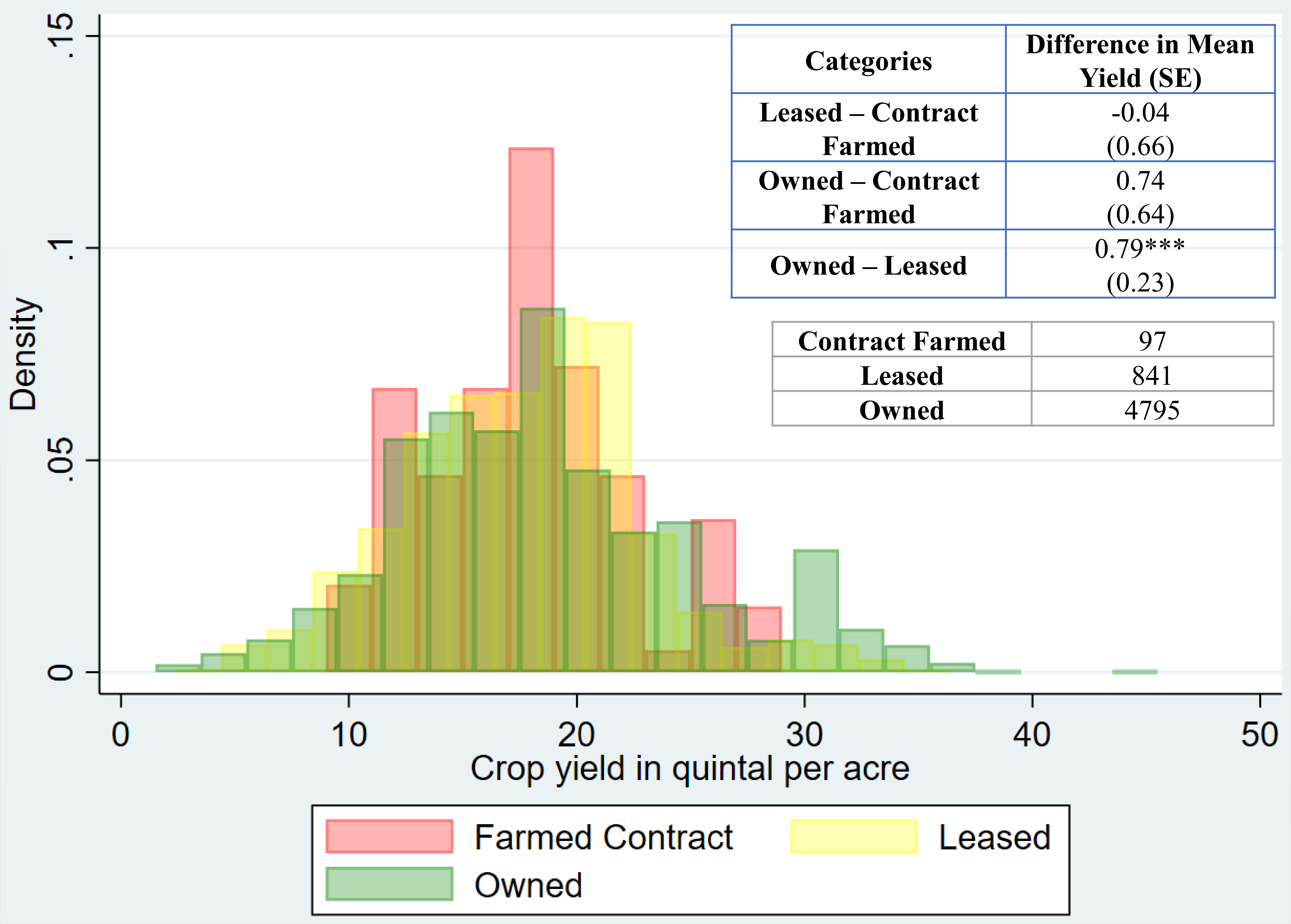


Categories	Difference in Mean Yield (Standard Error)
Hybrid – Local	1.62** (0.40)
Improved – Local	2.30*** (0.38)
Improved – Hybrid	0.68*** (0.19)

Local	287
Hybrid	1493
Improved	3943

Rice yields from hybrid and improved quality seeds are (on-average) higher than the yields from local seeds.

Crop Yield by Land Ownership



Categories	Difference in Mean Yield (SE)
Leased – Contract Farmed	-0.04 (0.66)
Owned – Contract Farmed	0.74 (0.64)
Owned – Leased	0.79*** (0.23)

Contract Farmed	97
Leased	841
Owned	4795

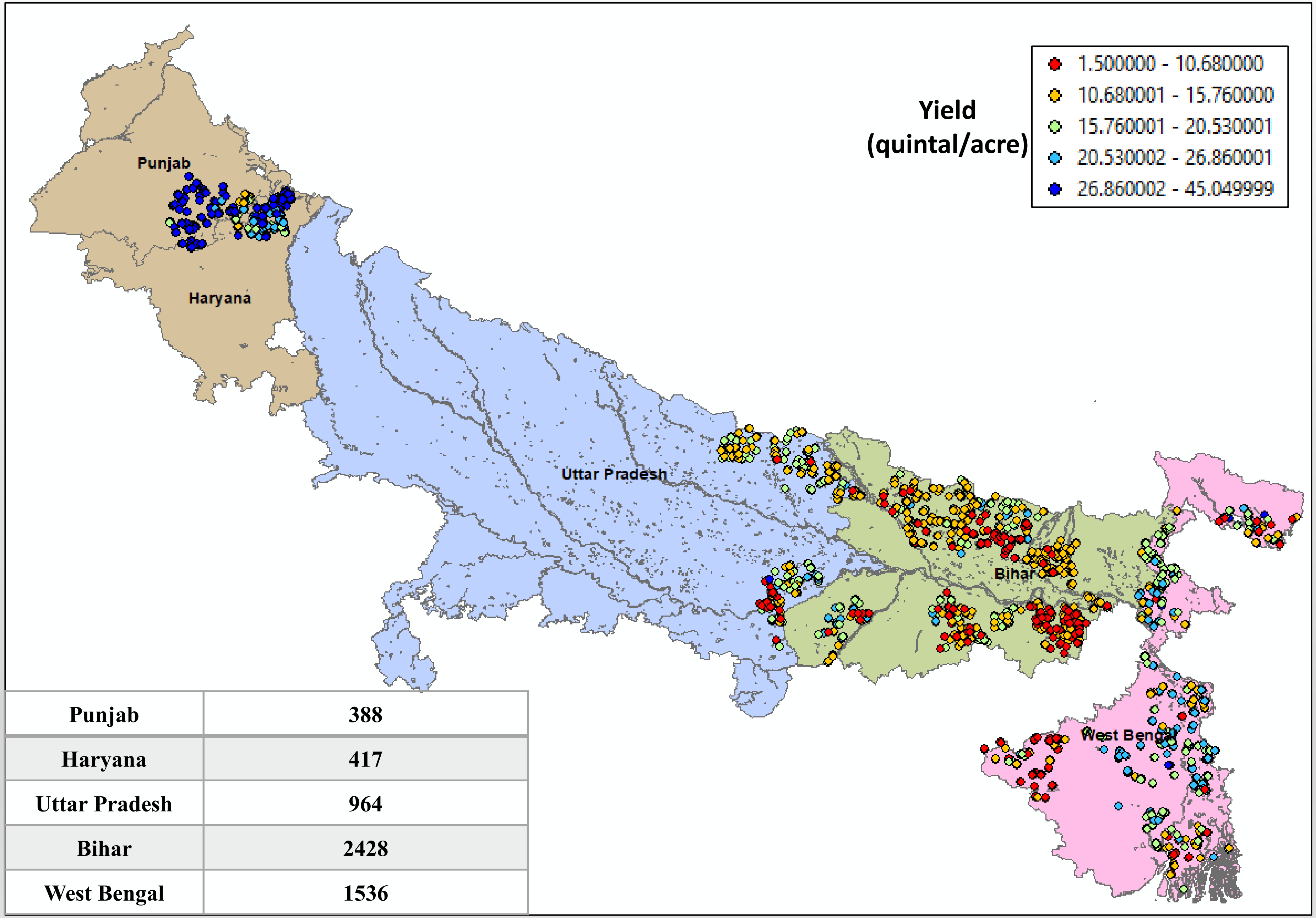
What have we learnt?

When modelling yields, we must account for:

- Soil Texture
- Perceived Soil Quality
- Social Categories
- Land Ownership
- For weed and pest management, we will rely on the application decision and not reported severity.

Now, we move to understanding spatial stationarity. For this, we must first ask, are yields spatially heterogenous – i.e., do we observe a trend in yields over space?

Spatial Heterogeneity in Crop Yields?



Punjab	388
Haryana	417
Uttar Pradesh	964
Bihar	2428
West Bengal	1536

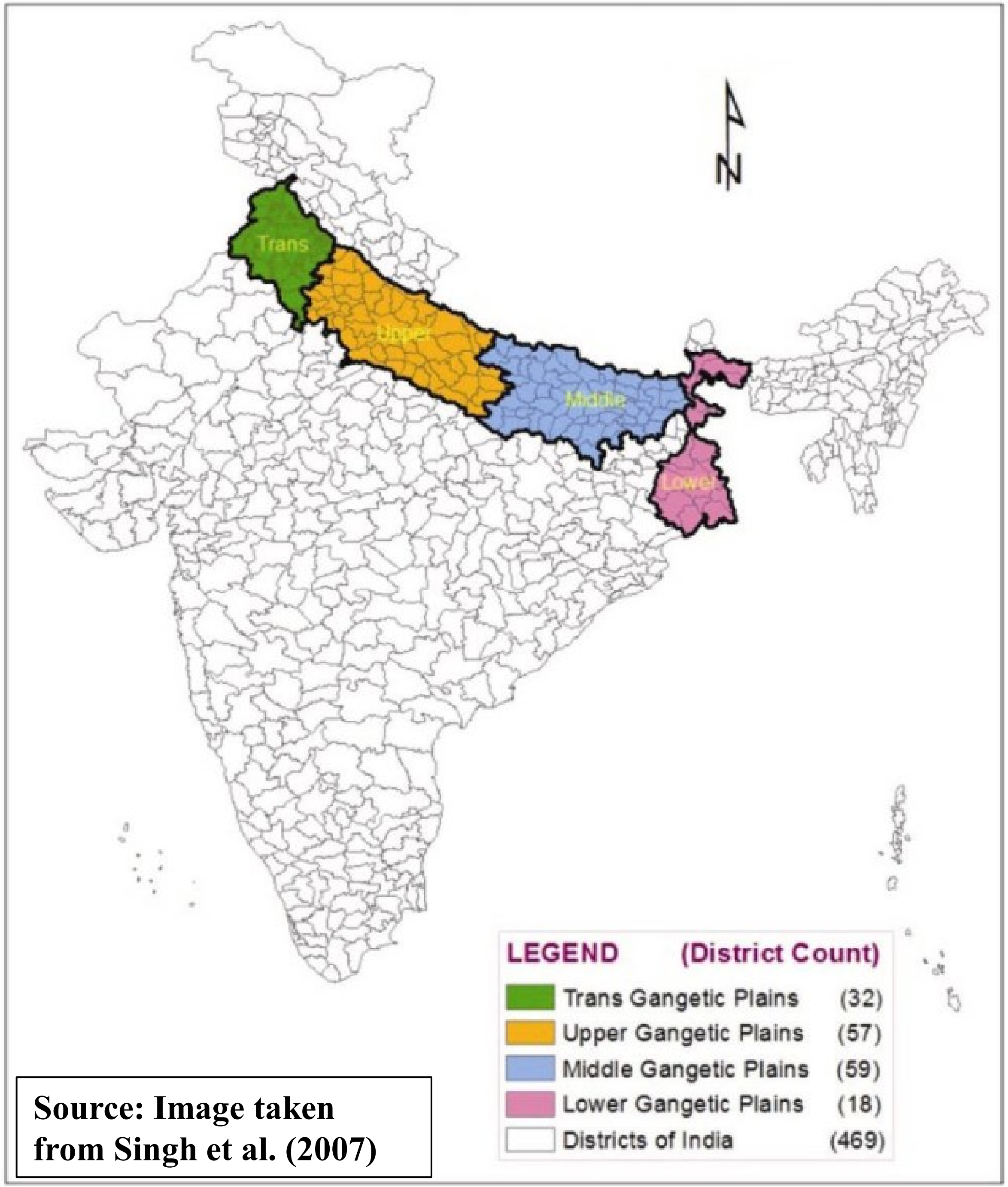
Yes!

Plots in Punjab and Haryana seem to have higher yields as compared to plots in Bihar and West Bengal.

Moreover, yields in certain seasons in West Bengal are higher than yields in other seasons.

This spatial heterogeneity indicates that yields may not be stationarity across the entire region comprising the Indo Gangetic Plains.

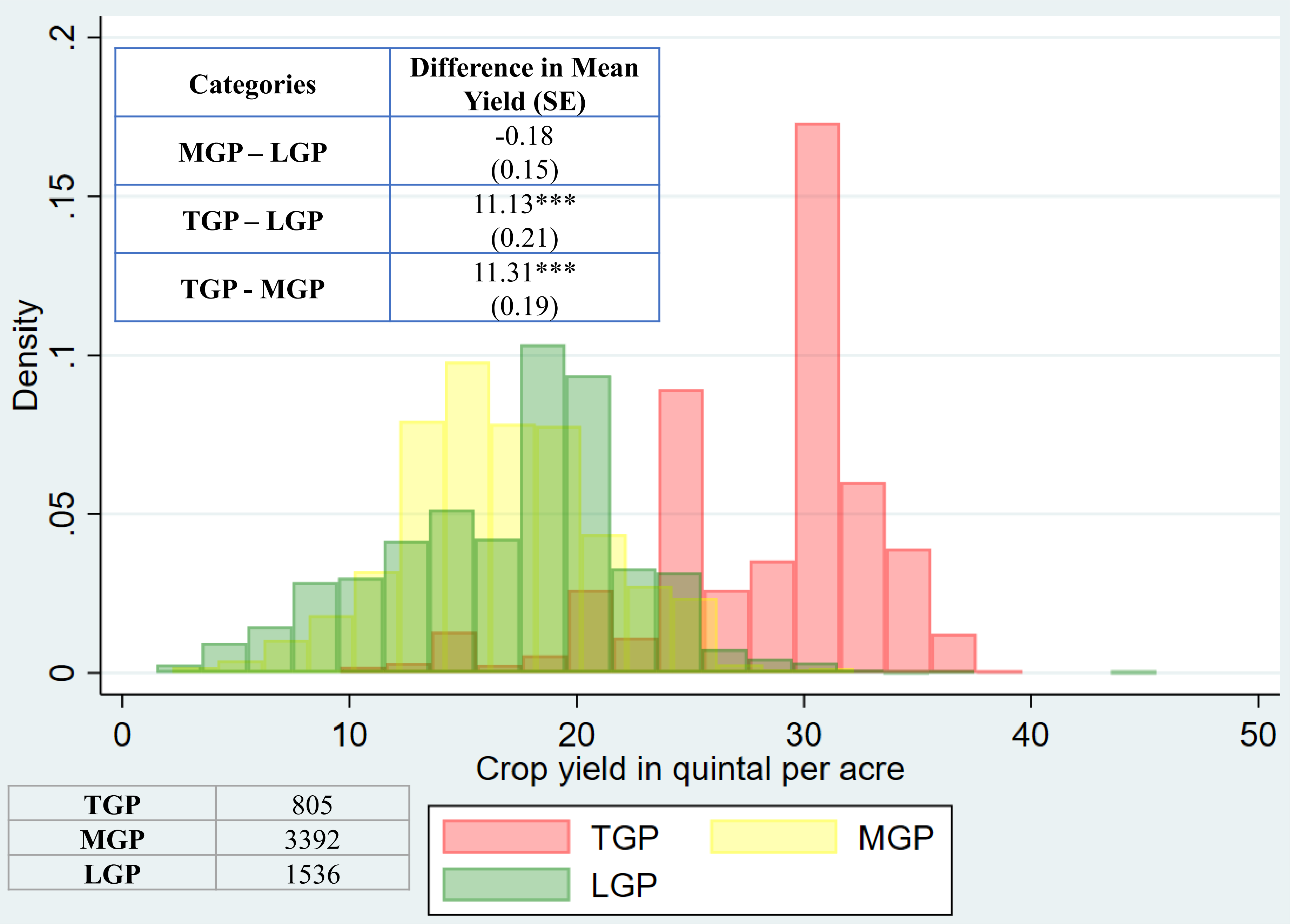
Instead, we will have to look at sub-agro ecological zones to arrive at regions where yields are expected to be stationary.



Source: Image taken from Singh et al. (2007)

According to this classification, in our dataset, plots in **Punjab and Haryana** belong to the **trans-Gangetic plains** whereas plots in **Uttar Pradesh and Bihar** belong to the **middle-Gangetic plains.** Finally, plots in **West Bengal** fall in the **lower-Gangetic plains** region.

Crop Yield by Sub-Agroecological Zones

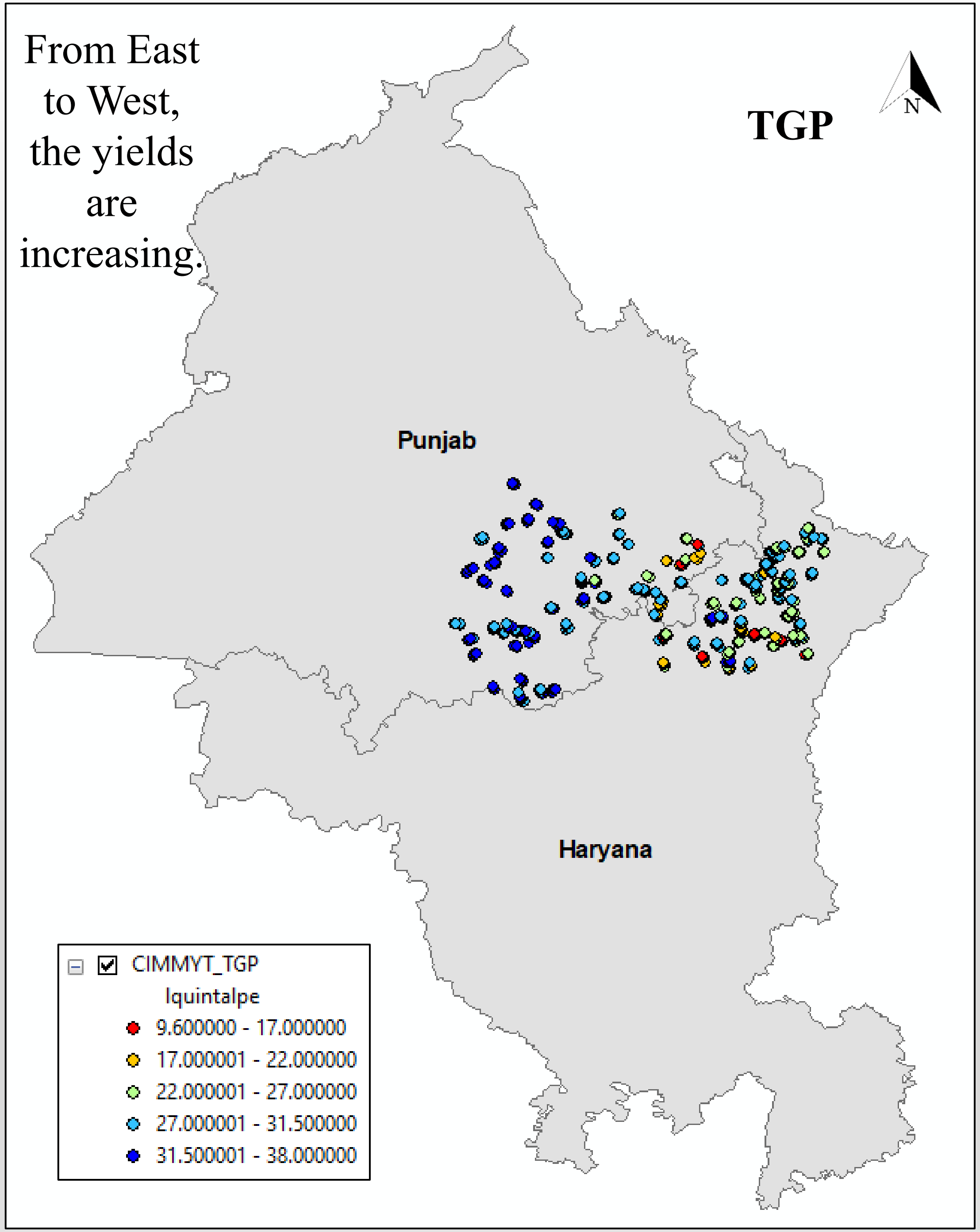


Categories	Difference in Mean Yield (SE)
MGP – LGP	-0.18 (0.15)
TGP – LGP	11.13*** (0.21)
TGP - MGP	11.31*** (0.19)

TGP	805
MGP	3392
LGP	1536

Clearly, yields are heterogenous across sub-agroecological zones. But, what about *within* the zone?

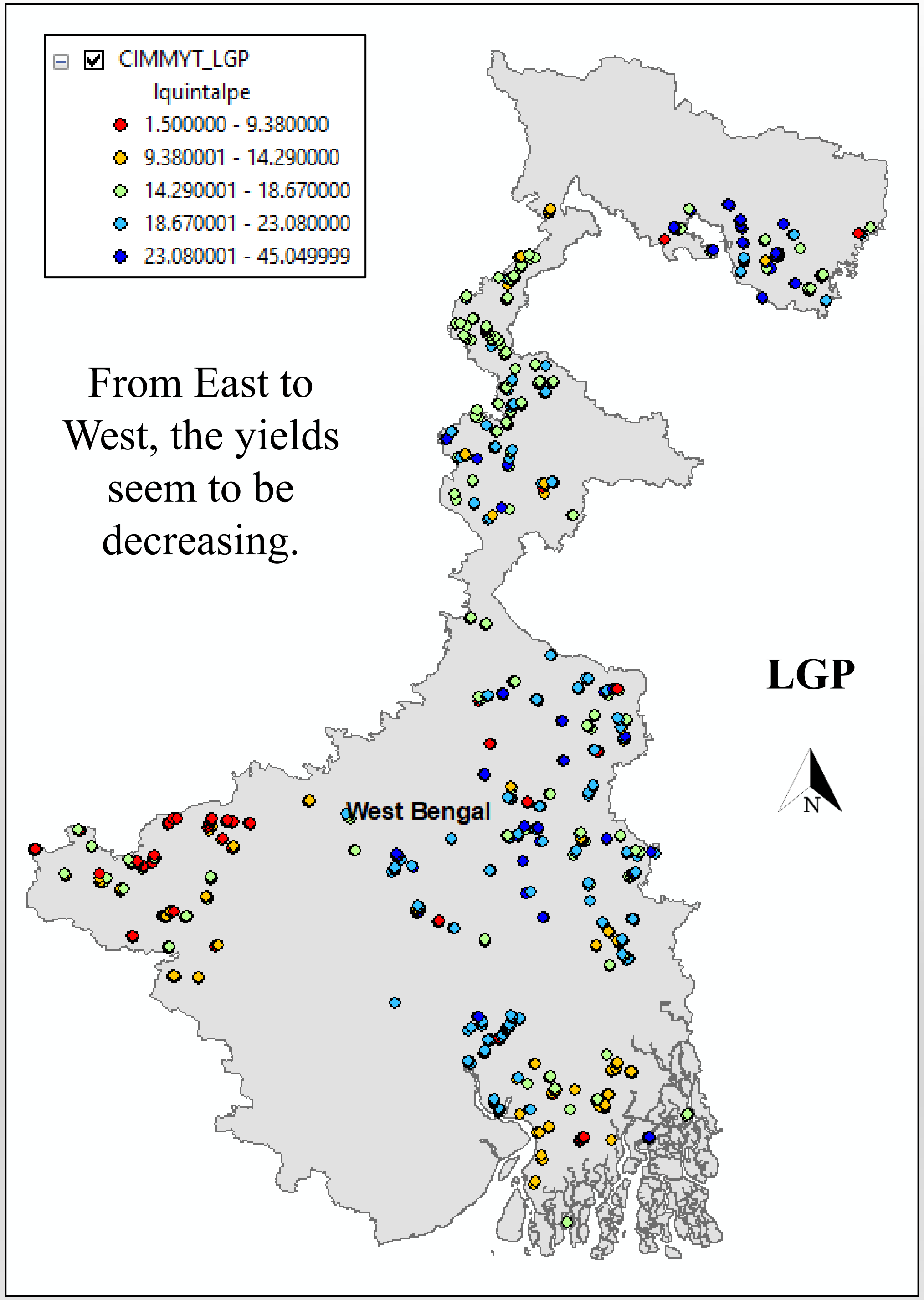
From East to West, the yields are increasing.



Plots with high yields seem to be clustered together.

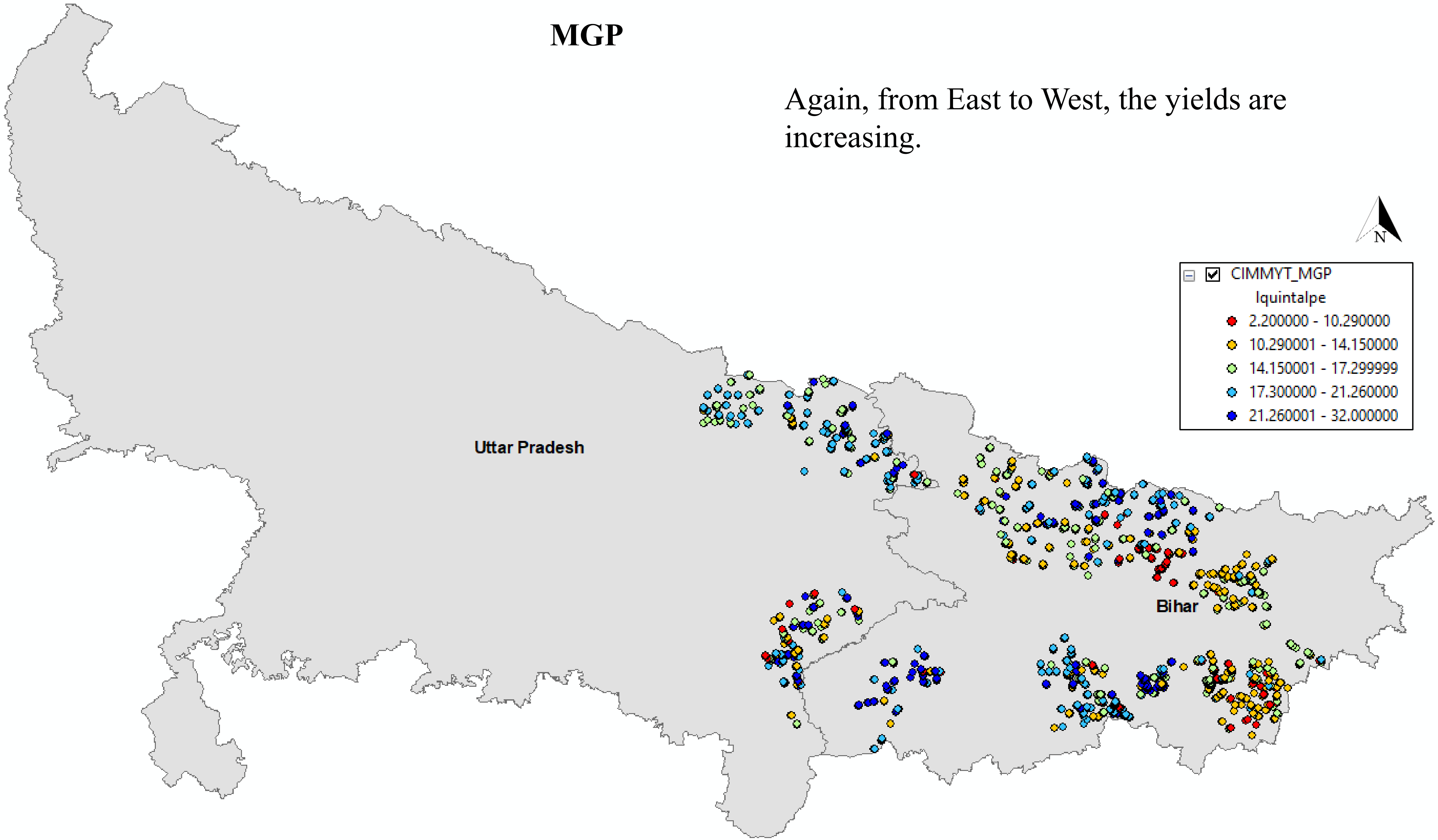
But, do we observe similar spatial dependence after accounting for the trend?

From East to West, the yields seem to be decreasing.



MGP

Again, from East to West, the yields are increasing.



What have we learnt?

Evidently, even within the sub agroecological zone, yields are non-stationary with a significant trend by different directions.

In order to further understand the sources of this trend, we will examine the influence of other additional (non-spatial) drivers of yield.

Now, we move to understanding input application. For this, we first present a chart explaining the levels of input application timeline and further the dis(similarity) in observed input application across the three zones.

Examination of the Trend: Input Decisions

Tillage (Land Preparation)
- Type for current and previous crop.
  - Directly Sown
  - Transplanted

Crop Establishment
- Date (Sowing + Nursery Establishment)
- Determinants of Timing of Transplanting and Nursery Establishment.

Fertilization
- Basal + First + Second + Third Topping

For each Topping:
- Amount
- Days after Sowing
- Type: FYM (Wet vs. Dry), DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron

Irrigation
- Frequency
- Availability
- Source
  - Depth of Tubewell
  - Energy Source of Pump
  - Flat Pipes
- Stage of Crop Growth at which Irrigation Application Occurs

Crop Protection
- Weed and Insect Severity
- Herbicide Application Timing and Frequency
- Manual Weeding (First + Second + Third) Timing

Harvesting
- Harvest Date
- Harvest Method
- Threshing Type

Examination of the Trend: Input Application

Crop Establishment

- Date (Sowing + Nursery Establishment)

Systematic differences in the timing of sowing and nursery establishment across the three zones.

- Determinants of Timing of Transplanting and Nursery Establishment.

Nursery Establishment Month	LGP	MGP	TGP
January	49	0	0
February	7	0	0
April	2	0	0
May	146	245	785
June	1,111	2,923	17
July	136	204	1
August	2	0	0
October	1	0	0
November	15	0	0
December	61	0	0
Total	1,530	3,372	803

Month of Sowing	LGP	MGP	TGP
January	62	0	0
February	58	0	0
March	11	0	0
May	4	4	2
June	131	296	767
July	1,020	2,947	36
August	240	145	0
September	3	0	0
November	1	0	0
December	6	0	0
Total	1,536	3,392	805

Seeds and Fertilization Type:

- FYM (Farm Yard Manure) : Wet vs. Dry
- Mineral: DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron

Basal Fertilizer	Count	Percentage of Total Plots
Reported None	1196	20.86
DAP	3337	73.55
Urea	1576	34.74
MoP	1382	30.46
NPK	659	14.53
ZnSO4	398	8.77
SSP	384	8.46
Other	66	1.45
NPKS	19	0.42
Boron	15	0.33
TSP	9	0.20
Gypsum	4	0.09
Potash	1	0.02

First Top Dressing	Count	Percentage
Not Reported	788	14.32
Reported None	33	
Urea	4803	97.78
MoP	731	14.88
ZnSO4	690	14.05
DAP	287	5.84
SSP	206	4.19
NPK	169	3.44
Other	77	1.57
NPKS	26	0.53
TSP	7	0.14
Boron	4	0.08
Gypsum	3	0.06

Second Top Dressing	Count	Percentage
Reported None	851	16.85
Not Reported	115	
Urea	4654	97.63
DAP	389	8.16
MoP	345	7.24
ZnSO4	236	4.95
NPK	105	2.20
SSP	57	1.20
Other	35	0.73
Gypsum	4	0.08
Boron	2	0.04
NPKS	1	0.02
TSP	1	0.02

Third Top Dressing	Count	Percentage
Not Reported	937	69.72
Reported None	3060	
Urea	1727	99.48
ZnSO4	26	1.50
MoP	23	1.32
NPK	19	1.09
DAP	15	0.86
SSP	4	0.23
Other	3	0.17
Boron	1	0.06

- Across different stages of fertilizer application, the focus is on Phosphorus, Nitrogen, Potassium, and Zinc.

Question: Does the nutrient type targeted and/or the fertilization frequency vary across the plain region?

Summary Nutrient Application (per acre) (After Outlier Removal)

Zone	LGP (N = 1425)								MGP (N = 3352)								TGP (N = 789)							
Fertilization	Mean	SD	Min	P25	P50	P75	P90	Max	Mean	SD	Min	P25	P50	P75	P90	Max	Mean	SD	Min	P25	P50	P75	P90	Max
Frequency	2.26	0.84	0	2	2	3	3	4	2.84	0.63	0	2	3	3	4	4	3.32	0.61	0	3	3	4	4	4
Nitrogen (kg/acre)	27.59	16.90	0	15.7	24.9	38.0	50.9	83.0	49.60	13.04	0	40.7	49.7	58.0	66.9	83.95	61.44	12.11	0	54.08	62.1	70.2	78.2	82.8
Phosphorus (kg/acre)	16.42	12.28	0	8.3	13.8	23.0	32.2	60.8	18.95	9.27	0	14.6	20.4	25.6	29.7	38.98	6.95	10.14	0	0	0	20.7	23	41.4
Potassium (kg/acre)	15.59	14.95	0	0	12.59	25.82	36.0	61.26	4.26	6.18	0	0	0	8.89	13.33	29.03	0.31	2.15	0	0	0	0	0	30
Zinc (kg/acre)	0.11	0.51	0	0	0	0	0	5.6	0.51	1.06	0	0	0	0	2.1	6.0	1.37	1.41	0	0	1.85	1.85	3.7	5.55

- Frequency of fertilizer application is highest in the TGP.
- Nitrogen application is high across all three zones.
- Phosphorus application is concentrated in the LGP and MGP.
- Potassium application is concentrated in the LGP. Barring this, the fertilizer application levels are generally lower in this region.
- Zinc application is concentrated in the TGP.

Irrigation

- Availability
- Source
  - Depth of Tubewell
  - Energy Source of Pump
  - Flat Pipes

Irrigation Availability on Largest Plot	Freq.	Percent
Yes	5,341	93.16
No	392	6.84

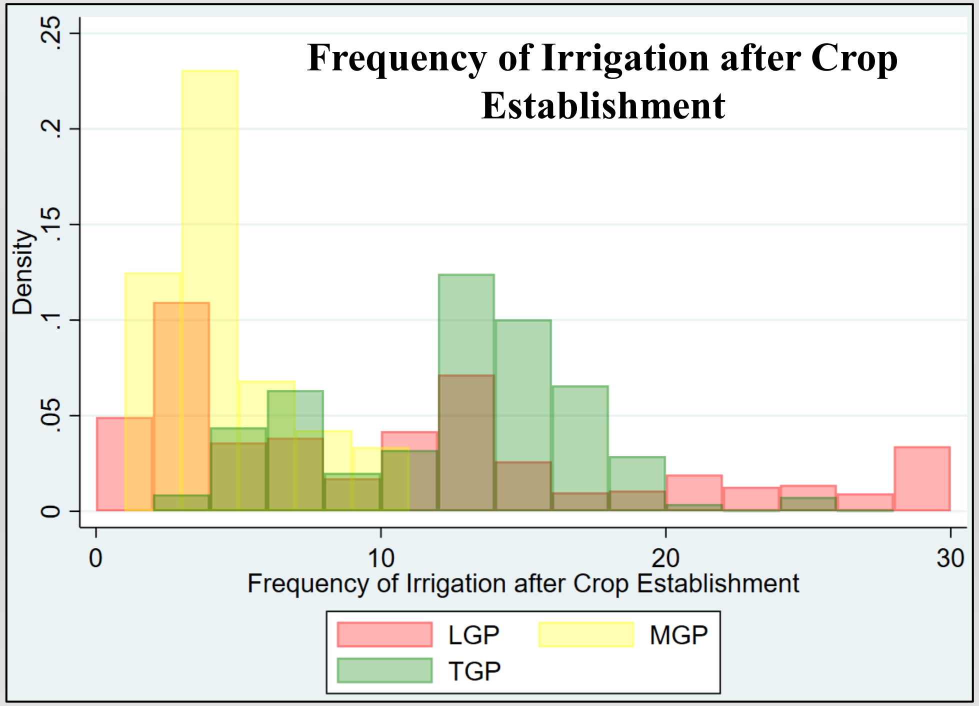
Pump Use	Count	Percent
Yes	4,689	87.81
No	651	12.19

Energy Source for Pump	Frequency	Percent
Both Electric & Diesel	113	2.41
Diesel	2,904	61.93
Electricity	1,672	35.66

Flat Pipes	Count	Percent
Yes	3,974	74.45
No	1,364	25.55

Source	N	Percentage
Deep Tube Well	2440	45.68
Shallow Tube Well	2212	41.42
Canal	839	15.71
Pond	222	4.16
River	180	3.37
Lift	88	1.65
Dugwell	26	0.49
Other	15	0.28
Tank	12	0.22

Other (Rain)	4
Other (Pump)	11



What is the difference between pesticides, herbicides, and insecticides?
(According to the United States Environmental Protection Agency (USEPA))
Pesticides are chemicals that may be used to kill fungus, bacteria, insects, plant diseases, snails, slugs, or weeds among others. These chemicals can work by ingestion or by touch and death may occur immediately or over a long period of time.

Insecticides are a type of pesticide that is used to specifically target and kill insects. Some insecticides include snail bait, ant killer, and wasp killer.

Herbicides are used to kill undesirable plants or "weeds". Some herbicides will kill all the plants they touch, while others are designed to target one species.

Application Decision	Insecticide Application	Pesticide Application	Herbicide Application
Yes	2,094	1,465	2,678
No	3,638	4,267	3,054

What have we learnt?

First, input application is significantly different across the Indo-Gangetic plains region.

Second, when modelling yields, we must account for:

- Fertilizer amount (by nutrient type)
- Fertilization Frequency: This reflect basal, first, second, and third topping.
- Irrigation Application – for now, taken as an indicator.
- Irrigation Frequency: Higher frequency of irrigation does not necessarily translate into higher water application per acre.
- Weedicide, Insecticide and Herbicide Application: Here, we would also like to account for the extent of problem weeds/insects present on the plot.

Finally, we come to understanding the spatial dependence in yields through a variogram (experimental + model)

The yield model also includes weather data (i.e., temperature and rainfall).

Weather Data

- Obtained from the AreNa DHS-GIS dataset provided by IFPRI linked here.
- For each plot, we obtain weather data for the five nearest points and take the relevant averages for temperature and rainfall.

- Importantly, our temperature and rainfall variables account for the different extent of cropping seasons across the three sub-agroecological zones.
- Kharif (for TGP): June – December + January (next Year)
- Kharif (MGP): June – December
- For LGP, we have cropping throughout the year, i.e., January – December.

Identifying drivers of spatial contiguity:

In order to identify the drivers of spatial contiguity in yields, we study the (simple) linear de-trending model systematically for each sub-agroecological zone:

- **First,** yields are modelled as a function of only the location – the results show that there is significant spatial trend within each of the zones.

- **Second,** yields are modelled as a function of the location and weather (temperature and rainfall) separately and together. Since weather is highly correlated over space, accounting for only weather would suffice.

- **Third,** yields are modelled as a function of weather and inputs (fertilizer, insecticide application, irrigation application).

- **Fourth,** in addition to the explanatory variables in model 3, we now also include soil controls (soil texture, perceived soil quality, and soil drainage class)

- **Finally,** we also include farmer demographics (caste) and land ownership.

The results for the final model are presented in the next few slides. All other results are included in the appendix.

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.24	3389 | 0.15	803 | 0.25
Farm Yard Manure Indicator	**-1.35***** **(0.26)**	**0.39**** **(0.15)**	**-1.16***** **(0.34)**
Fertilization Frequency	0.24 (0.19)	-0.01 (0.13)	-0.34 (0.42)
Nitrogen (kg/acre)	0.001 (0.001)	**0.07***** **(0.006)**	0.01 (0.01)
Phosphorous	**0.01**** **(0.001)**	**0.03***** **(0.009)**	0.003 (0.02)
Potassium	**0.03***** **(0.03)**	-0.003 0.011	0.03 (0.04)
Zinc	**0.17***** **(0.06)**	**0.014***** **0.001**	0.07 (0.07)
Irrigated Area	**3.50***** **(0.32)**	**3.39***** **(0.70)**	0.93 (1.86)
Insect Control **(No Presence + Control: 2** **Presence + Control: 1** **No Presence + No Control: 0** **Presence + No Control: -1)**	-0.16 (0.17)	**0.58***** **0.10**	0.25 (0.26)

*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

We find that irrigation and nutrient application (N, P, K and Zn) is positively associated with the average yield except in the TGP region.

Nitrogen Use Efficiency in the Trans-Gangetic Plain Region:

- In the TGP region, the effect of nitrogen is insignificant. This can be explained as follows:
- In general, the effect of nutrient application on yield increases until a certain threshold and then becomes constant. After this threshold, additional fertilizer application does not translate into additional nutrient absorption by the crop.

- This threshold, indicated by the optimal amount of nitrogen application per acre (obtained here) is ~55 kg/acre for the TGP region. (Note: I have taken a mean of the values suggested for Punjab and Haryana).

- However, only ~25 percent of our sample plots have a nitrogen application rate <= 55 kilograms/acre. Hence, the effect on average yields is insignificant.

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.24	3389 | 0.15	803 | 0.25
Total Rainfall (in mm.)	**0.001** **(0.0007)**	**0.003***** **(0.0003)**	**-0.02***** **(0.002)**
Average Temperature (in degree celsius)	**-0.42** **(0.21)**	**1.38***** **(0.25)**	**-2.36***** **(0.30)**
Soil Texture **(0,1,2: Light, Medium, Heavy)**	-0.21 (0.21)	**0.86***** **(0.22)**	-0.42 (0.35)
Perceived Soil Quality **(0,1,2: Low, Medium, High)**	**1.70***** **(0.27)**	**0.89***** **(0.28)**	-0.09 (0.91)
Drainage Class **(0,1,2: Low, Medium, Up)**	**0.47***** **(0.17)**	**-1.48***** **(0.15)**	-0.12 (0.44)
Caste **(0,1,2: SC/ST, OBC, General)**	0.12 (0.13)	**0.39***** **(0.13)**	**1.54***** **(0.46)**
Ownership **(Owned = 1, 0 otherwise)**	**0.99***** **(0.28)**	-0.10 (0.17)	**1.80***** **(0.56)**
Constant	19.91*** (6.53)	-30.48*** (6.51)	98.63*** (9.08)

*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

We find that rainfall is positively associated with the average yield except in the TGP region. This can be explained by the yield damage due flash floods (documented here and here) in Punjab and Haryana during 2018.

Soil Quality and Demographics:

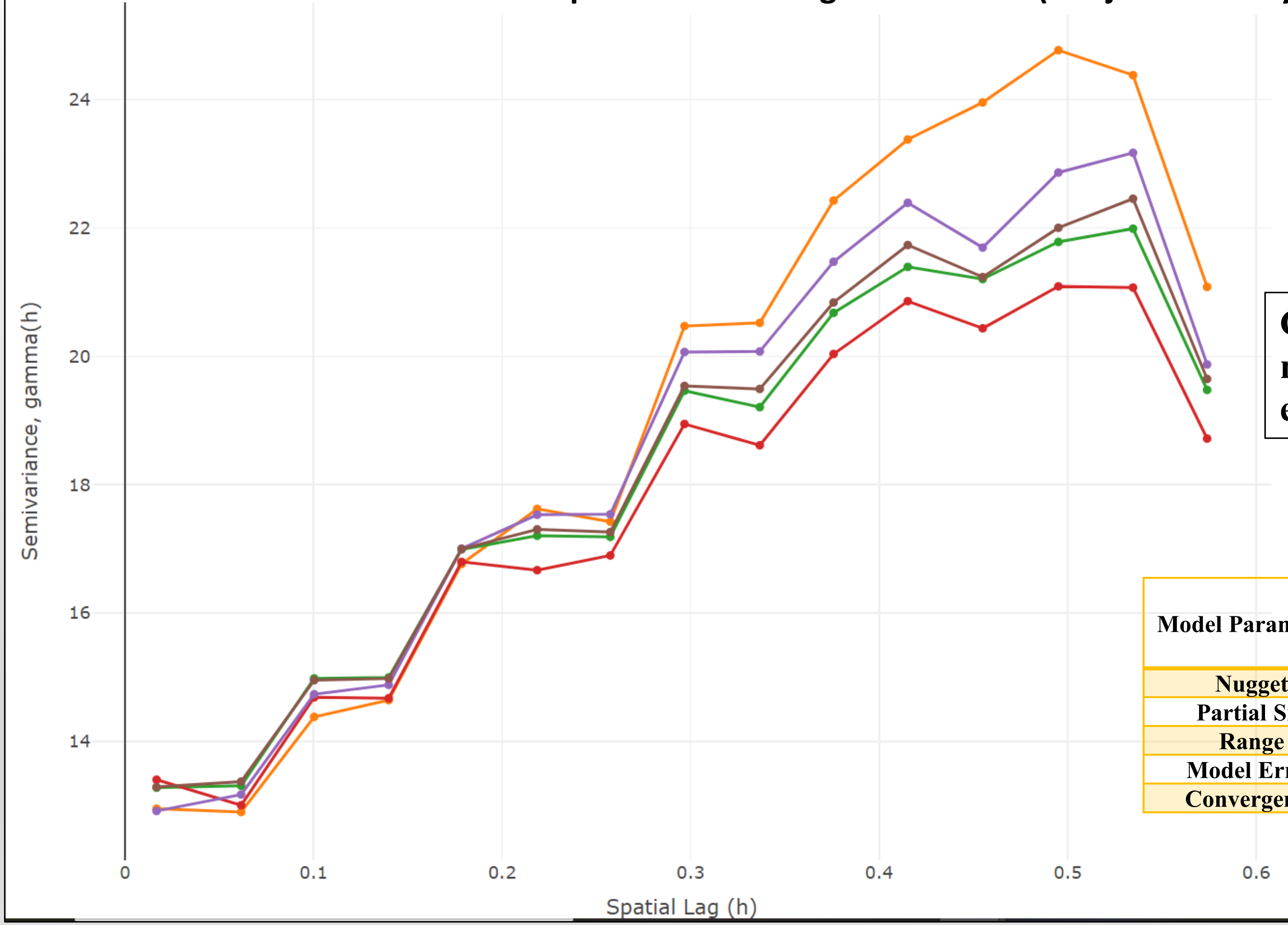
- Higher perceived soil quality is found to be associated with higher yields on-average for both LGP and MGP regions.

- For drainage class, the results are expected because, on the one hand upland soils are generally "well-drained" and do not have as much surface water accumulation – which is important for the rice crop. At the same time, up-land rice varieties are (largely) more drought tolerant as compared to low-land varieties which can also lead to higher yields.

- We also observe that farmers who own their plots have higher yields compared to farmers who do not own their plots.

- Finally, general caste households have higher yields as compared to their SC/ST and OBC counterparts. However, this difference is not significant in the LGP region.

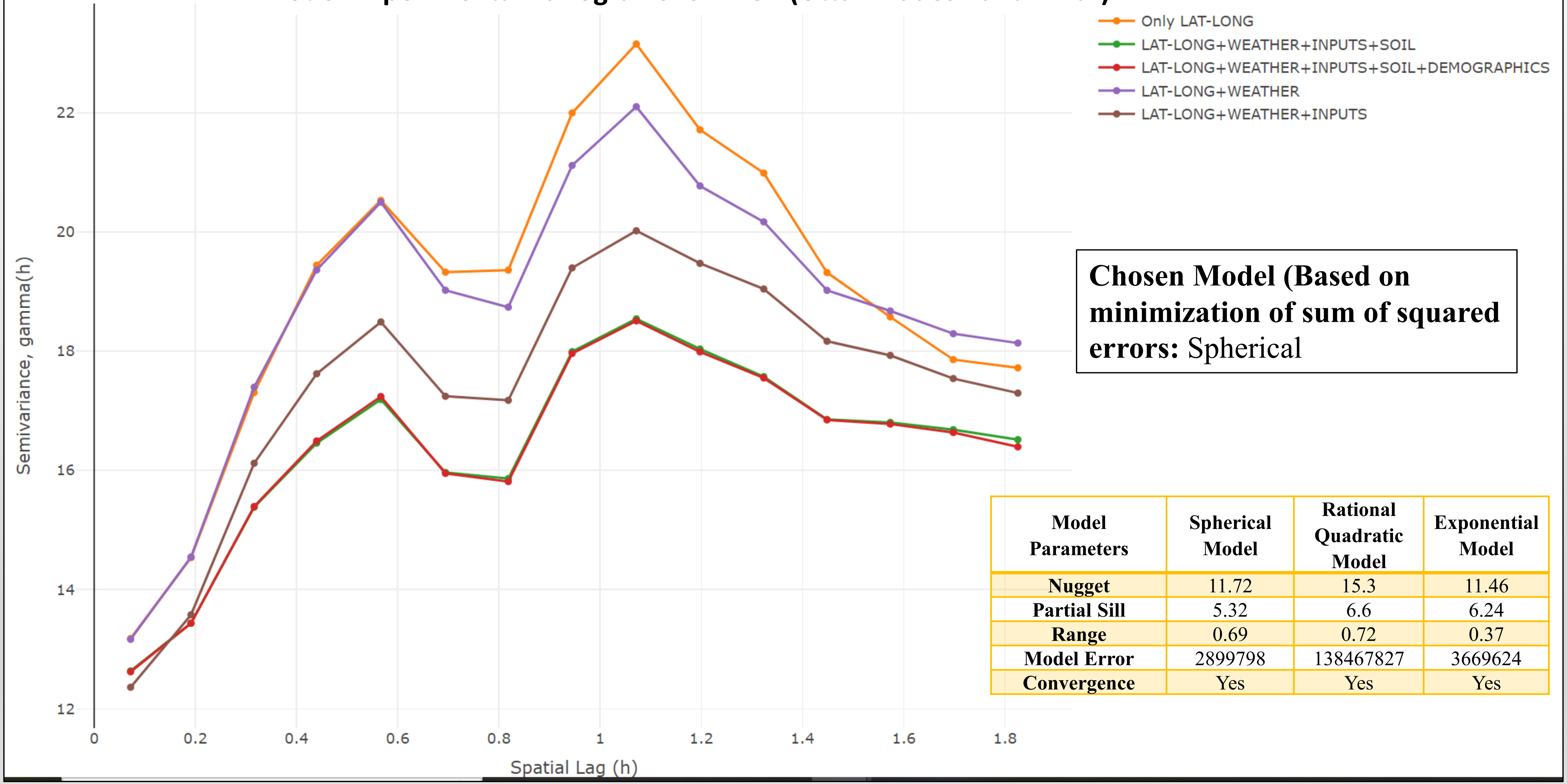
Plot of Experimental Variograms for TGP (Punjab and Haryana)



Chosen Model (Based on minimization of sum of squared errors: Spherical

Model Parameters	Spherical Model	Rational Quadratic Model	Exponential Model
Nugget	13.04	12.02	13.07
Partial Sill	11.13	8.7	10235.6
Range	0.99	0.55	630.5
Model Error	3012538	739655648	3080012
Convergence	Yes	Yes	No

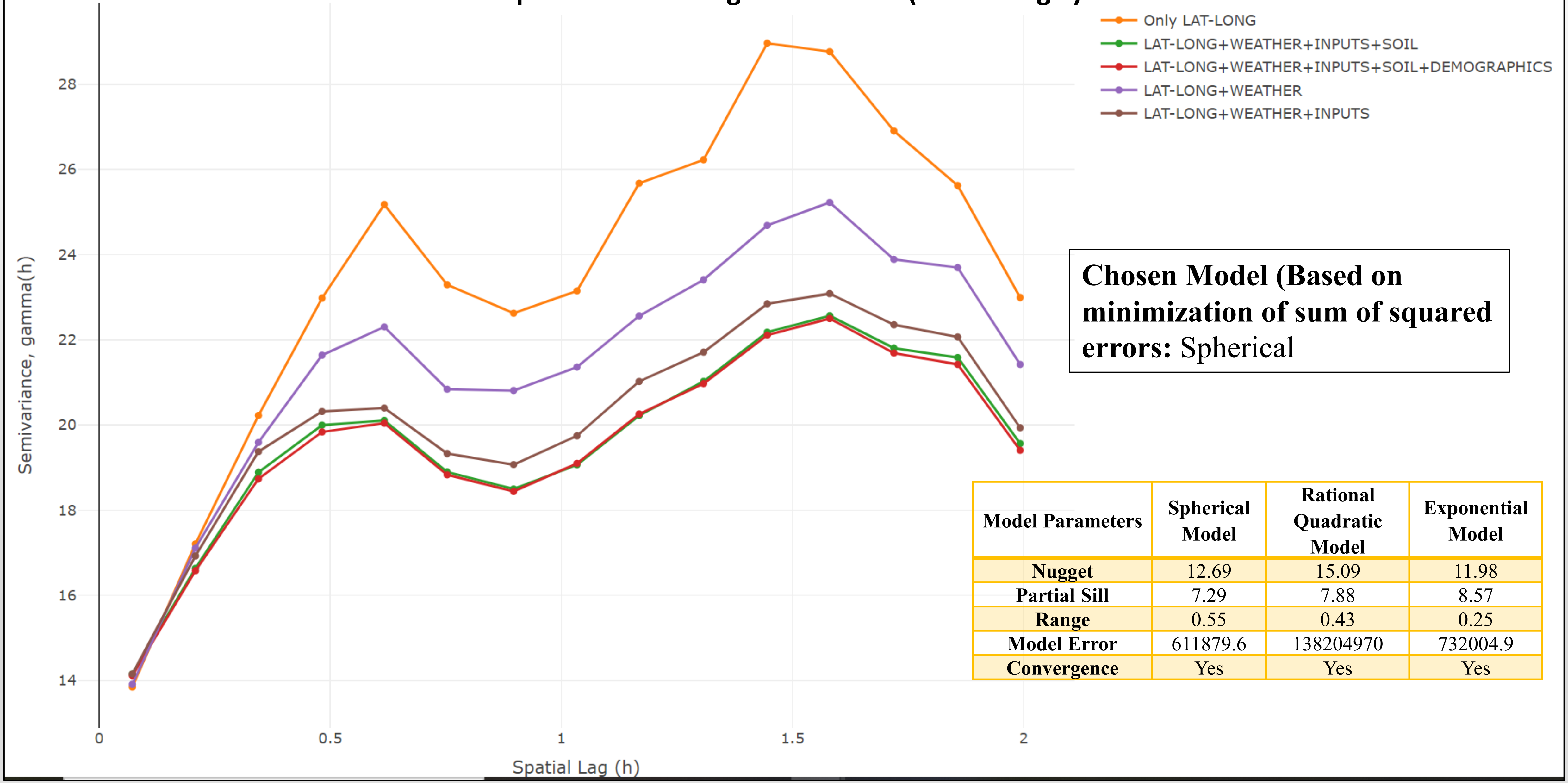
Plot of Experimental Variograms for MGP (Uttar Pradesh and Bihar)



Chosen Model (Based on minimization of sum of squared errors: Spherical

Model Parameters	Spherical Model	Rational Quadratic Model	Exponential Model
Nugget	11.72	15.3	11.46
Partial Sill	5.32	6.6	6.24
Range	0.69	0.72	0.37
Model Error	2899798	138467827	3669624
Convergence	Yes	Yes	Yes

Plot of Experimental Variograms for LGP (West Bengal)



Chosen Model (Based on minimization of sum of squared errors: Spherical

Model Parameters	Spherical Model	Rational Quadratic Model	Exponential Model
Nugget	12.69	15.09	11.98
Partial Sill	7.29	7.88	8.57
Range	0.55	0.43	0.25
Model Error	611879.6	138204970	732004.9
Convergence	Yes	Yes	Yes

Spatial Dependence in Yields:

- From the variogram plots and model results, it is clear that the spatial dependence structure for yields is different for the TGP, MGP, and LGP regions.

- The nugget effect values are close, reflecting that the micro-scale variation is similar (for the de-trended residual yields, i.e., after accounting for input application, soils, weather and demographics).

- However, the range values are ~ 1, 0.7, and 0.5 degrees, i.e., 111.1, 77.8, and 55.5 kilometres for the TGP, MGP, and LGP regions respectively. This reflects that spatial dependence in yields falls at a relatively lower rate in TGP as compared to MGP and LGP regions.

- Finally, the partial sill reflecting the absolute spatial dependence in yields are ~ 11, 5, and 7 for the TGP, MGP and LGP regions. Hence, yields in the MGP and LGP regions exhibit similar levels of spatial dependence which is higher than the TGP regions.

Thank you!

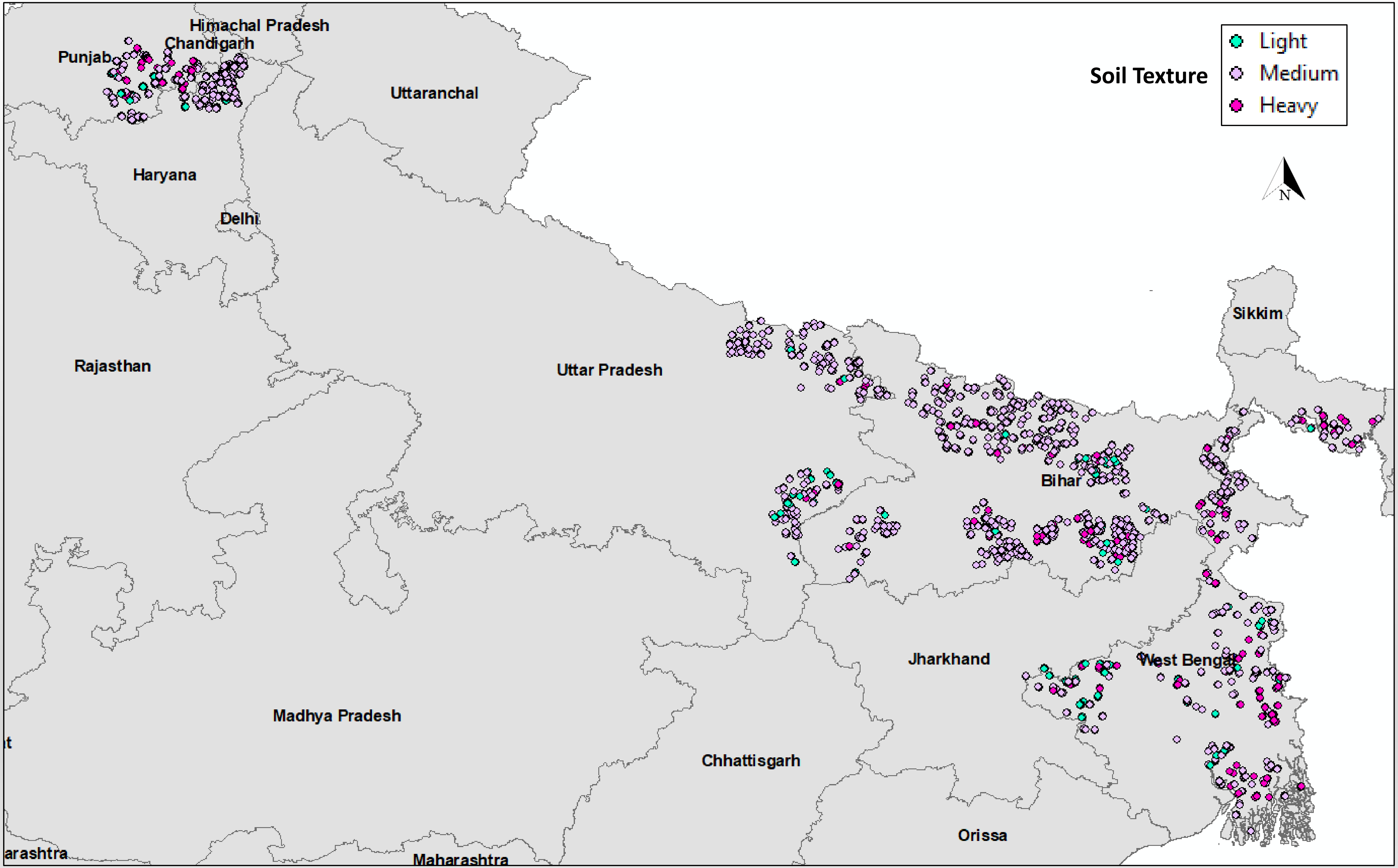
References:

Ajay A. et. al. 2022. *Large survey dataset of rice production practices applied by farmers on their largest farm plot during 2018 in India,* Data in Brief 45: 2352-3409

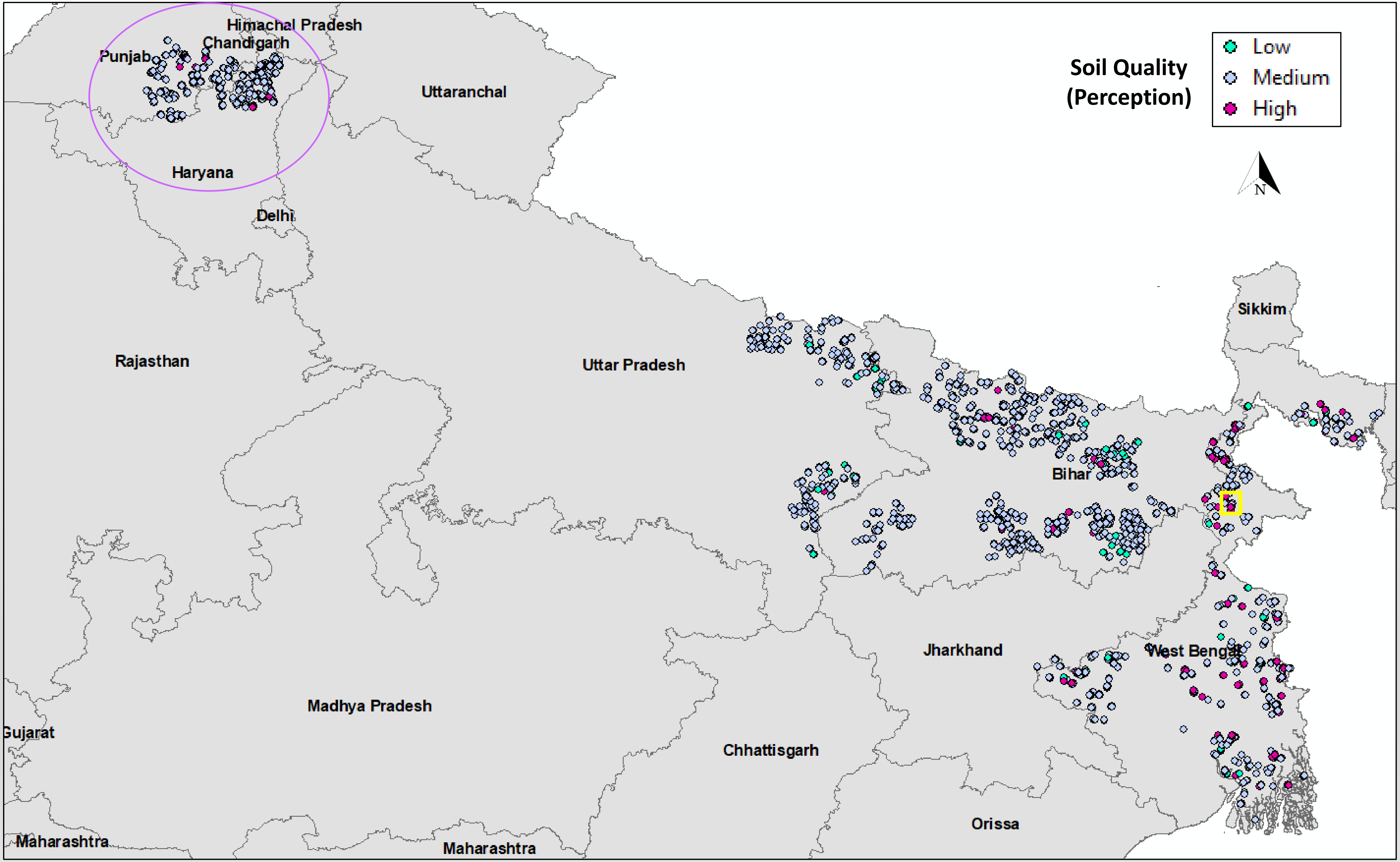
Ramadhani F. ; Pullanagiri R. ; Kereszturi G. ; Procter J. 2021. *Mapping a Cloud-Free Rice Growth Stages Using the Integration of PROBA-V and Sentinel-1 and Its Temporal Correlation with Sub-District Statistics,* Remote Sensing, 13:1498.

Singh J. ; Erenstein O. ; Thorpe W ; and Varma A. *Crop–livestock interactions and livelihoods in the Gangetic Plains of Uttar Pradesh, India*, International Livestock Research Institute

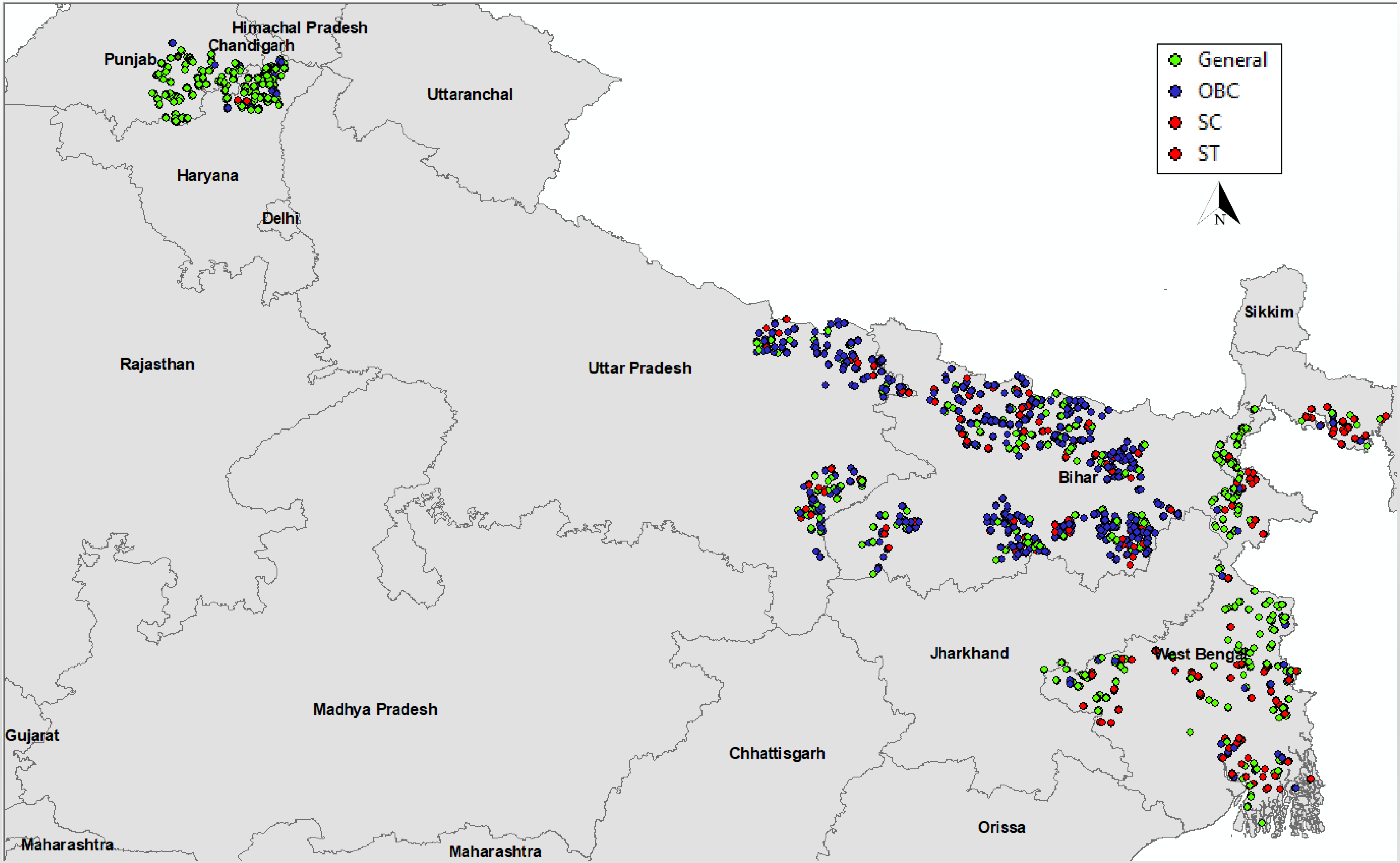
Appendix

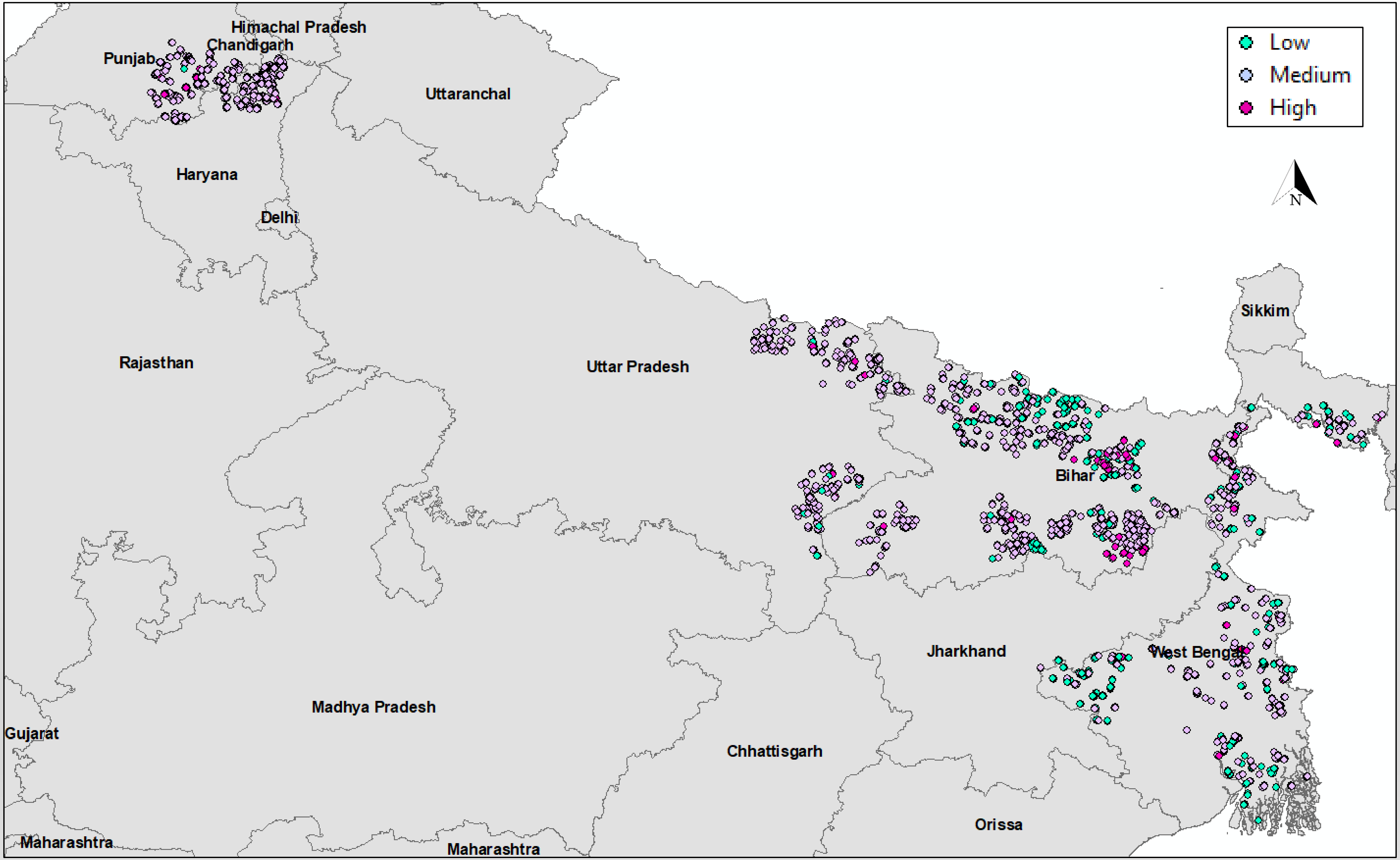


Himachal Pradesh
Chandigarh
Punjab
Uttaranchal
Haryana
Delhi
Rajasthan
Uttar Pradesh
Sikkim
Bihar
Jharkhand
West Bengal
Madhya Pradesh
Chhattisgarh
Orissa
Maharashtra
arashtra
Soil Texture
Light
Medium
Heavy
N

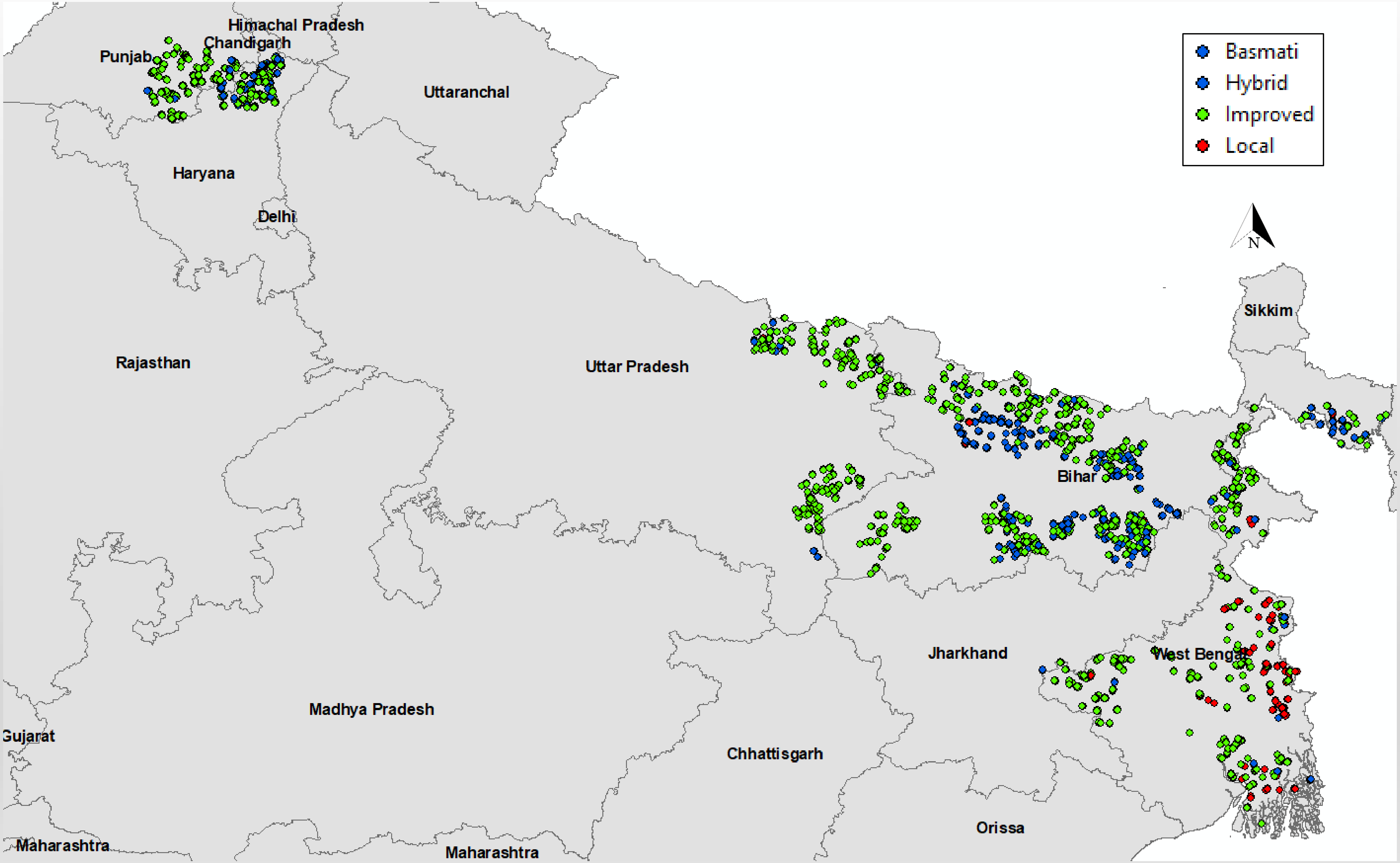


Himachal Pradesh
Chandigarh
Punjab
Uttaranchal
Haryana
Delhi
Rajasthan
Uttar Pradesh
Sikkim
Bihar
Jharkhand
West Bengal
Madhya Pradesh
Gujarat
Chhattisgarh
Orissa
Maharashtra
Maharashtra
Soil Quality (Perception)
Low
Medium
High
N


General
OBC
SC
ST
N
Himachal Pradesh
Chandigarh
Punjab
Uttaranchal
Haryana
Delhi
Rajasthan
Uttar Pradesh
Sikkim
Bihar
Jharkhand
West Bengal
Madhya Pradesh
Gujarat
Chhattisgarh
Orissa
Maharashtra
Maharashtra

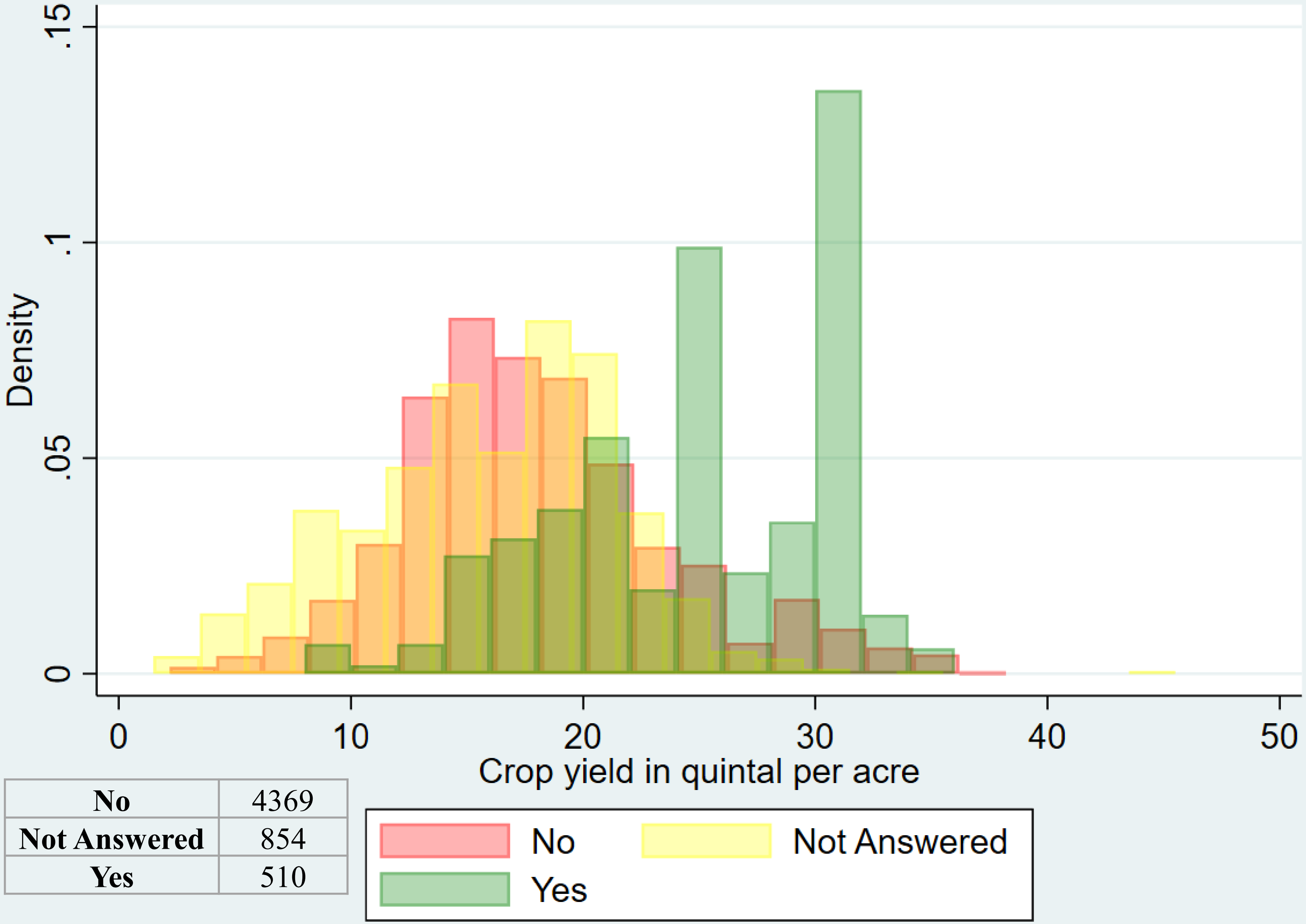


Low and Very Low Land soils are indeed clustered in the lower-Gangetic plains.



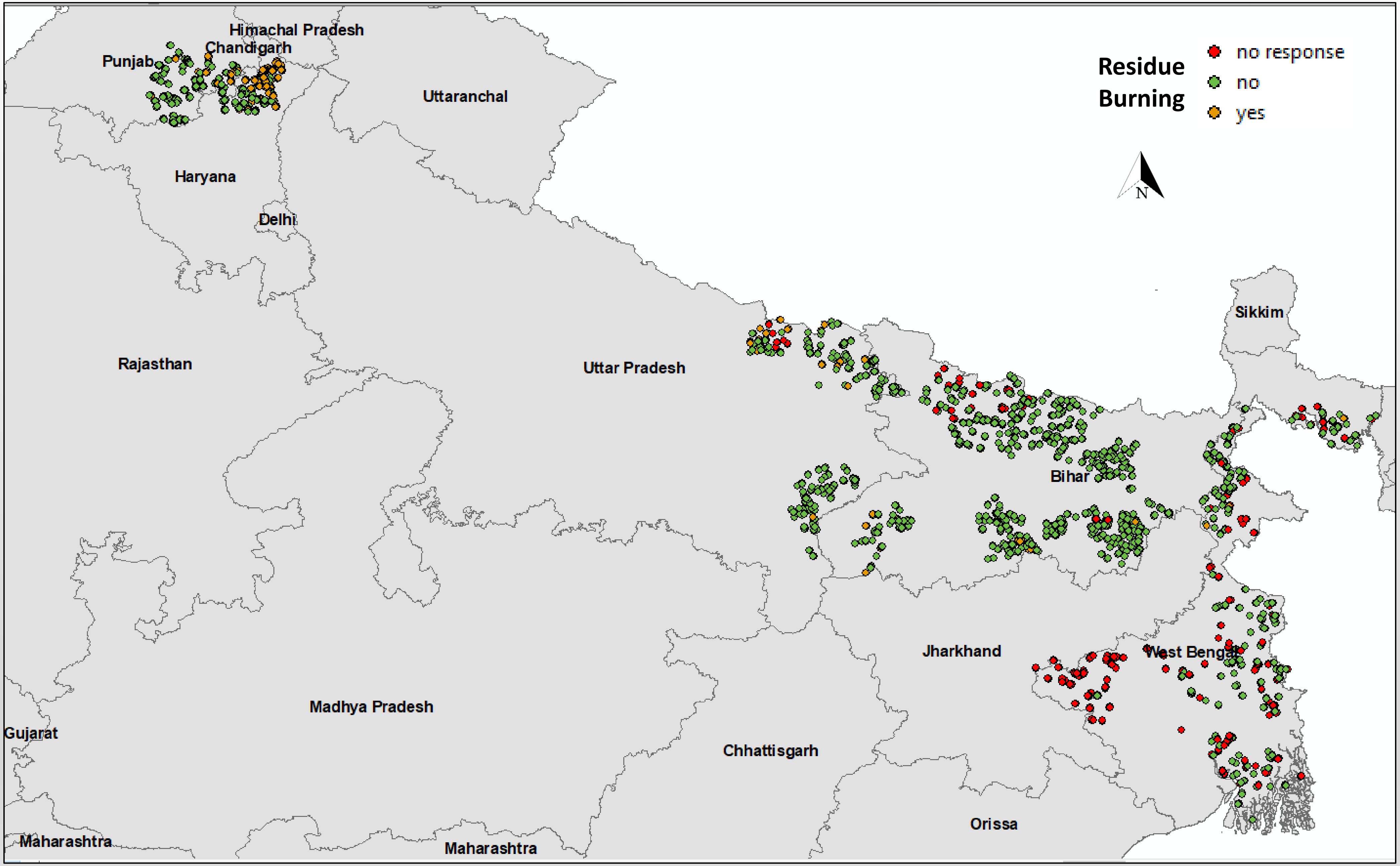
Interestingly, even though West Bengal has generally as high yields as the rest of the states, almost all the local seed use is also restricted within the state.

Crop Yield by Residue Burning (Previous Crop)



No	4369
Not Answered	854
Yes	510

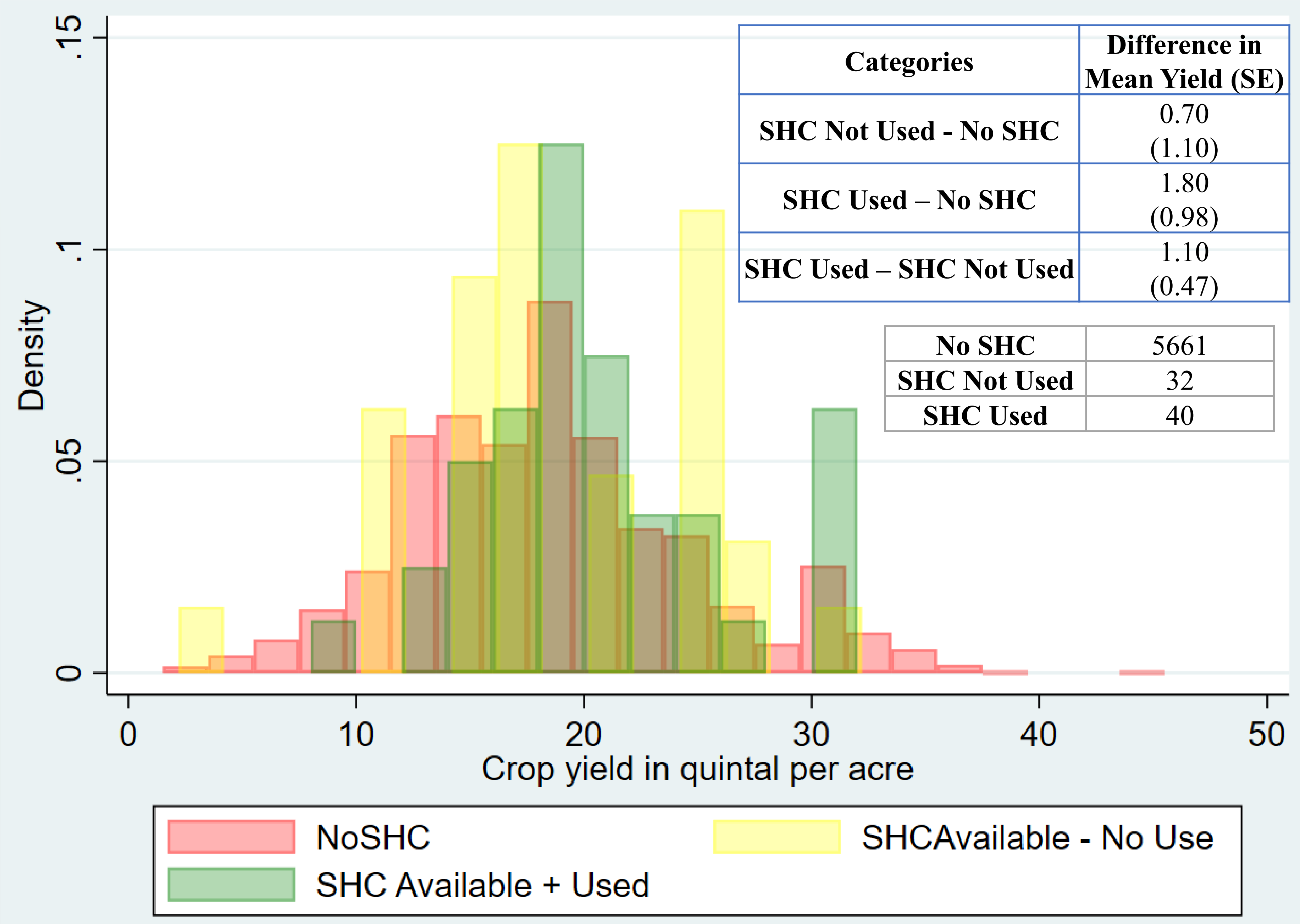
Categories	Difference in Mean Yield (SE)
Not Answered – No	-2.45*** (0.22)
Yes – No	6.12*** (0.27)
Yes – Not Answered	8.58*** (0.33)



Reds and Oranges are seemingly clustered together.

But, as it turns out the previous crop in case of "No Response" is fallow, i.e., there was no crop grown in the previous season.

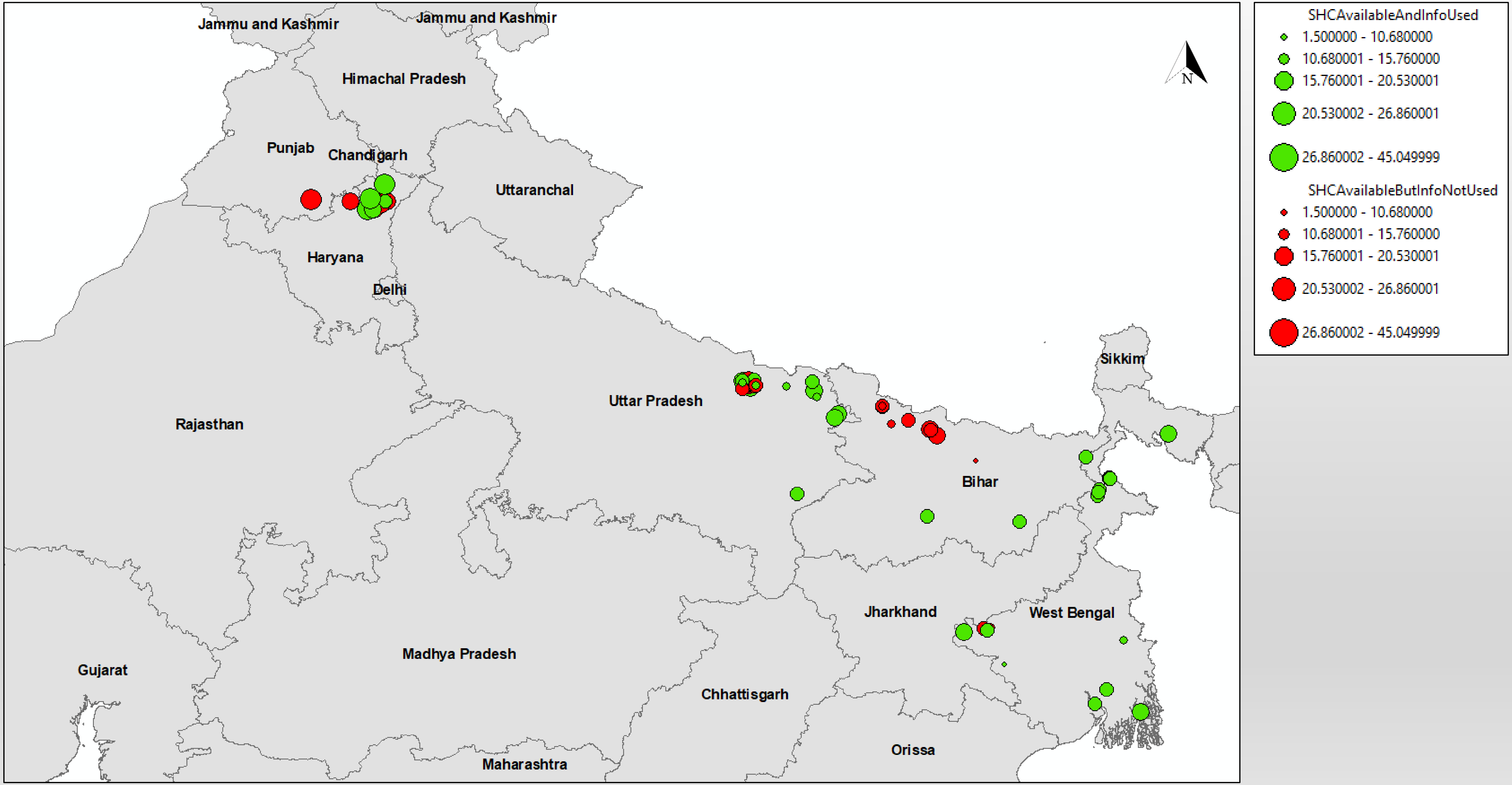
Crop Yield by Soil Health Card Information

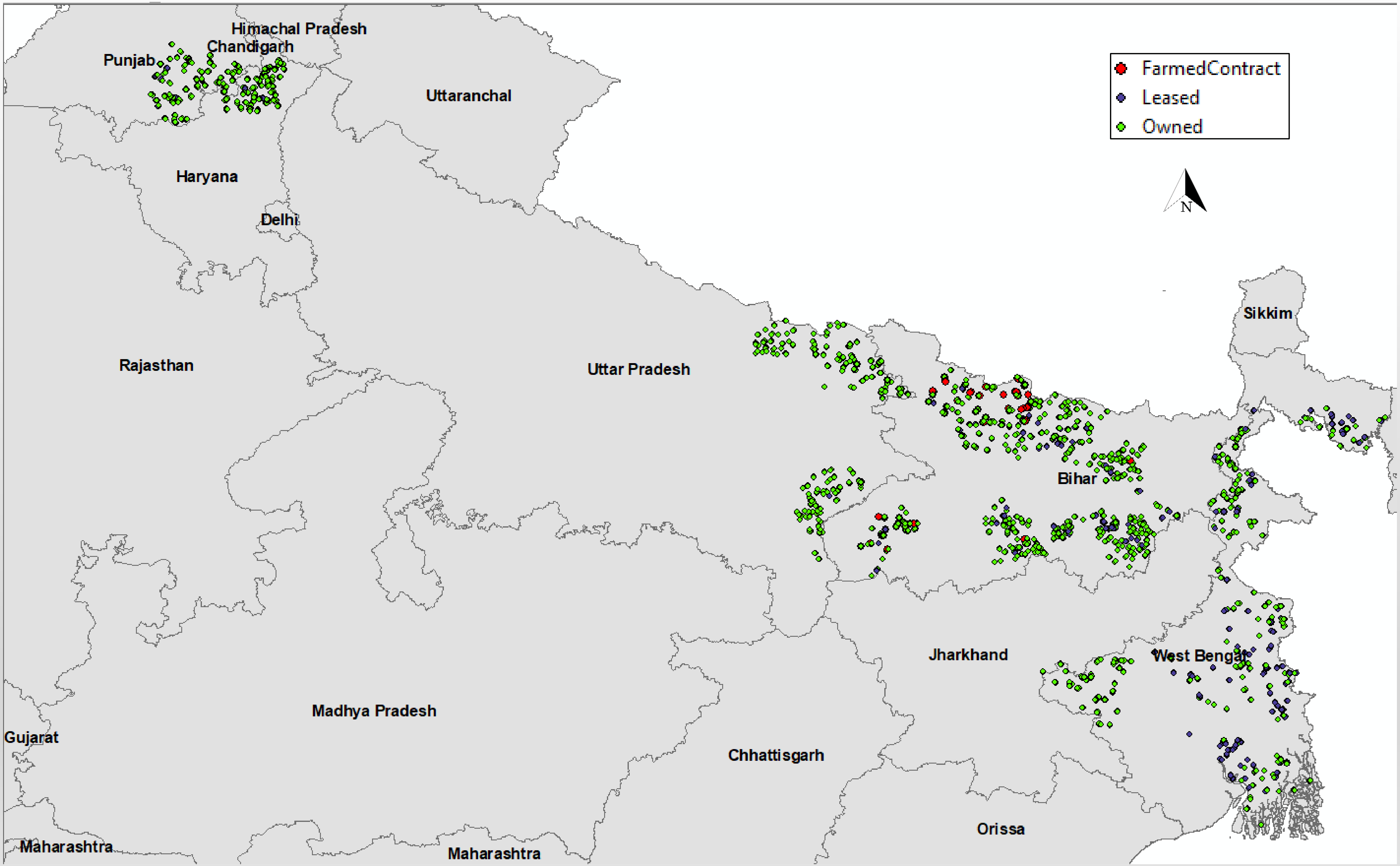


Categories	Difference in Mean Yield (SE)
SHC Not Used - No SHC	0.70 (1.10)
SHC Used – No SHC	1.80 (0.98)
SHC Used – SHC Not Used	1.10 (0.47)

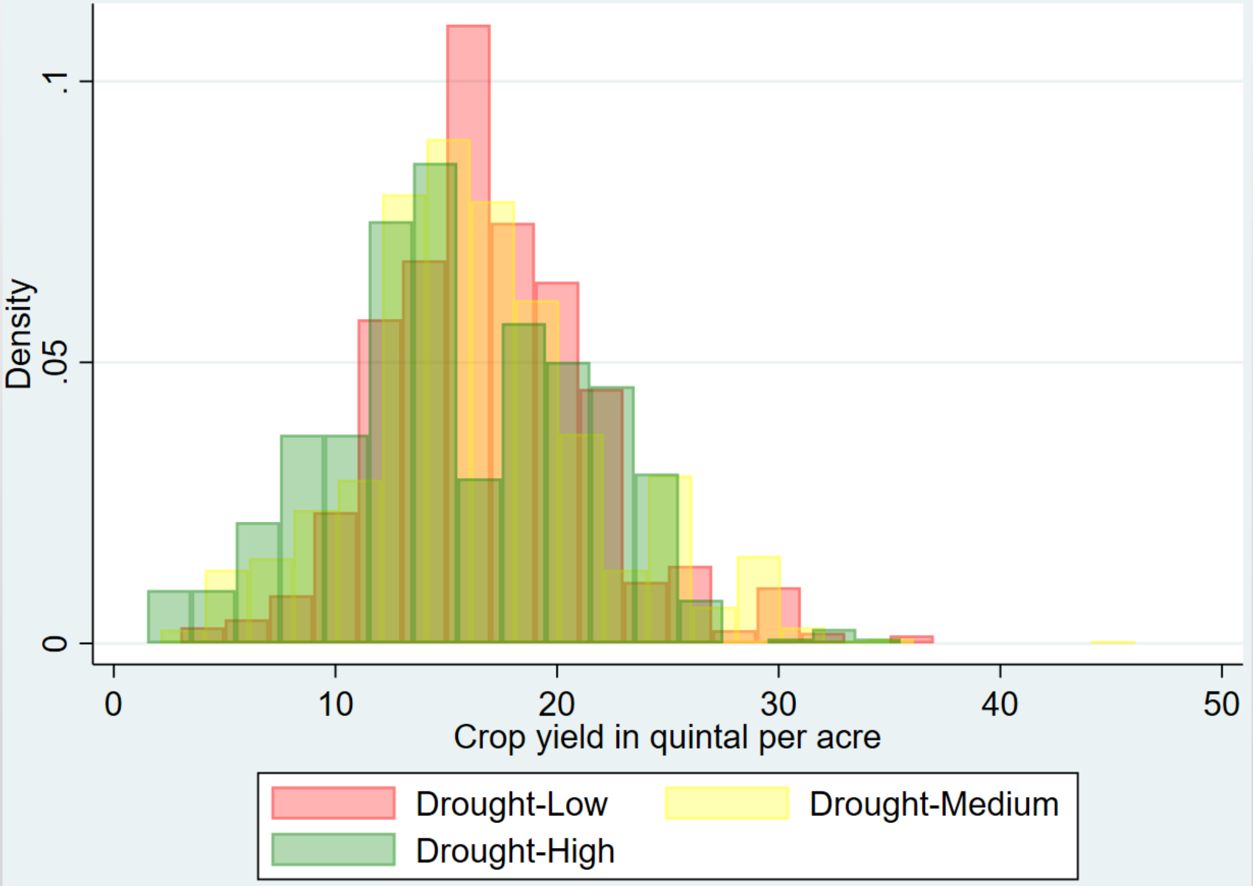
No SHC	5661
SHC Not Used	32
SHC Used	40

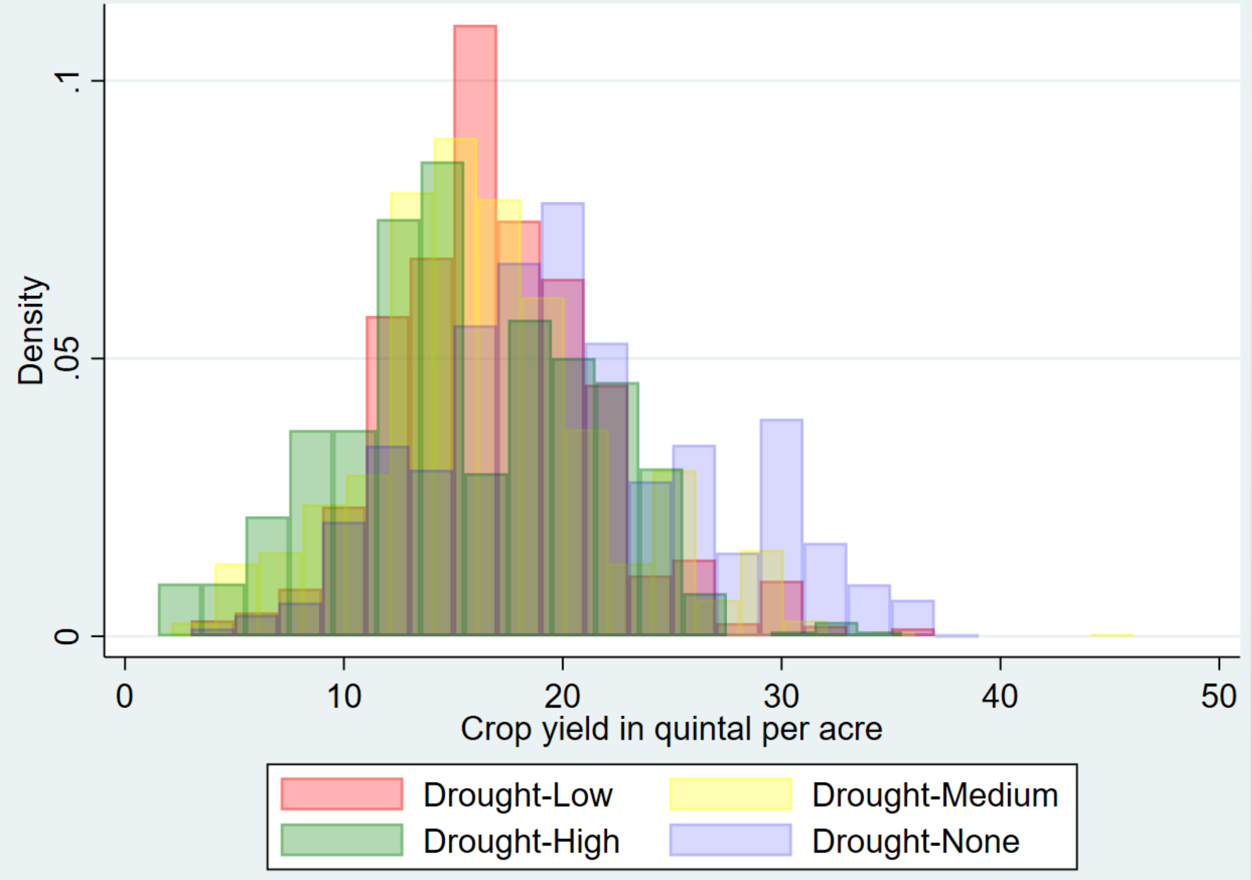
Crop Yield by Soil Health Card Information




FarmedContract
Leased
Owned
N
Himachal Pradesh
Chandigarh
Punjab
Uttaranchal
Haryana
Delhi
Rajasthan
Uttar Pradesh
Sikkim
Bihar
Jharkhand
West Bengal
Madhya Pradesh
Gujarat
Chhattisgarh
Orissa
Maharashtra
Maharashtra

Crop Yield by Drought Severity

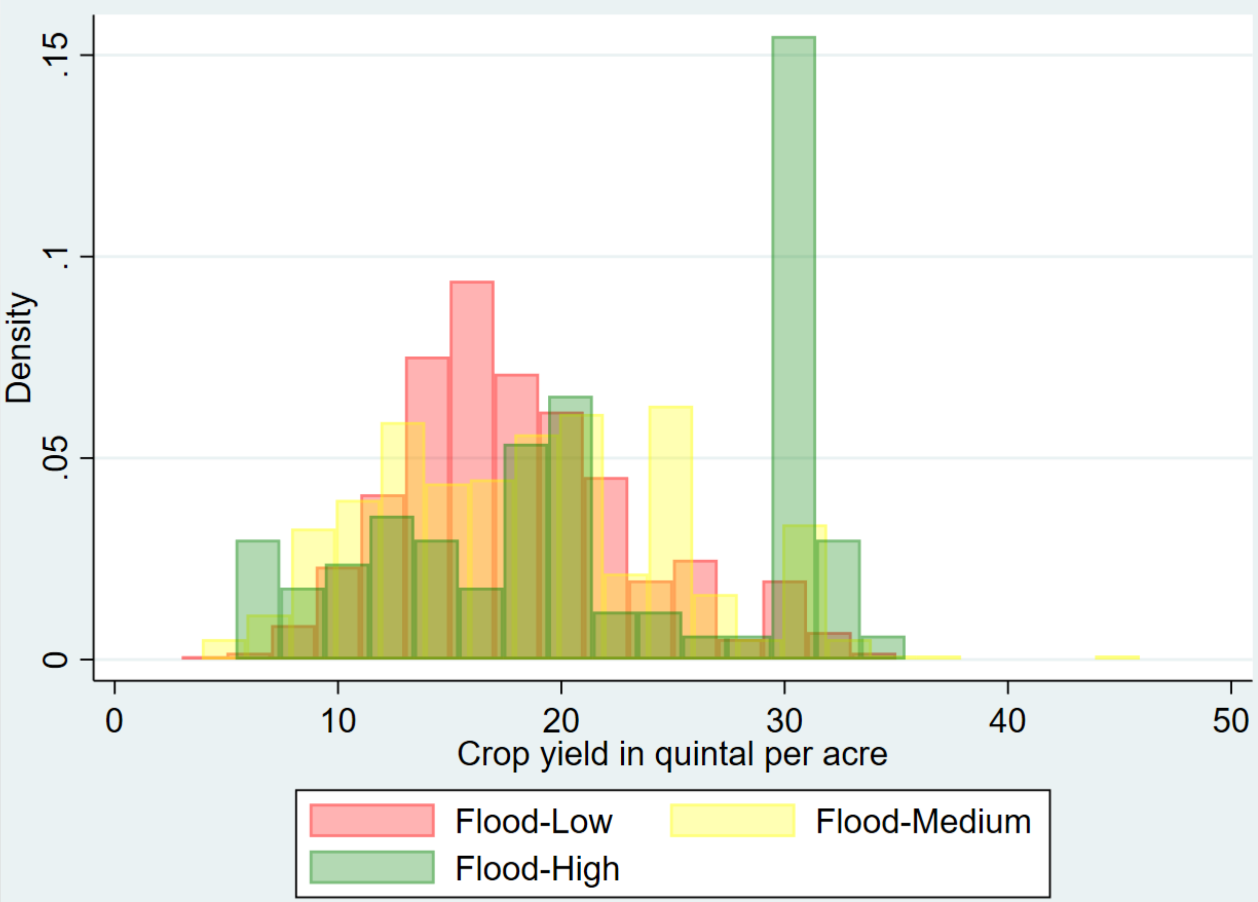


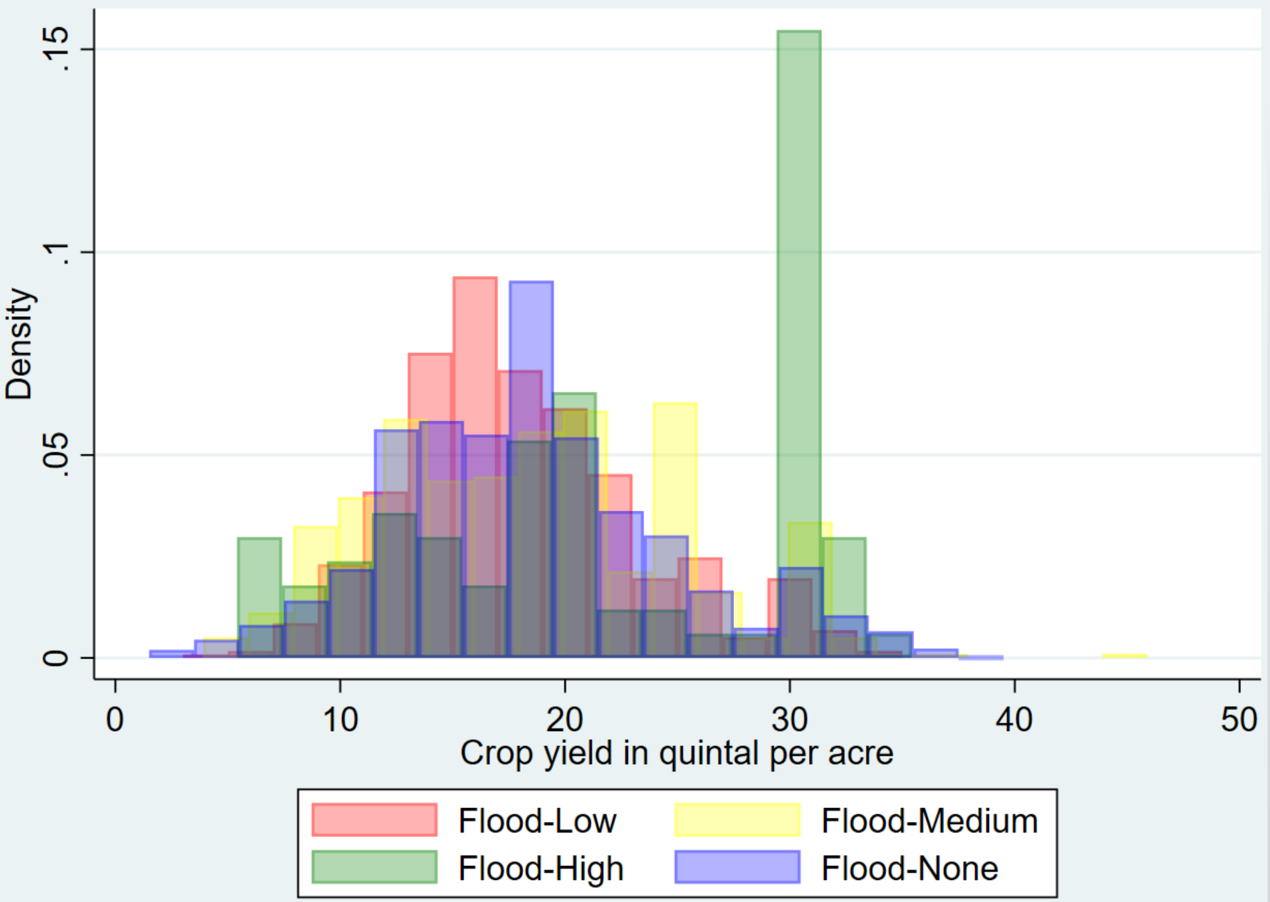


Drought Stress	Frequency	Percentage
High	579	10.1
Low	1,049	18.3
Medium	1,220	21.3
None	2,885	50.3

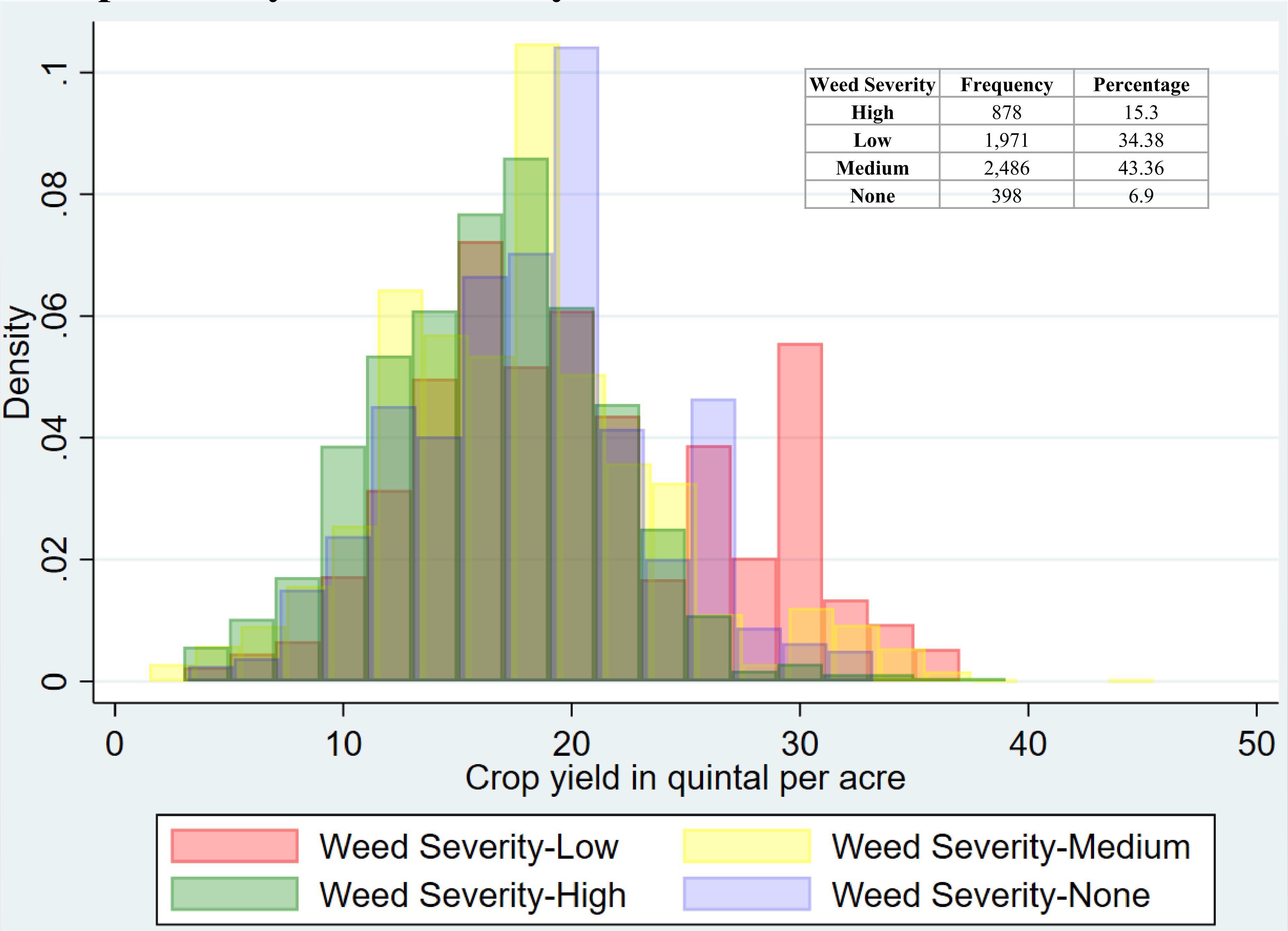
Crop Yield by Flood Severity

High	84	1.46
Low	585	10.2
Medium	492	8.6
None	4,571	79.7





Crop Yield by Weed Severity



Weed Severity	Frequency	Percentage
High	878	15.3
Low	1,971	34.38
Medium	2,486	43.36
None	398	6.9

Crop Establishment: Determinants of Timing of Nursery Establishment

Determinants of Timing of Nursery Establishment	Count	Percentage
Calendar Date	3643	33.98
Irrigation Water Availability	2438	22.74
Pre-Monsoon Showers	2116	19.74
Neighbouring Farmers Practice	1091	10.18
Seed Availability	927	8.65
Weather Forecast Information	495	4.62
Other	11	0.10
Total	10721	

Determinants of Nursery Establishment Timings	LGP (27%)	MGP (59%)	TGP (15%)
Water Availability	23.4	71.1	5.5
Seed Availability	**5.6**	**91.9**	**2.5**
Weather Forecast	**18.2**	**80.6**	**1.2**
Neighbourhood	**11.3**	**84.3**	**4.4**
Calendar Date	23.1	55.9	21.0

Determinants of Timing of Nursery Establishment by Agro-ecological Zones: All numbers are the row percentages.

Clearly, considerations of **seed availability, weather forecast and effect of neighborhood farm activities** on nursery establishment timings are concentrated in the Middle Gangetic Belt.

Crop Establishment: Determinants of Timing of Sowing/Transplanting

Determinants of Timing of Sowing/Transplanting	Frequency	Percentage
Seedling Age	3669	29.6
Irrigation Water Availability	2969	23.9
Calendar Date	2431	19.6
Rain Arrival	2305	18.6
Labour Availability	1032	8.3
Total	12406	

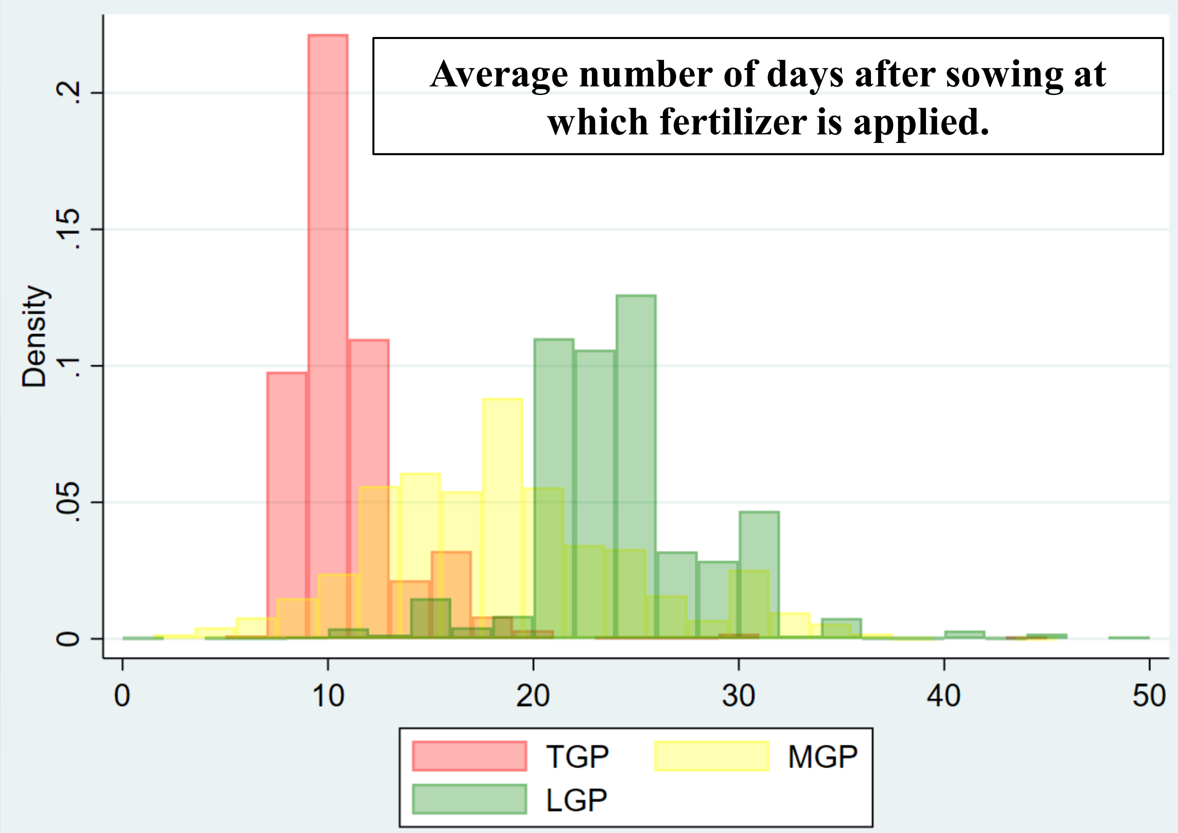
Determinants of Timing of Sowing/Transplanting	LGP (27%)	MGP (59%)	TGP (15%)
Water Availability	22.1	65.5	12.4
Seedling Age	22.3	67.3	10.3
Labour Availability	**3.8**	**60.5**	**35.8**
Calendar Date	17.8	65.4	16.8

Determinants of Timing of Sowing/Transplanting by Agro-ecological Zones: All numbers are the row percentages.

Considerations of **labour availability** are seemingly concentrated in the Trans-Gangetic zone.

Fertilizer Usage (Kg/Acre)	Mean	SD	Minimum	P25	P50	P75	Maximum
DAP	31.22	26.45	0	0	33.9	50	300.3
NPK	7.61	23.70	0	0	0	0	400
UREA	87.64	72.45	0	62.26	88.9	112.5	4654.7
NPKS	0.32	4.09	0	0	0	0	96.39
MoP	9.57	17.68	0	0	0	15.87	220
SSP	6.43	24.21	0	0	0	0	300.3
TSP	0.18	4.89	0	0	0	0	298.5
ZnSO4	1.96	18.88	0	0	0	0	1315.8
Gypsum	0.05	1.90	0	0	0	0	100
Boron	0.01	0.27	0	0	0	0	15.02

The amount of usage is highly skewed.



As of now, we don’t account for this when modelling yield.

Summary Nutrient Application (per acre)

Zone	LGP (N = 1536)								MGP (N = 3392)								TGP (N = 805)							
Fertilization	Mean	SD	Min	P25	P50	P75	P90	Max	Mean	SD	Min	P25	P50	P75	P90	Max	Mean	SD	Min	P25	P50	P75	P90	Max
Frequency	2.30	0.85	0	2	2	3	3	4	2.84	0.63	0	3	3	3	4	4	3.34	0.61	0	3	3	4	4	4
Nitrogen (kg/acre)	32.46	60.04	0	16.8	26.5	41.1	56.5	2195.2	49.78	13.21	0	40.9	49.8	58.5	67.3	92	61.82	12.61	0	55.2	62.1	70.2	80.5	142.2
Phosphorus (kg/acre)	19.24	17.87	0	8.8	13.8	24.5	39.6	214.8	19.02	9.31	0	14.6	20.4	25.6	29.7	38.98	7.03	10.23	0	0	0	20.7	23	46
Potassium (kg/acre)	18.60	20.17	0	0	15	27.95	44.4	236	4.25	6.18	0	0	0	8.89	13.33	29.03	0.31	2.13	0	0	0	0	0	30
Zinc (kg/acre)	0.20	1.18	0	0	0	0	0	22.2	0.79	8.99	0	0	0	0	2.1	486.8	1.47	1.83	0	0	1.85	1.85	3.7	18.5

Examination of the Trend: Input Decisions

Tillage (Land Preparation)
- Type for current and previous crop.
  - Directly Sown
  - Transplanted

Crop Establishment
- Date (Sowing + Nursery Establishment)
- Determinants of Timing of Transplanting and Nursery Establishment.

Fertilization
- Basal + First + Second + Third Topping

For each Topping:
- Amount
- Days after Sowing
- Type: FYM (Wet vs. Dry), DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron

Irrigation
- Frequency
- Availability
- Source
  - Depth of Tubewell
  - Energy Source of Pump
  - Flat Pipes
- Stage of Crop Growth at which Irrigation Application Occurs

Crop Protection
- Weed and Insect Severity
- Herbicide Application Timing and Frequency
- Manual Weeding (First + Second + Third) Timing

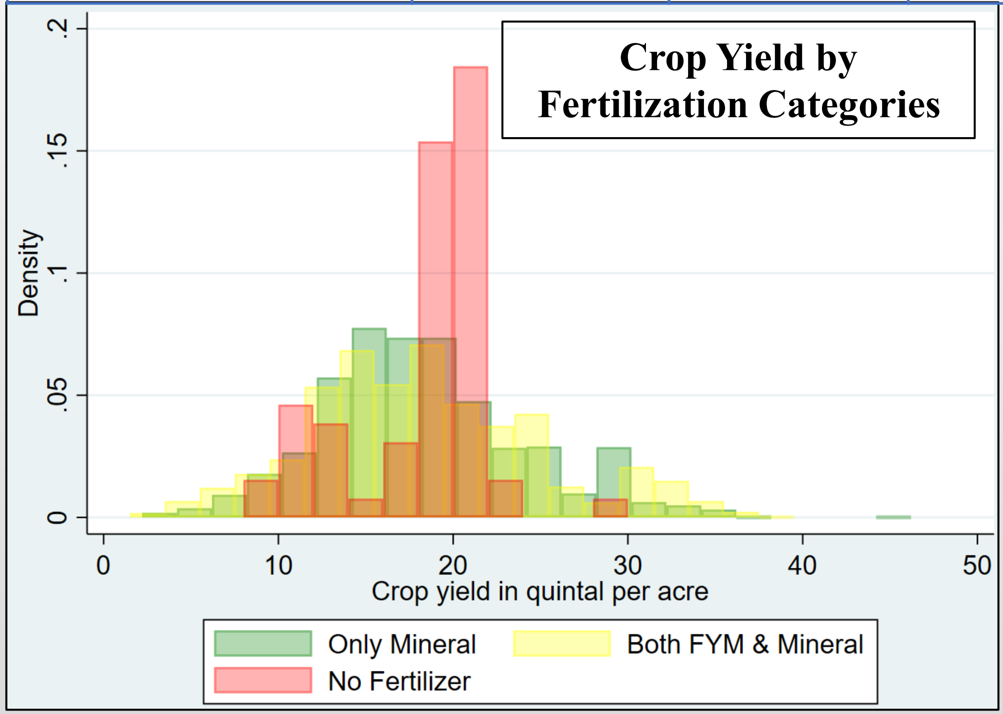
Harvesting
- Harvest Date
- Harvest Method
- Threshing Type

Seeds and Fertilization Type:

- FYM (Farm Yard Manure) : Wet vs. Dry
- Mineral: DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron

Category	LGP (27%)	MGP (59%)	TGP (15%)
Only Mineral	21.3	65.2	13.5
Both FYM & Mineral	35.1	49.5	15.4
No Fertilizer	89.2	9.2	1.5
Only FYM	50	25	25

Seed Source	Frequency	Percentage
Cooperative	58	1.01
Government KVKSAU	137	2.39
Neighbour/Relative	183	3.19
Other	1	0.02
Private Seed Dealer	4,839	84.41
Self Saved	515	8.98



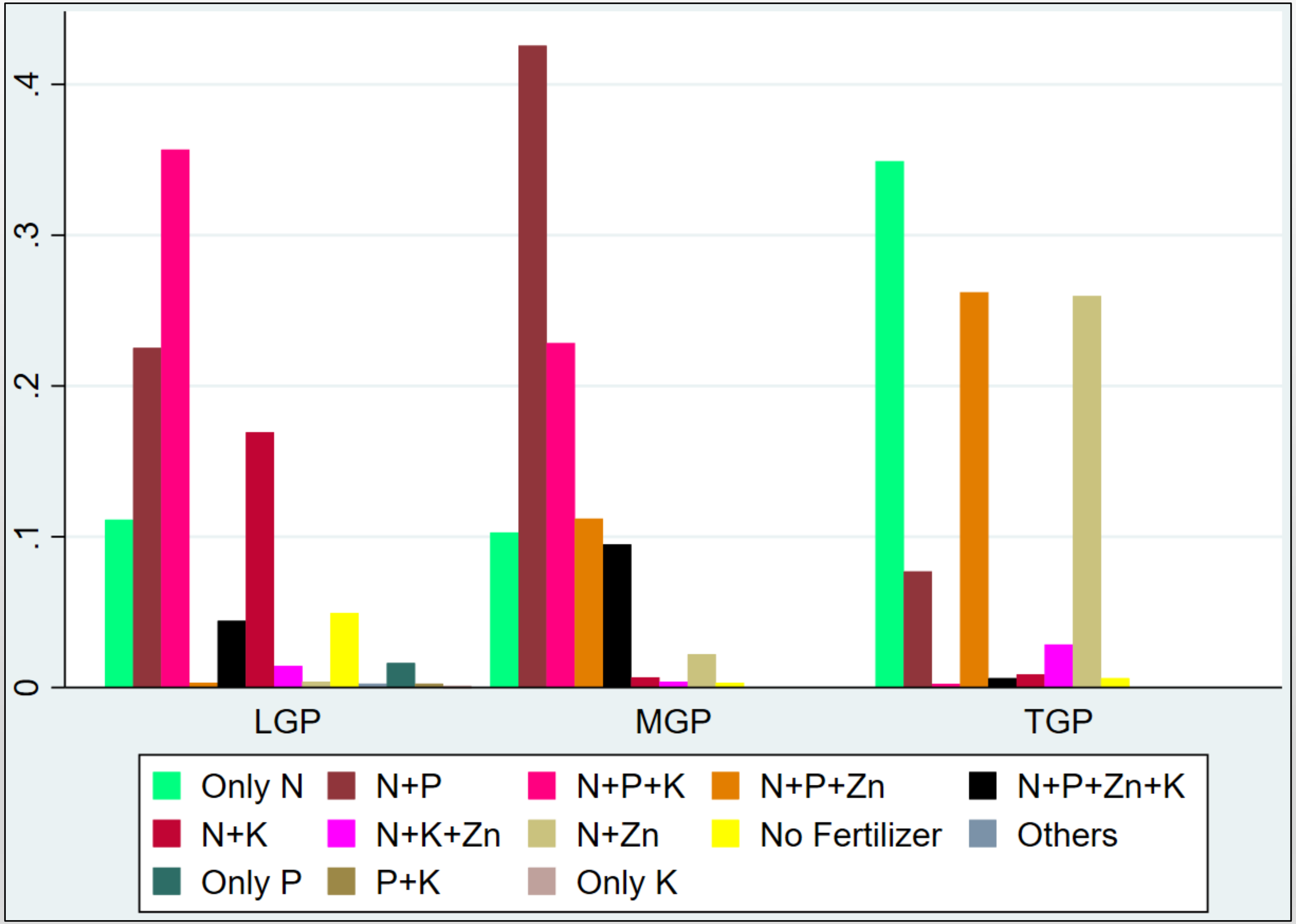
Fertilizer Type	Frequency	Percentage
Only Mineral	3718	64.9
Both FYM & Mineral	1934	33.7
Only FYM	16	0.28
No Fertilizer	65	1.14

Yields on plots with Dry FYM usage are on-average lower compared to plots with Wet FYM usage.

Type of FYM	Frequency	Percentage
Dry	3,783	65.99
Wet	1,950	34.01

While FYM usage is observed on only ~34% of plots, mineral fertilizers are used on ~98% of plots.

A majority of the plots using no fertilizer are concentrated in the Lower Gangetic plains.



Fertilization Categories	Count	Percent
Nitrogen + Phosphorus	1,852	32.3
Nitrogen + Phosphorus + Potassium	1,325	23.11
Nitrogen Only	801	13.97
Nitrogen + Phosphorus + Zinc	596	10.4
Nitrogen + Phosphorus + Zinc + Potassium	395	6.89
Nitrogen + Potassium	290	5.06
Nitrogen + Zinc	290	5.06
No Fertilizer	92	1.6
Nitrogen + Potassium + Zinc	58	1.01
Phosphorus Only	25	0.44
Others	4	0.07
Phosphorus + Potassium	4	0.07
Potassium Only	1	0.02

The plots using no fertilizer and only phosphorus are concentrated in the LGP.

Zinc usage is restricted largely to the TGP.

Phosphorus usage is very low in TGP while reliance on Nitrogen is very high.

Potassium usage is concentrated in the LGP.

Clearly, type of fertilizers used differ by the zones.

Examination of the Trend: Input Decisions

Tillage (Land Preparation)
- Type for current and previous crop.
  - Directly Sown
  - Transplanted

Crop Establishment
- Date (Sowing + Nursery Establishment)
- Determinants of Timing of Transplanting and Nursery Establishment.

Fertilization
- Basal + First + Second + Third Topping

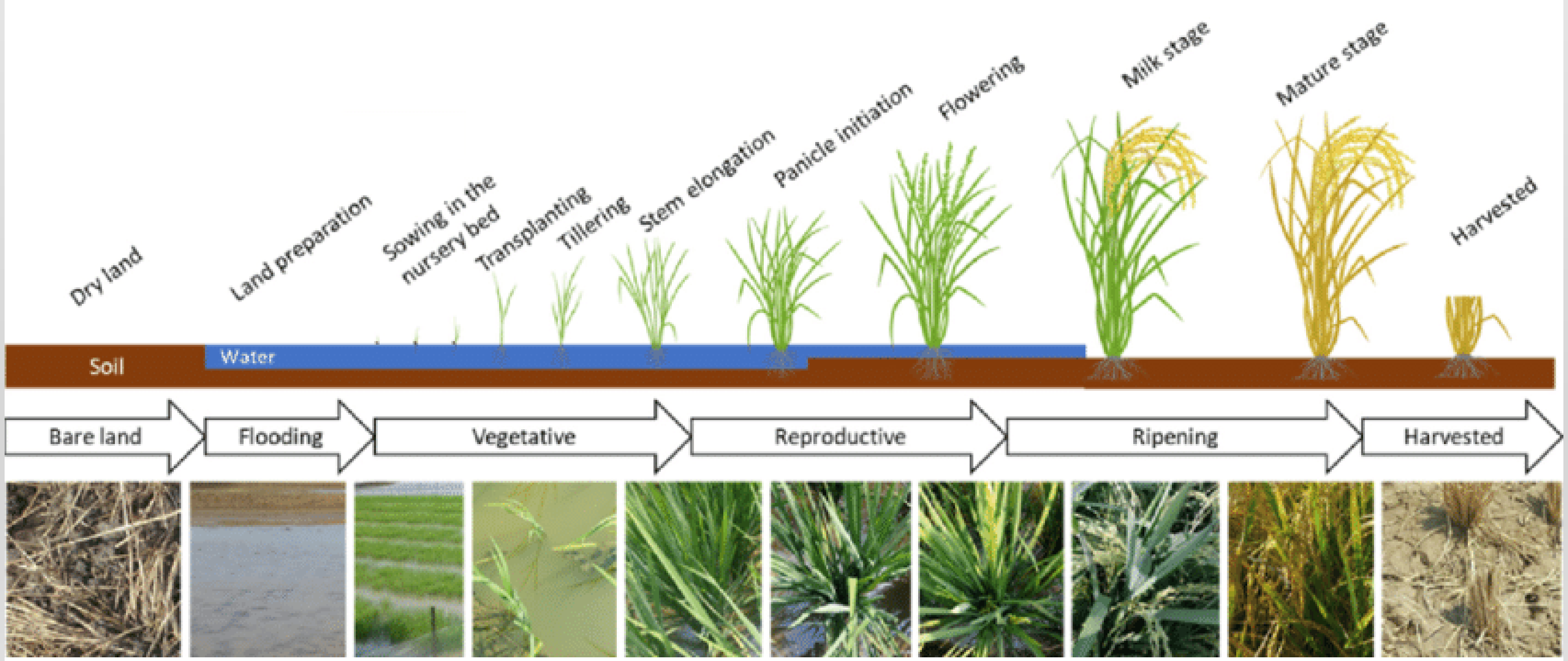
For each Topping:
- Amount
- Days after Sowing
- Type: FYM (Wet vs. Dry), DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron
- Source of Information (IGNORED)

Irrigation
- Frequency
- Availability
- Source
  - Depth of Tubewell
  - Energy Source of Pump
  - Flat Pipes
- Stage of Crop Growth at which Irrigation Application Occurs

Crop Protection
- Weed and Insect Severity
- Herbicide Application Timing and Frequency
- Manual Weeding (First + Second + Third) Timing

Harvesting
- Harvest Date
- Harvest Method
- Threshing Type

Crop Growth Stage	Count	Percentage
Pre Seeding	3611	67.61
Tillering Initiation	3336	62.46
Tillering	3486	65.27
Panicle Initiation	3070	57.48
Heading: Panicle Emergence	2143	40.12
Boot Leaf Stage	2883	53.98
Flowering	2907	54.43
Flowering To Maturity	1746	32.69
None	178	
CRI	1	
Jointing	1	



Source: Image taken from Ramadhani et. al. 2021

Irrigation:
Stage of Crop Growth at which Irrigation Application Occurs

From here, it is clear that irrigation is more prevalent in the early stages of the crop growth cycle.

Examination of the Trend: Input Decisions

Tillage (Land Preparation)

- Type for current and previous crop.
    - Directly Sown
    - Transplanted

Crop Establishment

- Date (Sowing + Nursery Establishment)
- Determinants of Timing of Transplanting and Nursery Establishment.

Fertilization

- Basal + First + Second + Third Topping

For each Topping:

- Amount
- Days after Sowing
- Type: FYM (Wet vs. Dry), DAP, NPK, Urea, MoP, SSP, TSP, ZnSo4, Gypsum, Boron
- Source of Information (IGNORED)

Irrigation

- Frequency
- Availability
- Source
    - Depth of Tubewell
    - Energy Source of Pump
    - Flat Pipes
- Stage of Crop Growth at which Irrigation Application Occurs

Crop Protection

- Weed and Insect Severity
- Herbicide Application Timing and Frequency
- Manual Weeding (First + Second + Third) Timing

Harvesting

- Harvest Date
- Harvest Method
- Threshing Type

Only LAT+LONG:

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.18	3389 | 0.07	803 | 0.21
Longitude	2.55*** (0.18)	-0.96*** (0.07)	-4.70*** (0.26)
Latitude	0.64*** (0.09)	-0.24** (0.11)	3.03*** (0.78)
Constant	-233.28*** (16.43)	105.37*** (7.79)	296.19*** (26.54)

*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

Only Weather:

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.0065	3389 | 0.0459	803 | 0.20
Total Rainfall (in mm.)	0.0015** (0.0007)	0.0031*** (0.0003)	-0.027*** (0.0015)
Average Temperature (in degree celsius)	-0.14 (0.20)	1.755*** (0.23)	-2.44*** (0.25)
Constant	18.76*** (5.93)	-32.64*** (6.11)	110.31*** (7.35)

*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

LAT+LONG+WEATHER:

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.2549	3389 | 0.0928	803 | 0.2514
Longitude	1.62*** (0.30)	-0.85*** (0.09)	-0.32 (1.48)
Latitude	3.14*** (0.32)	-1.82*** (0.22)	5.49*** (1.04)
Total Rainfall (in mm.)	-0.0009 (0.0009)	0.007*** (0.0009)	-0.02*** (0.006)
Average Temperature (in degree celsius)	4.65*** (0.68)	-0.43 (0.32)	-2.50*** (0.27)
Constant	-319.81*** (18.60)	140.19*** (15.48)	-28.39 (130.50)

*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

Accounting for either weather or latitude, longitude should suffice. (Correlation is 0.7***)

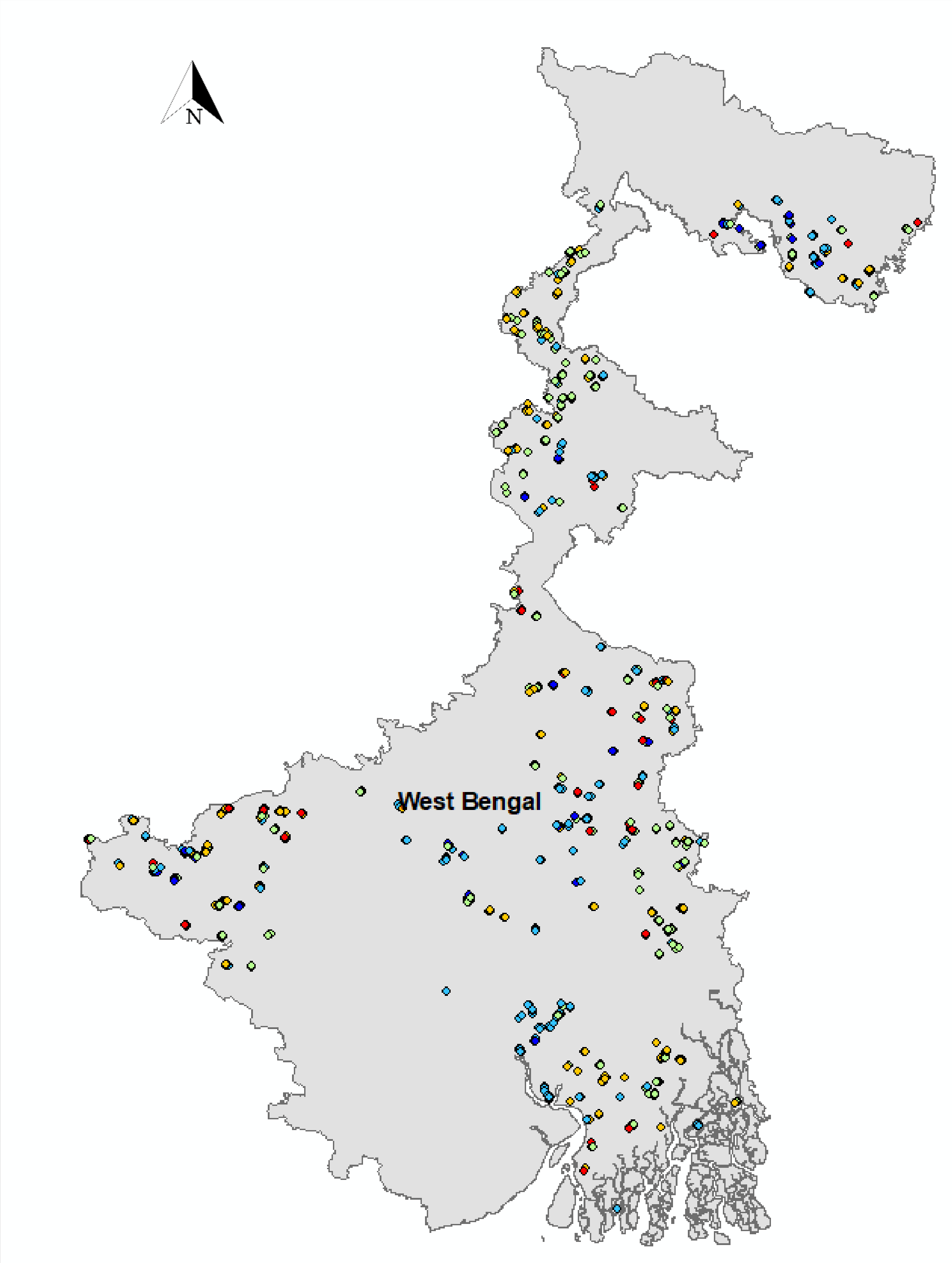
Weather + Inputs:

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
N | R^2	1535 | 0.21	3389 | 0.11	803 | 0.22
Farm Yard Manure Indicator	-1.61*** (0.25)	0.40** (0.16)	-1.14*** (0.33)
Fertilization Frequency	0.34* (0.19)	0.02 (0.13)	-0.44 (0.40)
Nitrogen (kg/acre)	0.0010 (0.0015)	0.06*** (0.007)	0.02 (0.01)
Phosphorous (kg/acre)	0.019** (0.008)	0.02*** (0.009)	-0.001 (0.02)
Potassium (kg/acre)	0.039*** (0.007)	-0.004 (0.011)	0.03 (0.04)
Zinc (kg/acre)	0.17*** (0.06)	0.01*** (0.002)	0.09 (0.07)
Irrigation Availability (Yes = 1, 0 otherwise)	3.69*** (0.33)	3.25*** (0.66)	0.59 (1.81)
Insect Control (No Presence + Control: 2 Presence + Control: 1 No Presence + No Control: 0 Presence + No Control: -1)	-0.14 (0.16)	0.65*** (0.10)	0.24 (0.26)
Total Rainfall (in mm.)	0.0021*** (0.0007)	0.003*** 0.0003	-0.02*** (0.002)
Average Temperature (in degree celsius)	-0.37* (0.21)	1.12*** (0.25)	-2.25**** (0.26)
Constant	19.92*** (6.17)	-23.72*** (6.69)	104.52*** (7.95)

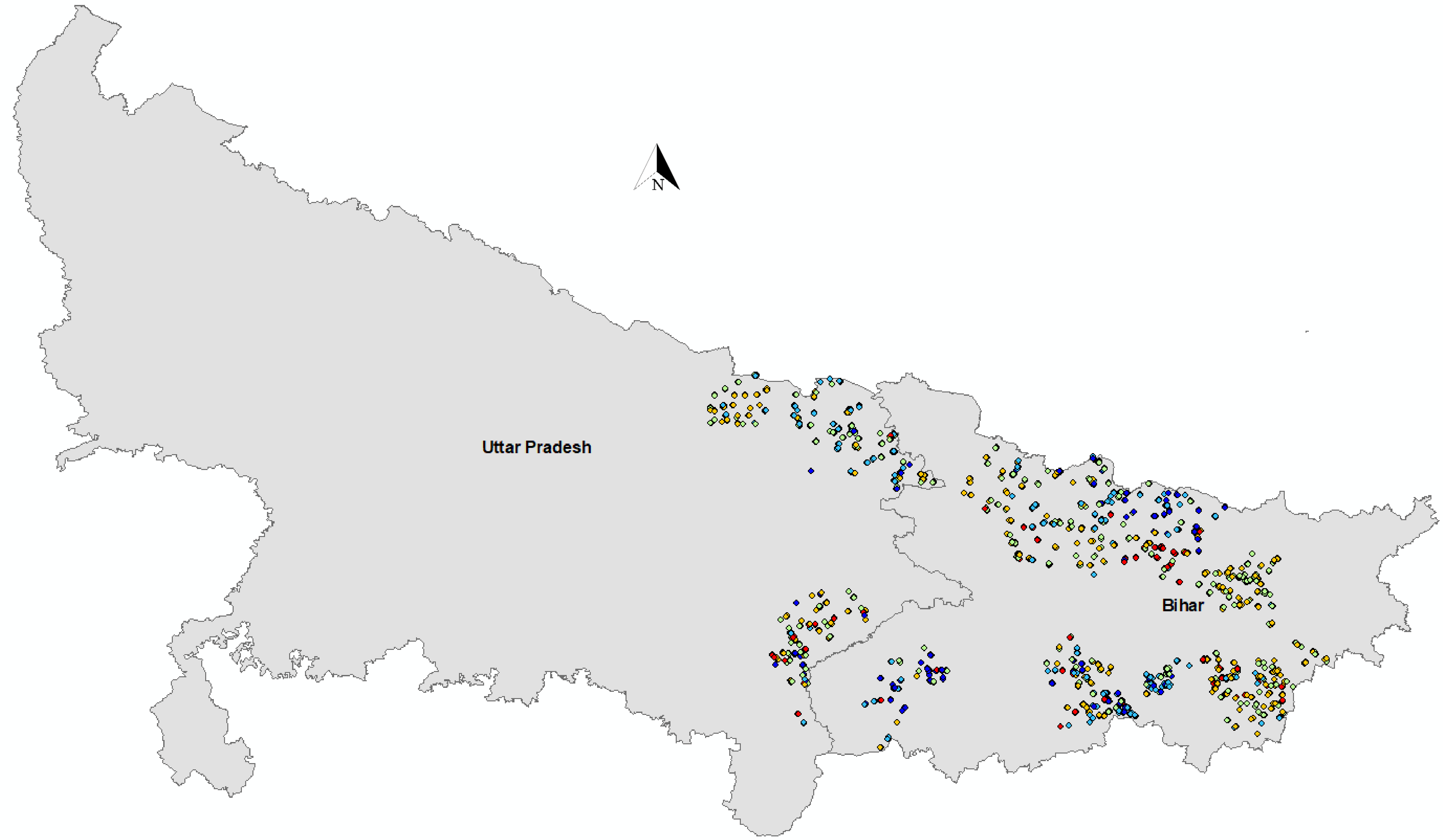
*: $p < 0.10$ | **: $p < 0.05$ | ***: $p < 0.01$

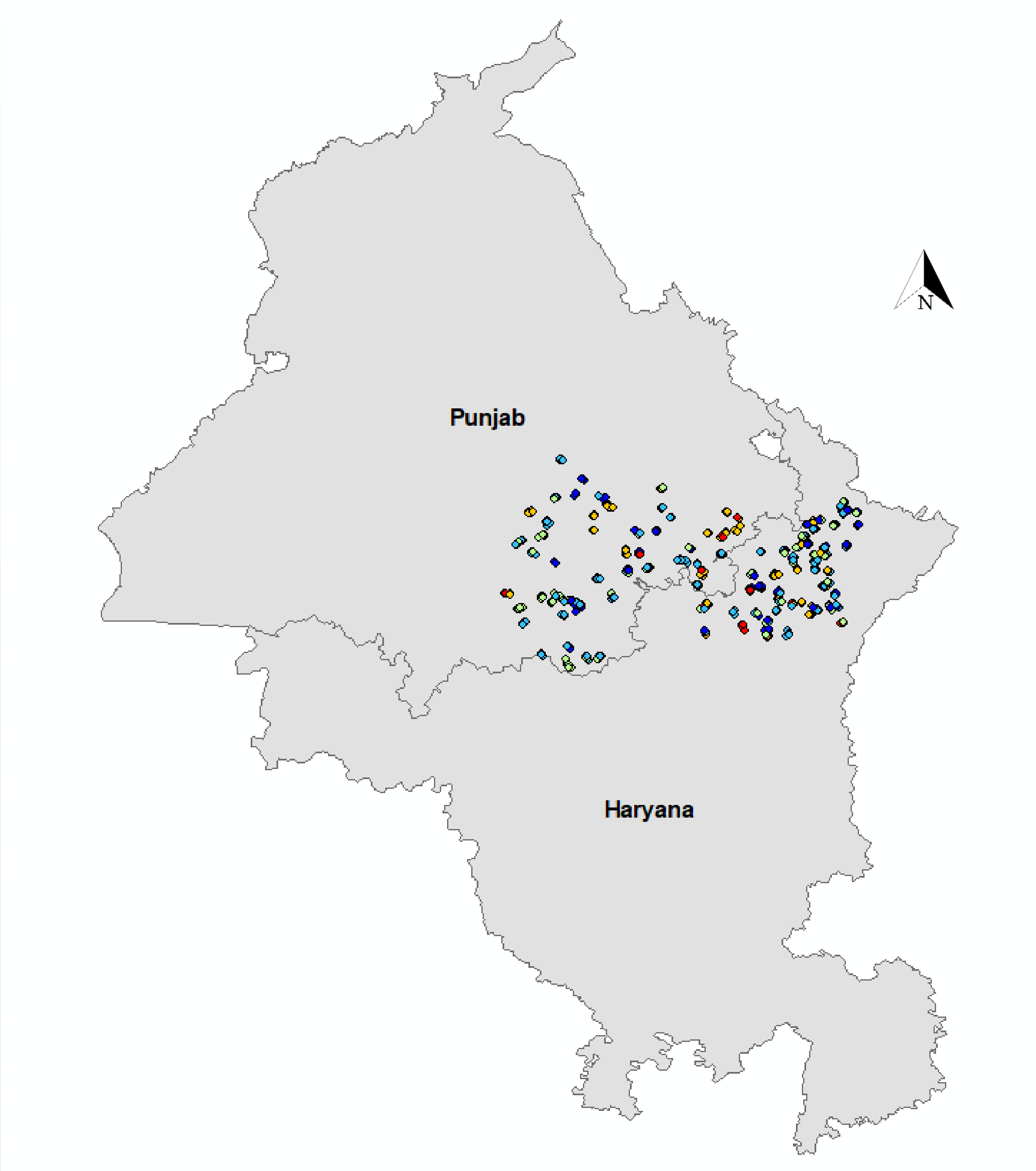
Weather + Inputs + Soil

Dependant Variable: Rice Yield (quintal/acre)	Lower Gangetic Plains	Middle Gangetic Plains	Trans-Gangetic Plains
$N \mid R^2$	1535 | 0.24	3389 | 0.15	803 | 0.22
Farm Yard Manure Indicator	-1.51*** (0.25)	0.38** (0.15)	-1.12*** (0.33)
Fertilization Frequency	0.23 (0.19)	-0.001 (0.13)	-0.44 (0.42)
Nitrogen (kg/acre)	0.001 (0.001)	0.07*** (0.006)	0.02 (0.01)
Phosphorous	0.021*** (0.007)	0.03*** (0.01)	-0.003 (0.02)
Potassium	0.03*** (0.007)	-0.003 (0.01)	0.03 (0.04)
Zinc	0.19*** (0.06)	0.01*** (0.002)	0.08 (0.07)
Irrigated Area	3.53*** (0.32)	3.46*** (0.71)	0.57 (1.84)
Insect Control (No Presence + Control: 2 Presence + Control: 1 No Presence + No Control: 0 Presence + No Control: -1)	-0.20 (0.17)	0.58*** (0.10)	0.31 (0.27)
Total Rainfall (in mm.)	0.0019*** (0.0007)	0.003*** (0.0003)	-0.02*** (0.002)
Average Temperature (in degree celsius)	-0.28 (0.21)	1.40*** (0.25)	-2.21*** (0.26)
Soil Texture (0,1,2: Light, Medium, Heavy)	-0.16 (0.21)	0.87*** (0.23)	-0.37 (0.35)
Perceived Soil Quality (0,1,2: Low, Medium, High)	1.67*** (0.27)	0.91*** (0.28)	-0.18 (0.97)
Drainage Class (0,1,2: Low, Medium, Up)	0.52*** (0.17)	-1.50*** (0.15)	-0.09 (0.44)
Constant	13.57** (6.15)	-30.77*** (6.52)	104.64*** (8.19)



What does detrending do?


N
Uttar Pradesh
Bihar



N
Punjab
Haryana