Understanding the spatial variation in prices as a function of irrigation input (groundwater levels and rainfall).

21 states
Kharif season
(May-December)
2017

Sajag Agarwal (ECE 2016185)

Sumedha Shukla (Research Associate, Econometrics Lab)

Background:

- Groundwater is the primary source of irrigation for ~62% of the net irrigated area (Government of India, 2016)
- Groundwater decline can affect economic outcomes in several ways.
- For example, reduction in groundwater levels leads to an increase in poverty and poses a threat to food production through declining land productivity (Sekhri, 2014; Seckler, et al., 1998).
- In this project, we seek to understand the relationship between groundwater levels and agricultural prices.
- Prices are an inherent component of farm revenue and have invited a significant amount of government intervention to help farmers realize better returns.

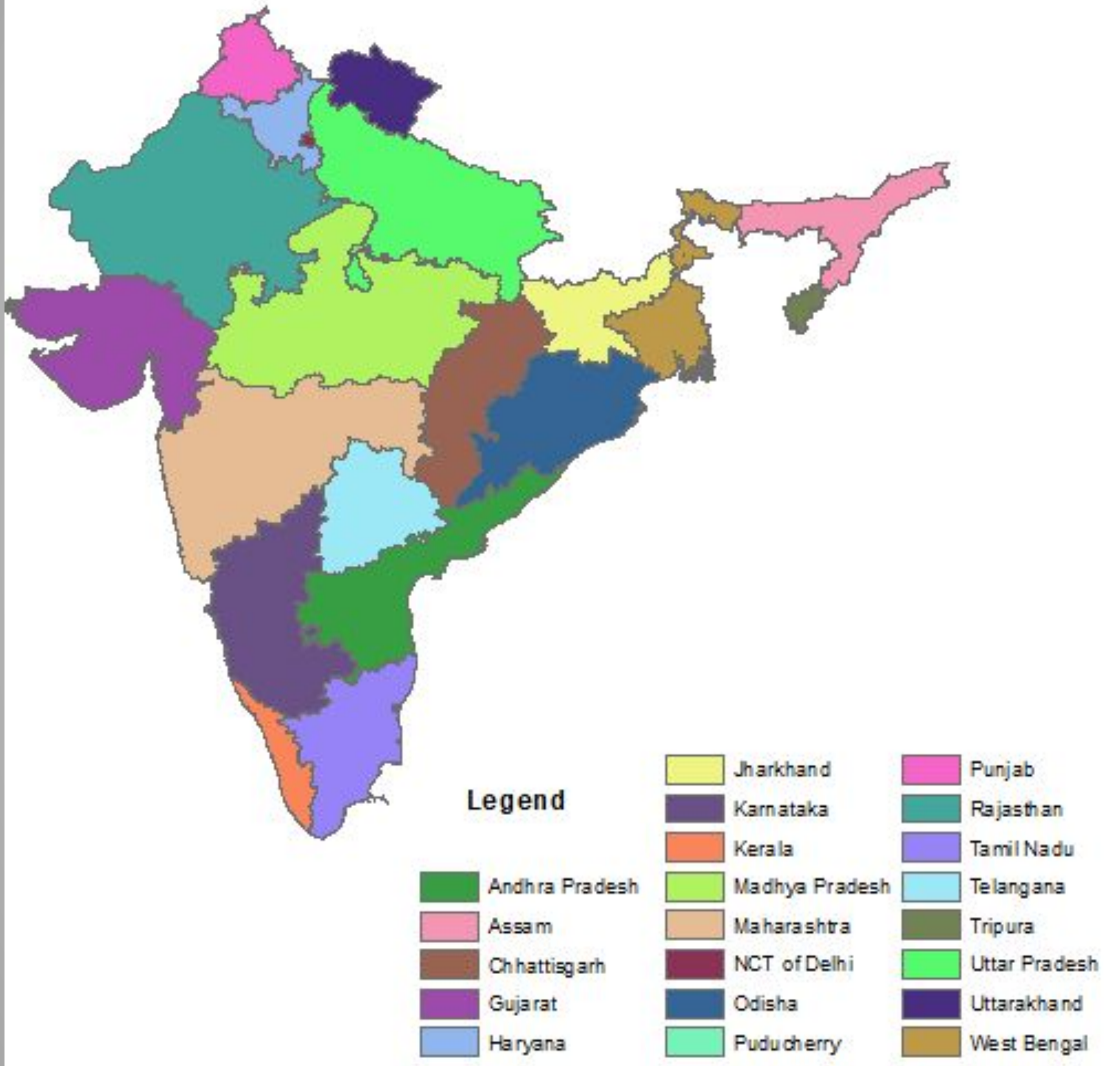
Background:

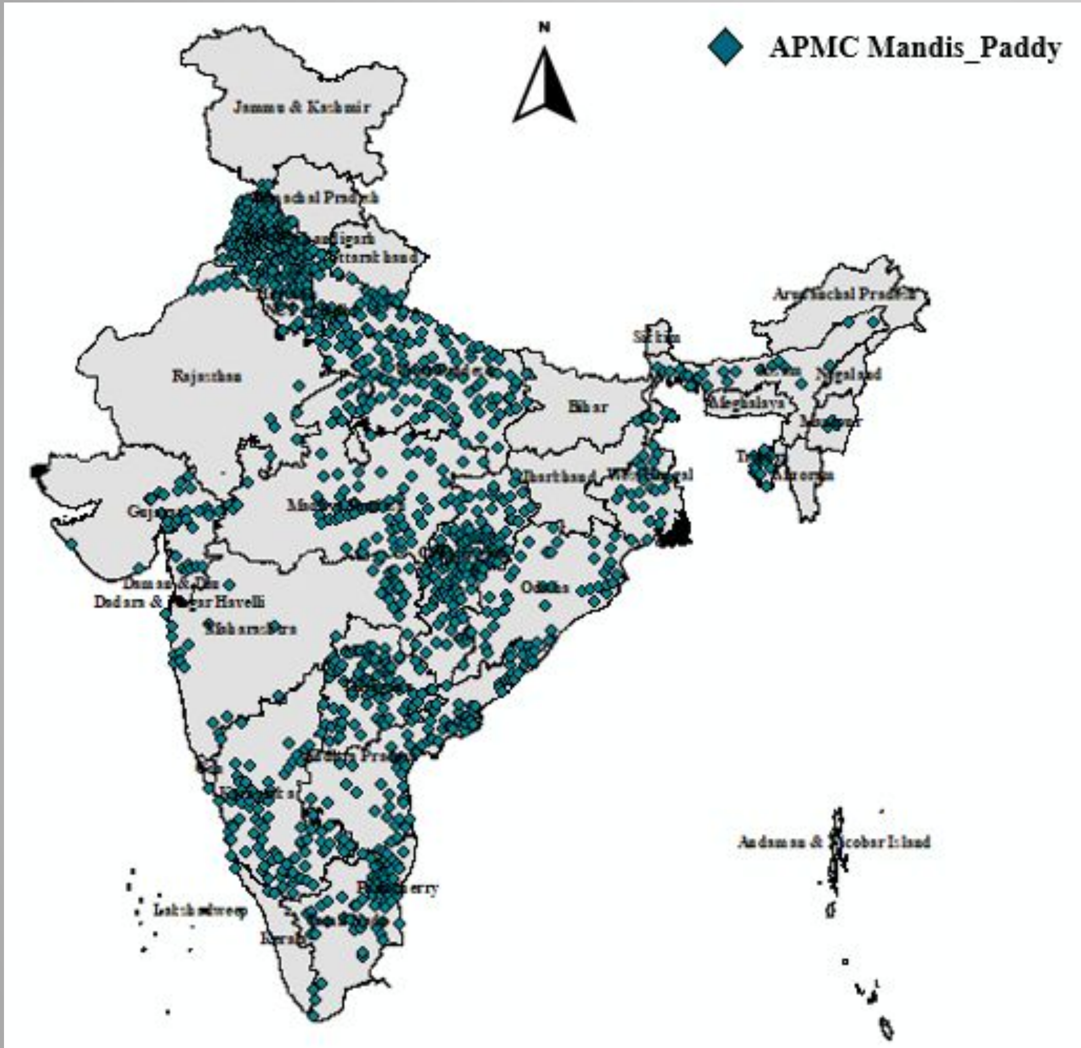
- Under the Agricultural Produce Marketing Acts state governments notify and designate commodities and market areas where regulated trade can take place.
- In APMC mandies, prices are discovered through open auction(s).
- Critically, once an area is declared a market area and falls under the jurisdiction of a Market Committee, no person or agency is allowed freely to carry on wholesale marketing activities elsewhere.

Data Sources:

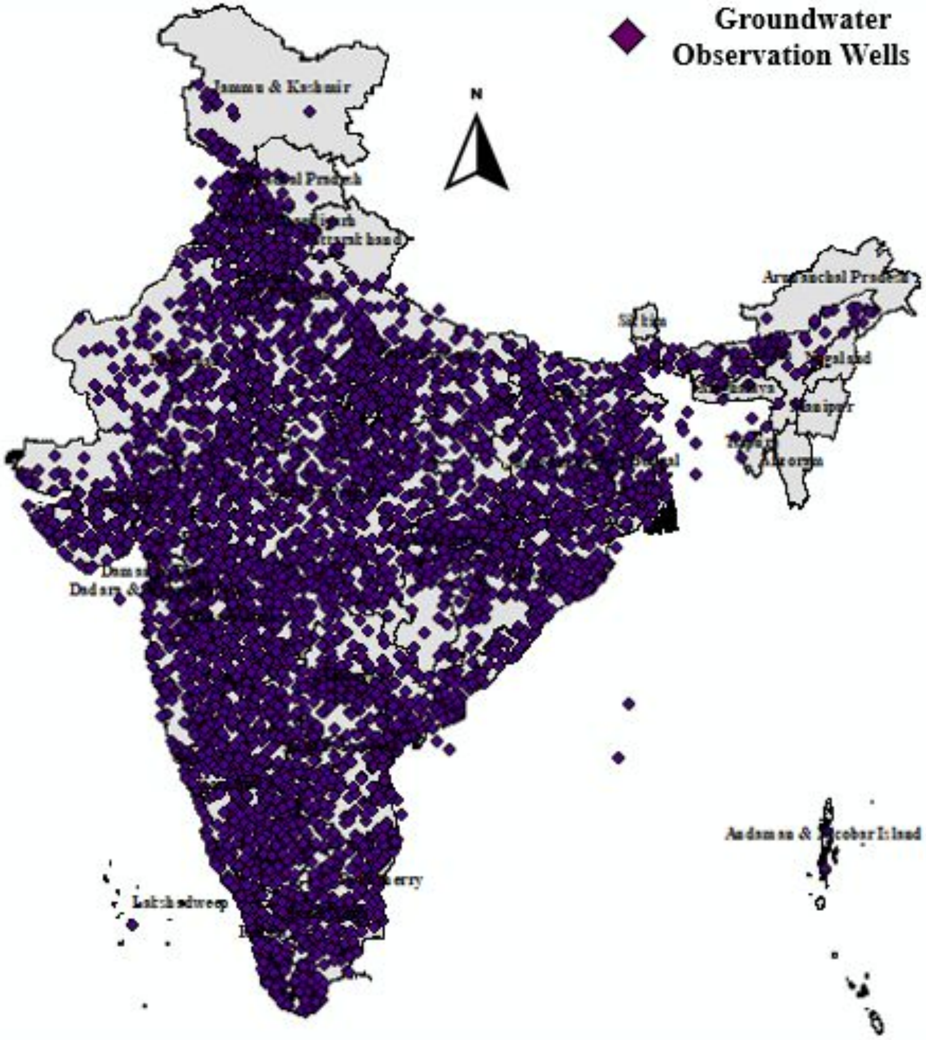
- Date wise prices for Paddy: Scraped from AgMarknet, an e-governance portal managed by Ministry of Agriculture and Farmers' Welfare. We use the modal price for our analysis.
- Groundwater (GW) data from India-Water Resources Information System (WRIS) which uses multiple agencies like CGWB and APWRIMS to compile the data. The data lists the depth to groundwater (meters below ground level) for each well at a monthly frequency.
- States: 21 | Period of analysis: Kharif, 2017 (GW: May-September | Modal Price: October-December)
- We find GW observation wells near each of the mandis (within 1.5 kilometers of each mandi). In case of multiple such wells, we take an average value.

States where paddy is exchanged in APMCs.





Visualizing the APMC Mandis where paddy is exchanged.
N = 1054



Visualizing the groundwater observation wells. | N=5573

Plan:

Exploratory Data Analysis:

- Basic summary statistics
- Why consider a spatial analysis at all?
- Stationarity : agro-ecological zones
- Zone-wise summary statistics

Spatial Regression Analysis:

- OLS model (Use Near tool to assign a GW well near to each Mandi (less than 1 kilometer))
- Limitations of the OLS model
- Justifying the use of Agro-ecological zones as a basis for stationarity
- Alternative model specifications
- Choosing the ‘best’ model

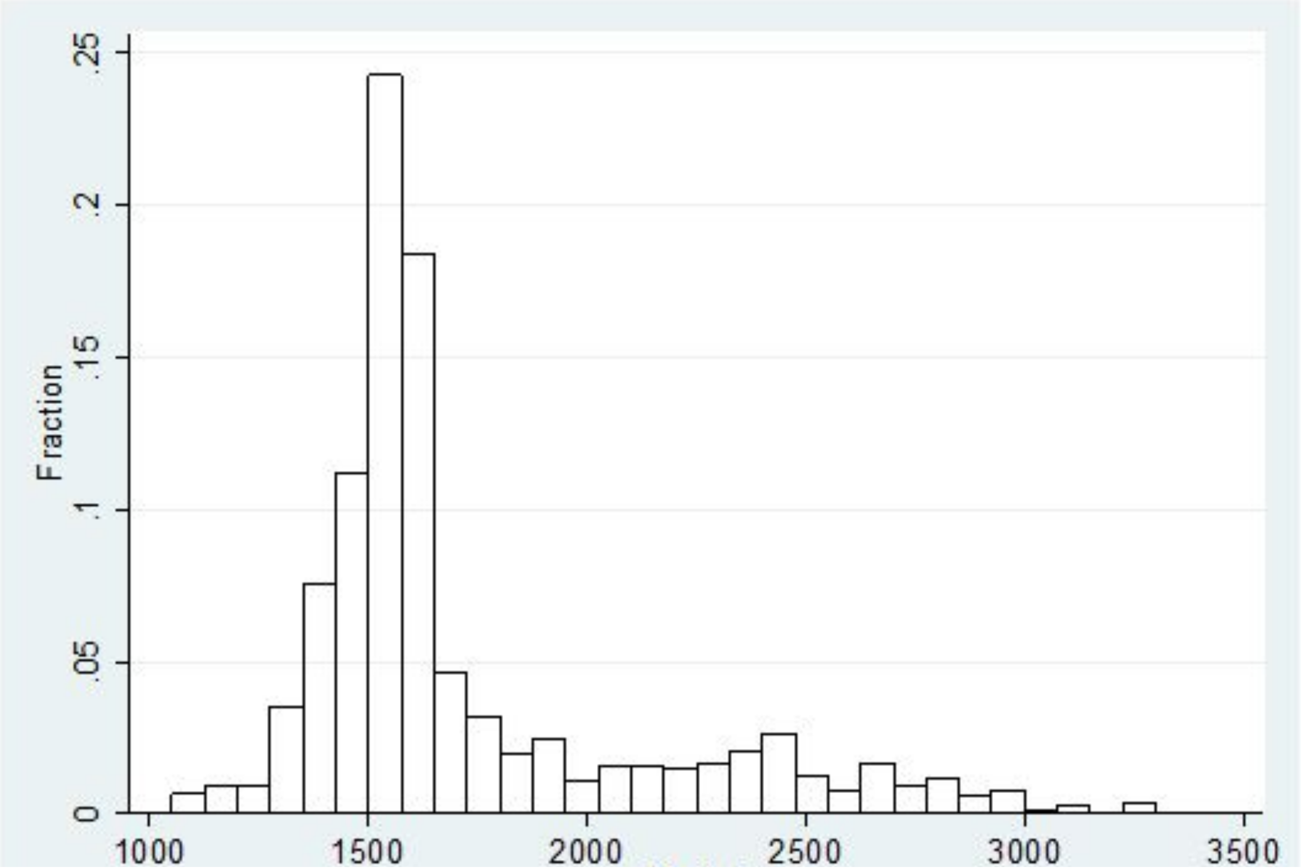
Revised Results

Major Changes:

- Modal Price was considered to be for post-harvest season: October - December (2017) while the groundwater and rainfall data was considered for the Kharif season: May - September (2017).
- Instead of using a Queen’s weight matrix, a 6-neighbor inverse-distance weight matrix was taken.

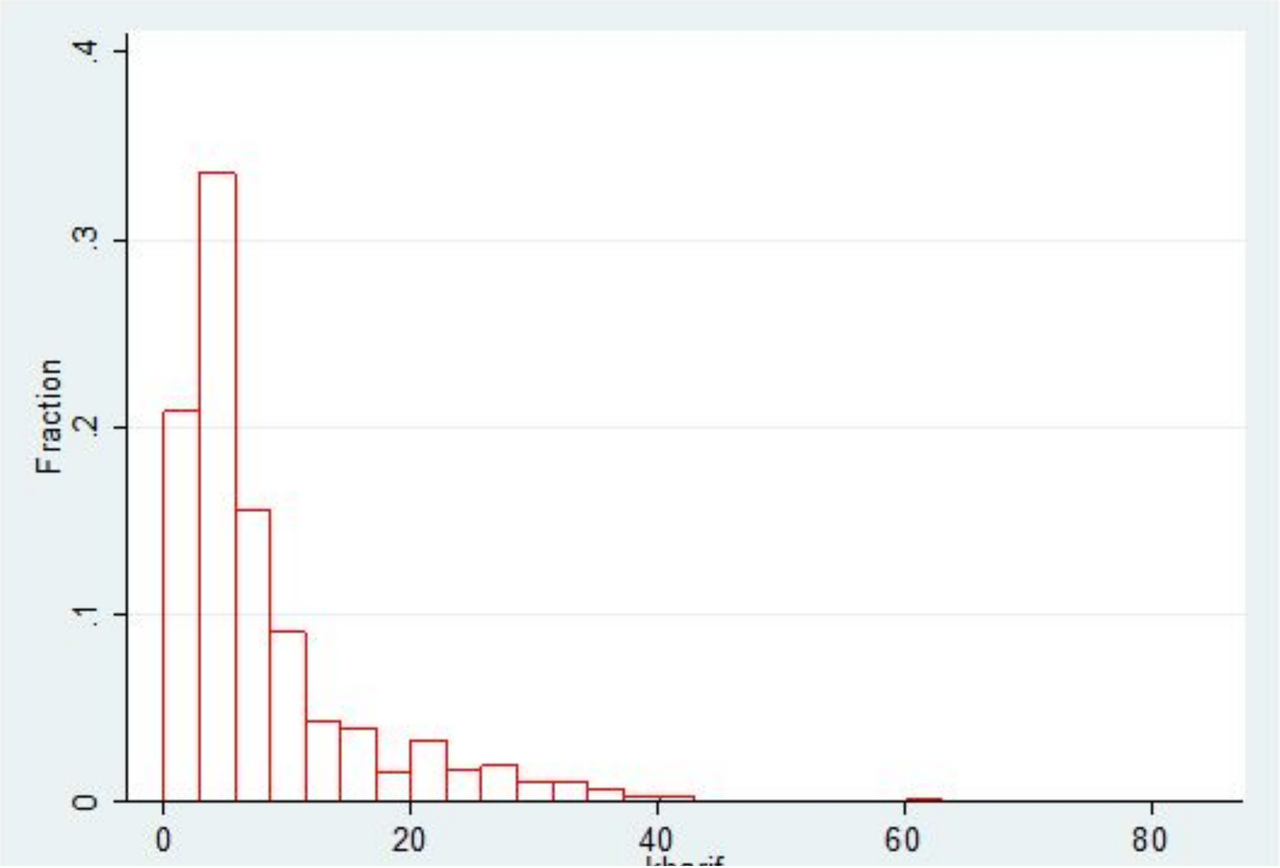
Summary statistics

Variable	N	Mean	SD	Minimum	Maximum
Groundwater	1025	8.46	9.01	0.04	85.89
Modal Price (Rupees/ Quintal)	1025	1725.98	407.53	1052	3301.25
Rainfall	1025	740.36	310.37	355.9	1926.1



Histogram for modal price, Kharif(2017)

- Outlier value: 3301.25
- Corresponds to: Haryana



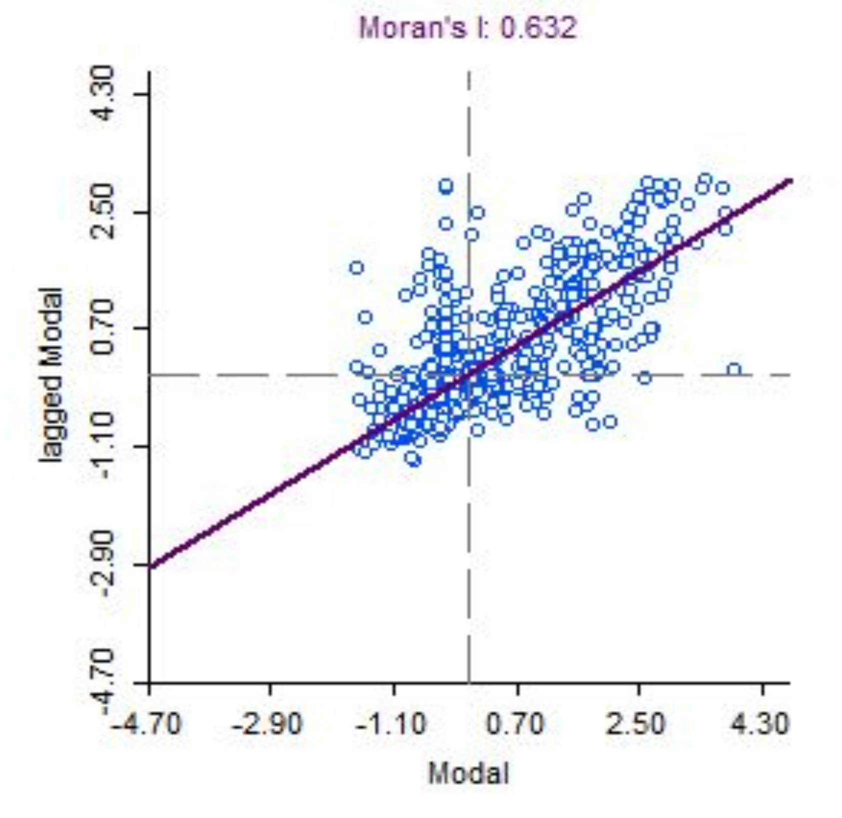
Histogram for depth to groundwater, Kharif(2017)

- Outlier values: >43.68 (P99)

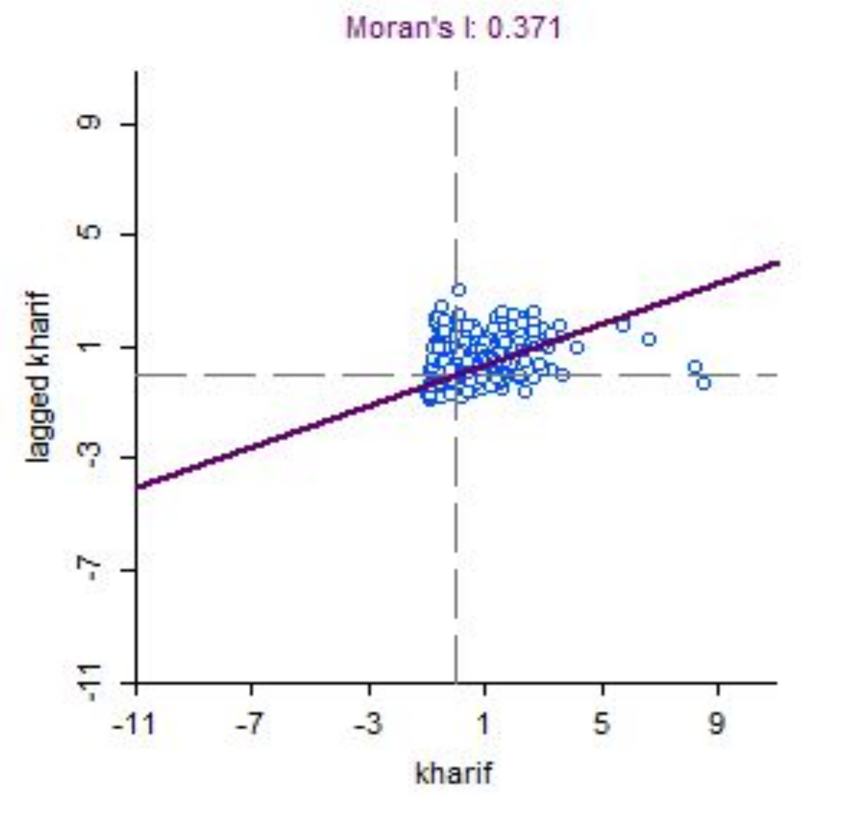
- Corresponds to:
- Karnataka (4 obs.)
- Gujarat (3 obs.)
- Punjab (2 obs.)

Variable	Moran's I statistic (K-Neighbors Inverse Distance Weights Matrix)
Modal Price	0.63***
GW	0.37***

- The Moran's I statistic is an indicator of spatial correlation. Higher the value, higher is the correlation between value at point i and value for neighbour(s) j.
- The high Moran's I statistics clearly indicate substantial spatial autocorrelation in the variables included in the analysis.

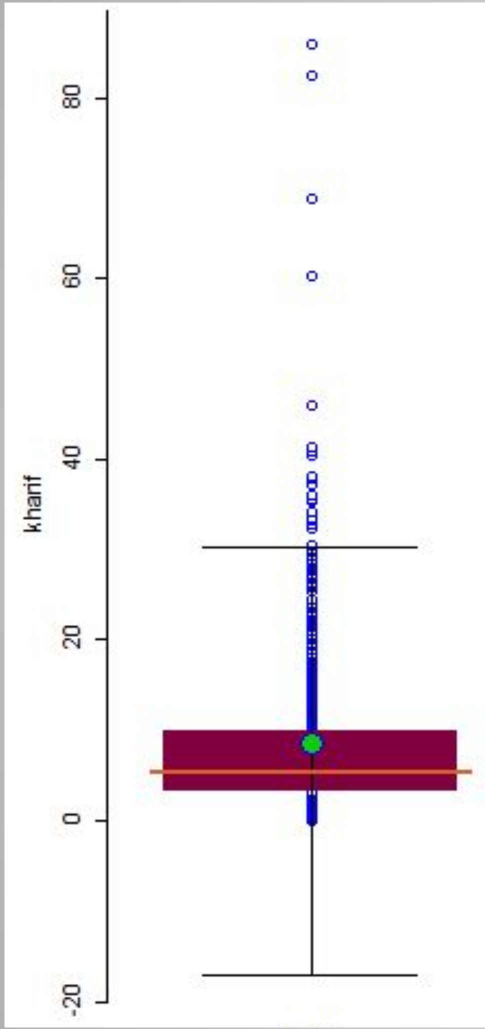


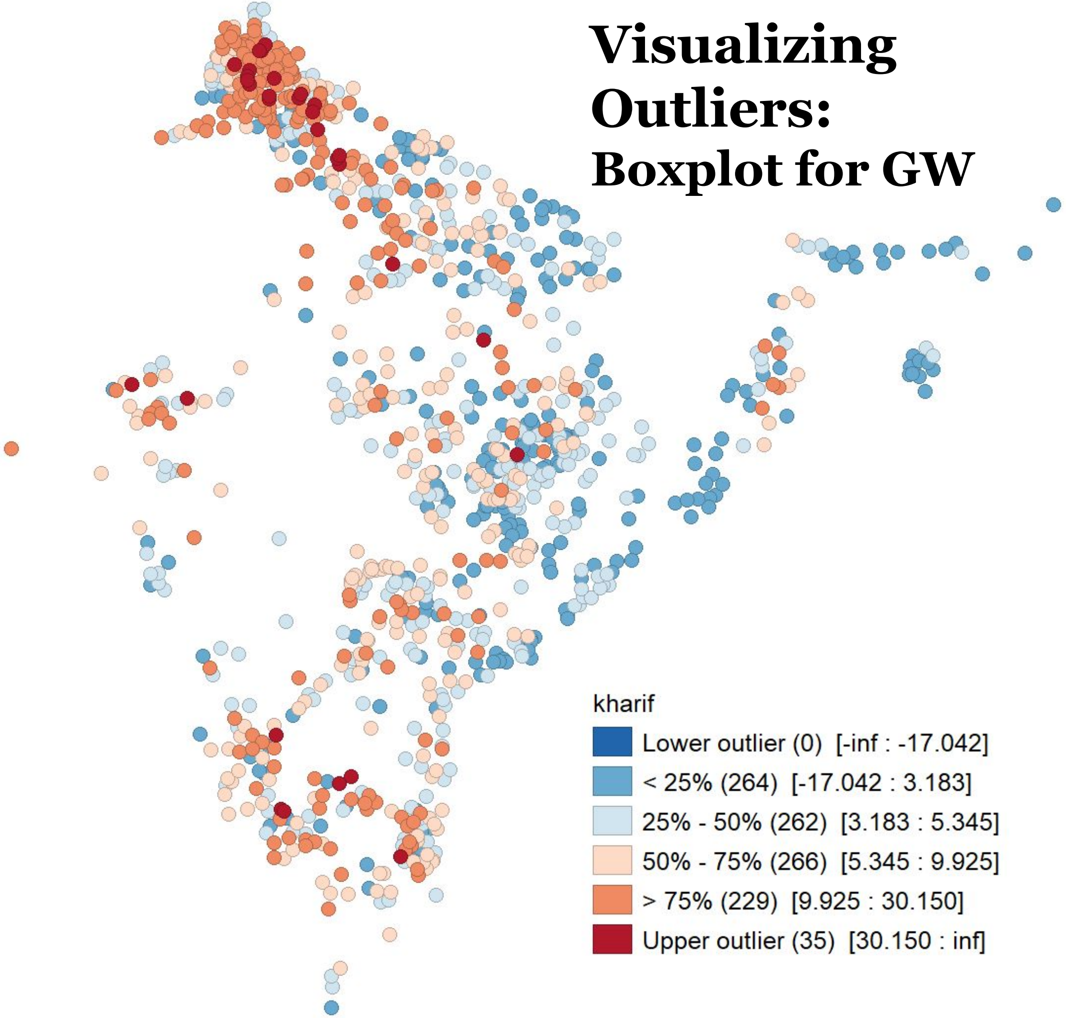
Moran's I (K-Neighbors Inverse Distance Weights Matrix) for GW



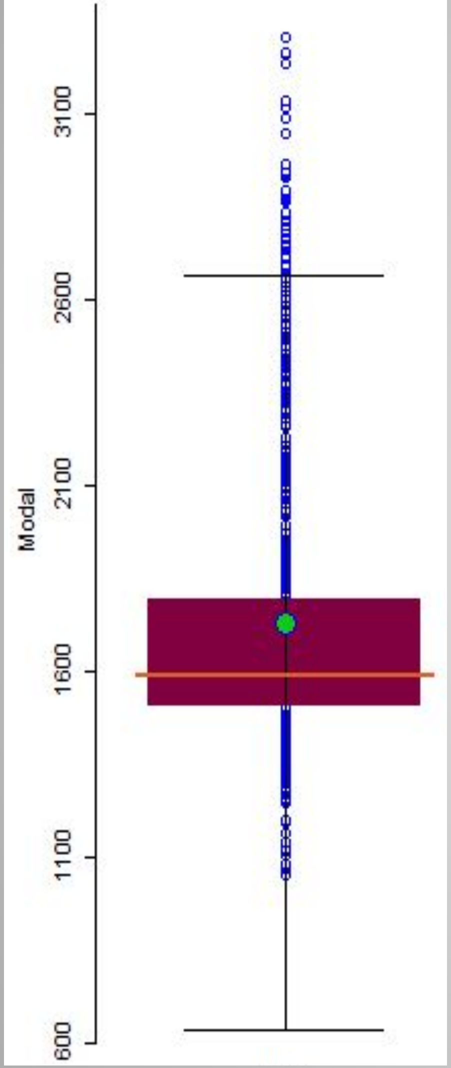
Moran's I (K-Neighbors Inverse Distance Weights Matrix) for Modal Price

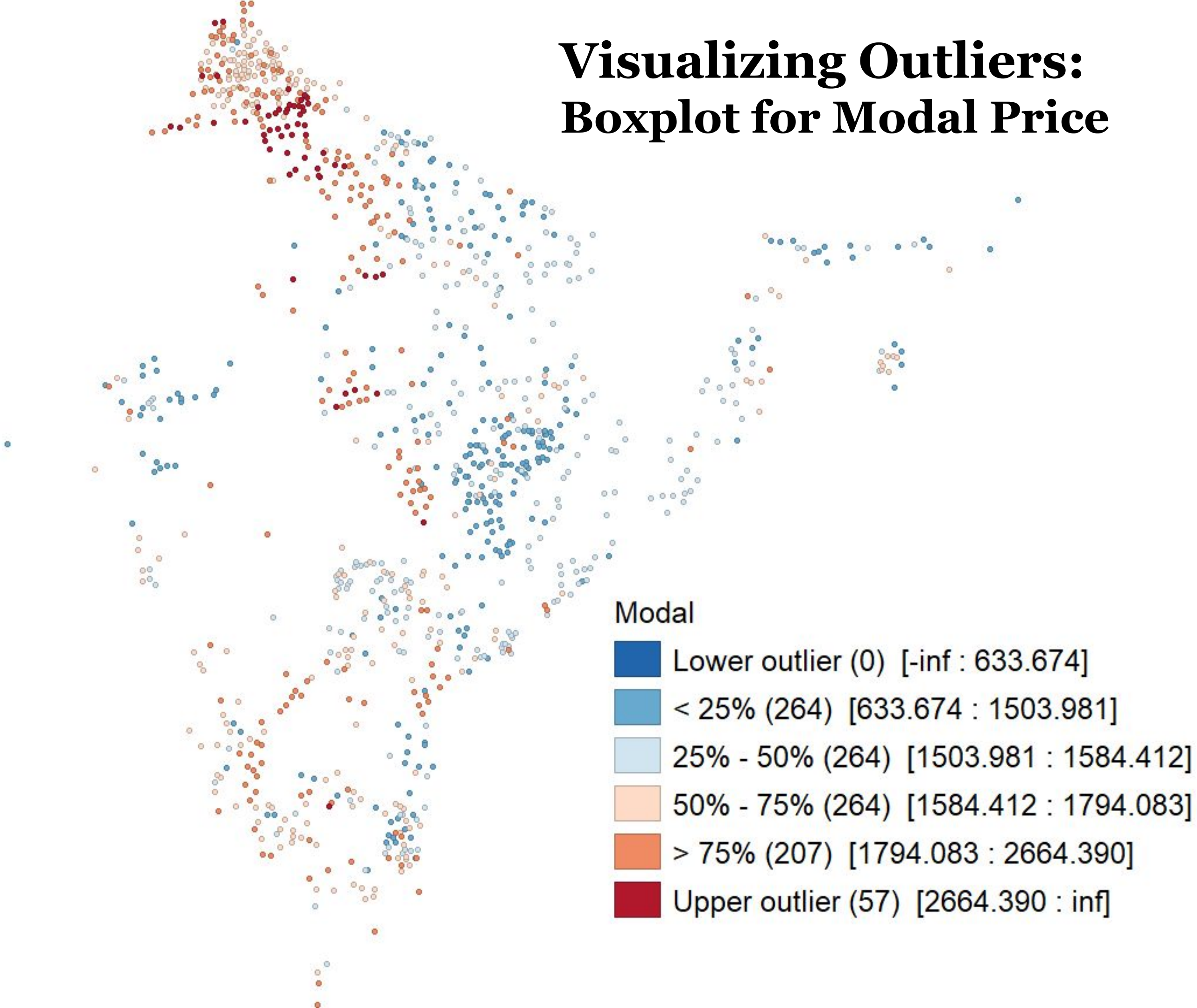
Visualizing Outliers: Boxplot for GW

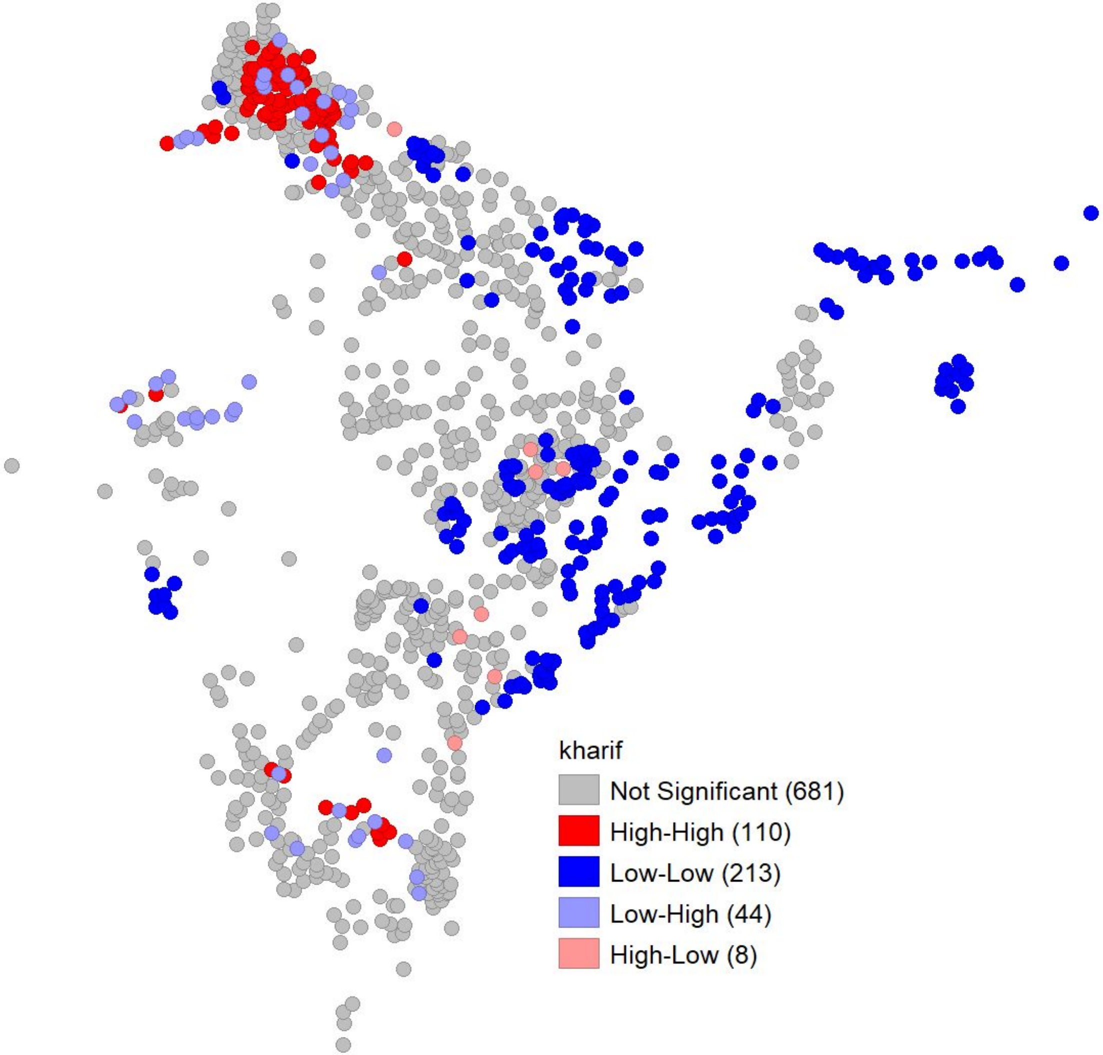




Visualizing Outliers: Boxplot for Modal Price







Local Indicators of Spatial Autocorrelation (LISA) Plot Groundwater

- The cluster map shows locations with a significant local Moran's statistic classified by the type of spatial correlation

- High-high GW level implies regions with high values (i.e. higher GW stress) are surrounded by regions with high GW stress.

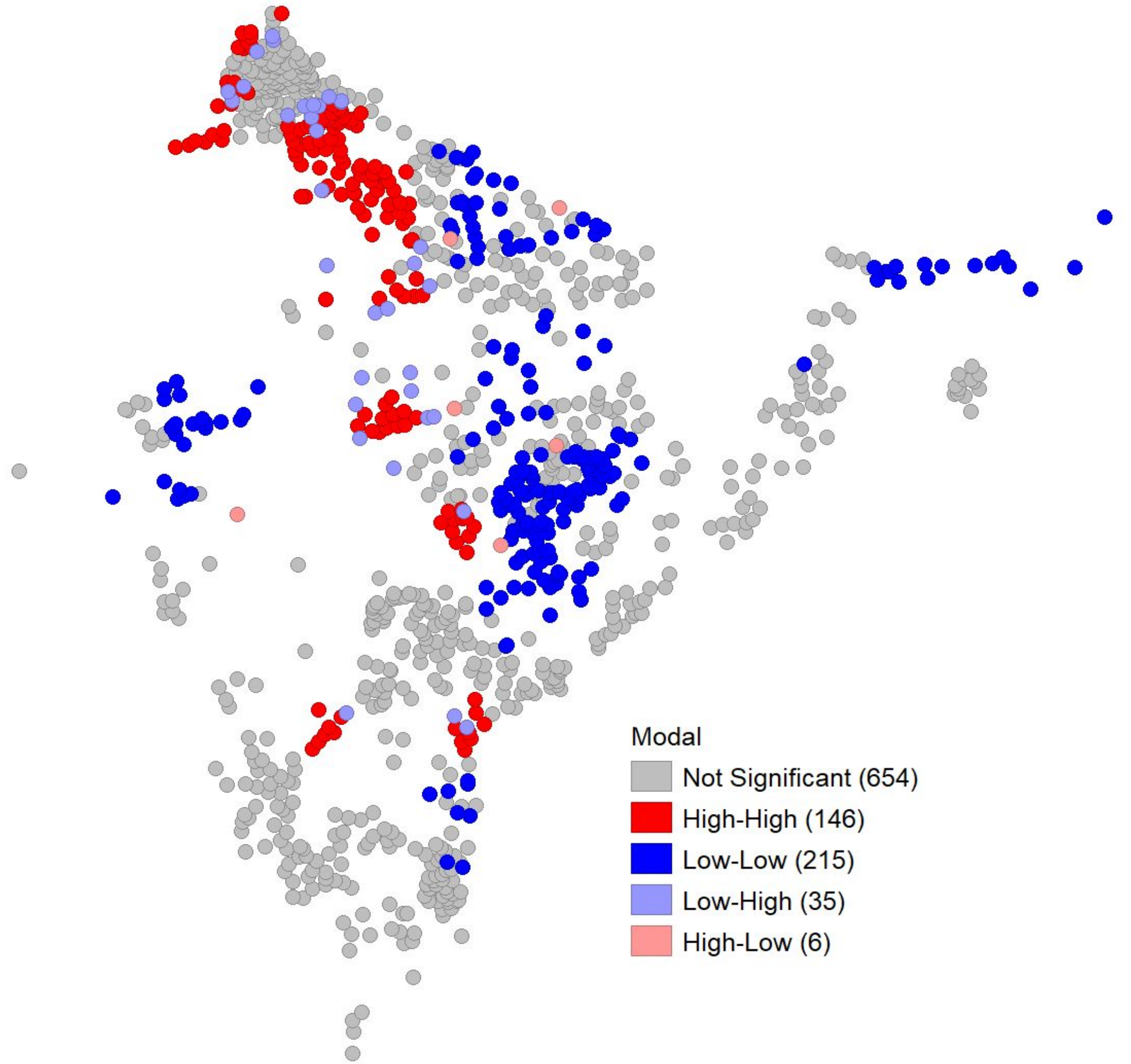
LISA Plot Groundwater

Low-low regions were usually in observed in states like Odisha, West Bengal and Assam.

- Here, West Bengal and Odisha seemed strange. Turns out, 2017 is a flood year for West Bengal with heavy rains in July-August.
- (Source:https://economictimes.indiatimes.com/news/economy/agriculture/over-2-lakh-hectares-land-damaged-in-west-bengal-rains/articleshow/48336815.cms?from=mdr)

High-high regions were usually observed in states like Punjab, Haryana, Delhi and some parts of Andhra Pradesh/Karnataka.

- Punjab, Haryana and Delhi have heavily stressed groundwater table as expected.



LISA Plot Modal Price

Low-low regions were usually in observed in states like Chhattisgarh, Assam, Uttar Pradesh and Jharkhand.

LISA Plot Modal Prices

High-high regions were observed in states like Punjab, Haryana, Delhi, and some parts of Andhra Pradesh.

- Interestingly, we don’t observe a cluster of high or low prices in West Bengal. This is because most of the crop itself got destroyed. But, in case of Jharkhand, we do observe a low-low cluster.

- The high-high regions for GW stress coincide with the high-high clusters of prices.

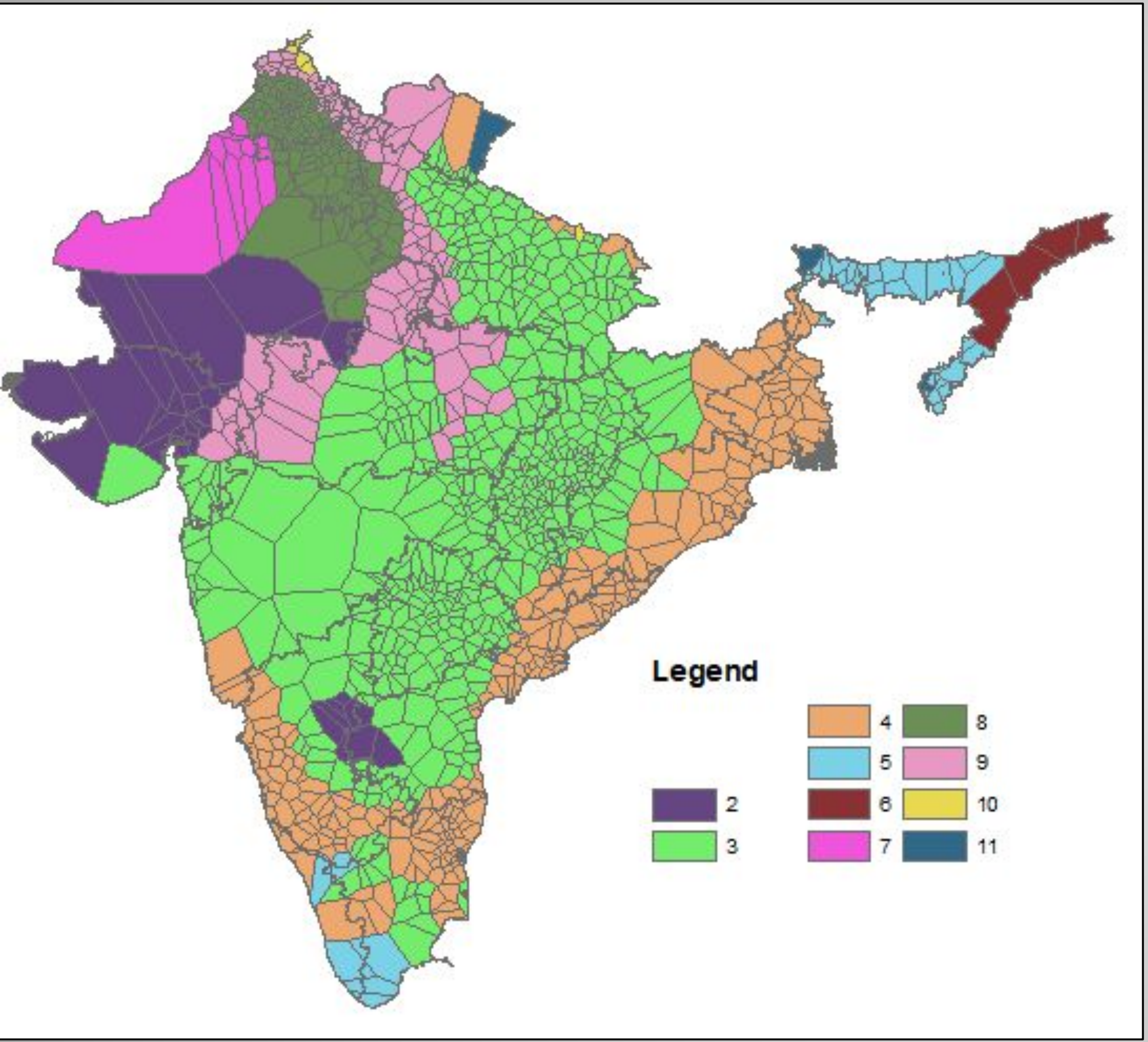
Stationarity Considerations: Agro-Ecological Zones (AEZs)

- Released by The Center for Sustainability and the Global Environment (SAGE) at the University of Wisconsin in 2007
- Divides the world into 18 zones dependent on:
- Temperature
- Rainfall
- Length of growing period
- Soil suitability
- Crop type

Agro-Ecological Zone List

Grid Code	Moisture Regime
1	Arid
2	Dry semi-arid
3	Moist semi-arid
4	Sub-humid
5	Humid
6	Humid; year round growing season
7	Arid
8	Dry semi-arid

Grid Code	Moisture Regime
9	Moist semi-arid
10	Sub-humid
11	Humid
12	Humid; year round growing season
13	Arid
14	Dry semi-arid
15	Moist semi-arid
16	Sub-humid



Visualizing the AEZs:

Broader classification used:

Arid/ Semi-arid regions: 2,3,7,8,9

- 770 observations

Humid/ Sub-humid regions: 4,5,6,10,11

- 255 observations

Zone Wise Summary Statistics:

Zone	Variable	Mean	SD	Minimum	Maximum
ARID	Groundwater	9.38	9.66	0.34	85.89
	Modal Price (Rupees/ Quintal)	1757.86	445.29	1052	3301.25
	Rainfall	694.29	244.28	355.9	1199
HUMID	Groundwater	5.67	5.91	0.04	37.88
	Modal Price (Rupees/ Quintal)	1629.7	238.4	1078.25	2783.33
	Rainfall	879.47	428.18	389.9	1926.1

Insights:

Mean(Humid Regions) - Mean(Arid Regions)	T-Value
Groundwater	-5.76***
Modal Price (Rupees/ Quintal)	-4.39***
Rainfall	8.52***

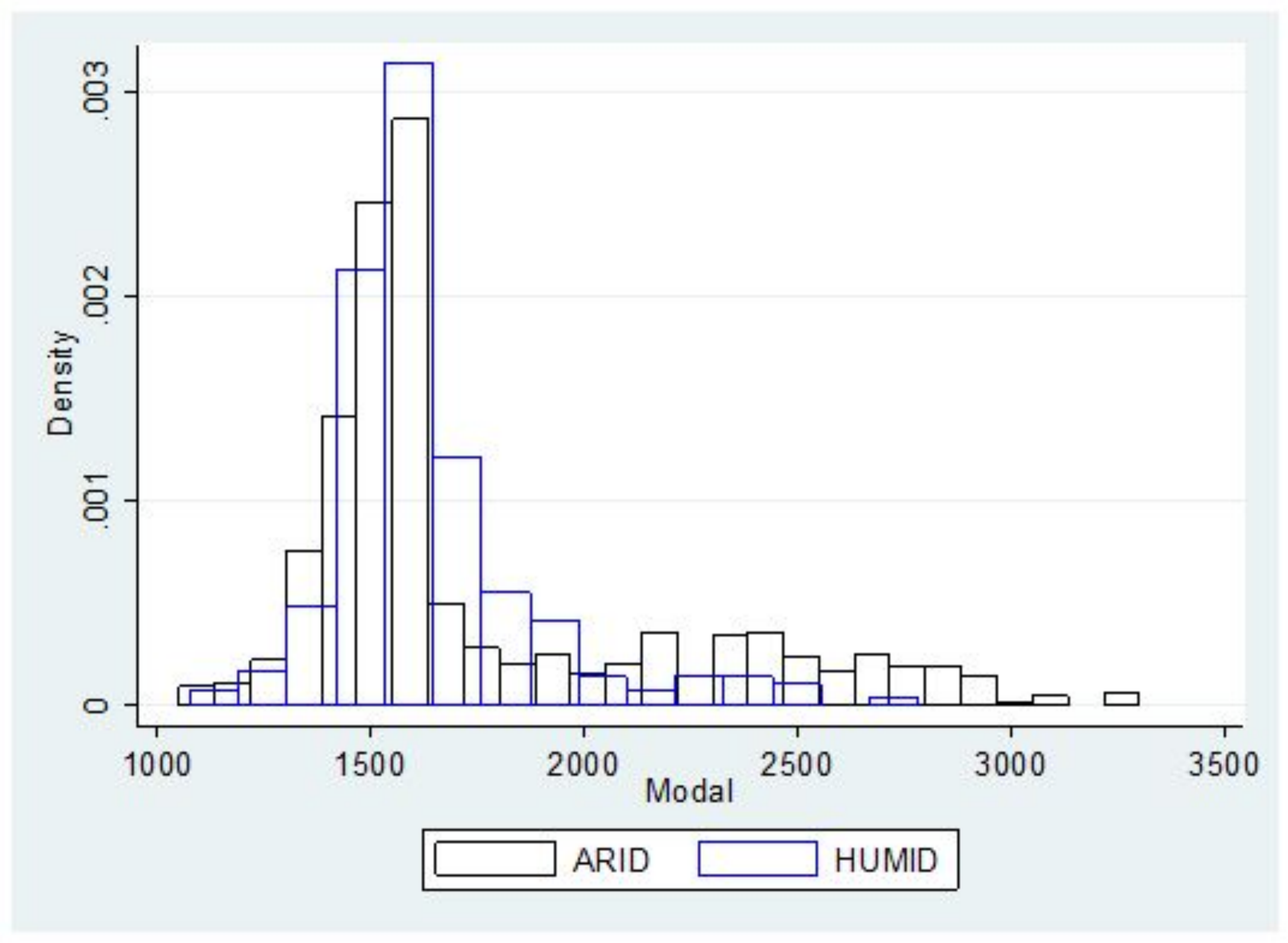
***$p \leq 0.01$ | **$p \leq 0.05$ | *$p \leq 0.10$

Arid regions exhibit higher GW stress on average as compared to humid regions.

Modal price in arid regions is higher on average as compared to humid regions.

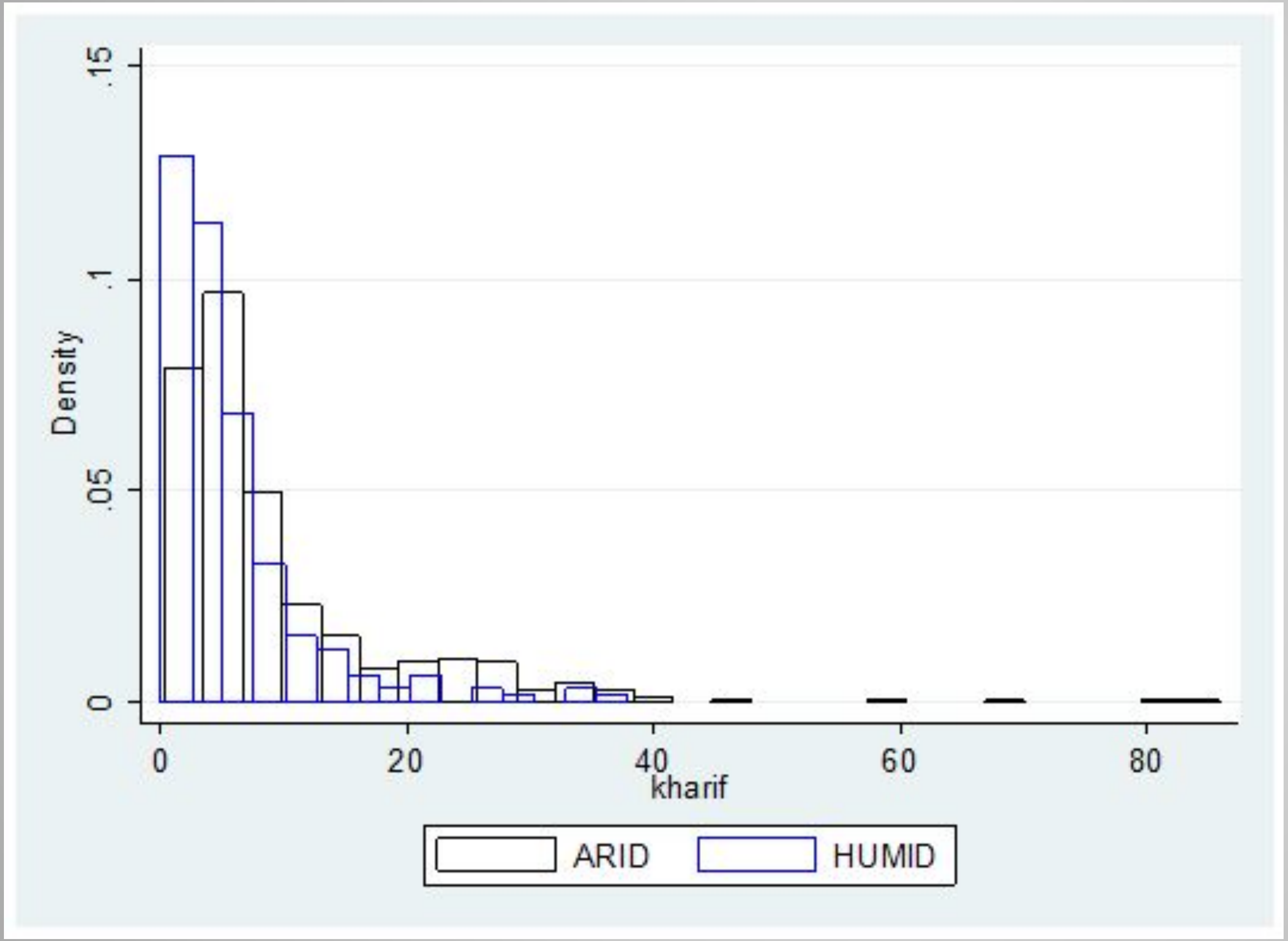
Rainfall is on average higher in humid regions as compared to arid regions.

Zone Wise Histogram: (Modal Price)



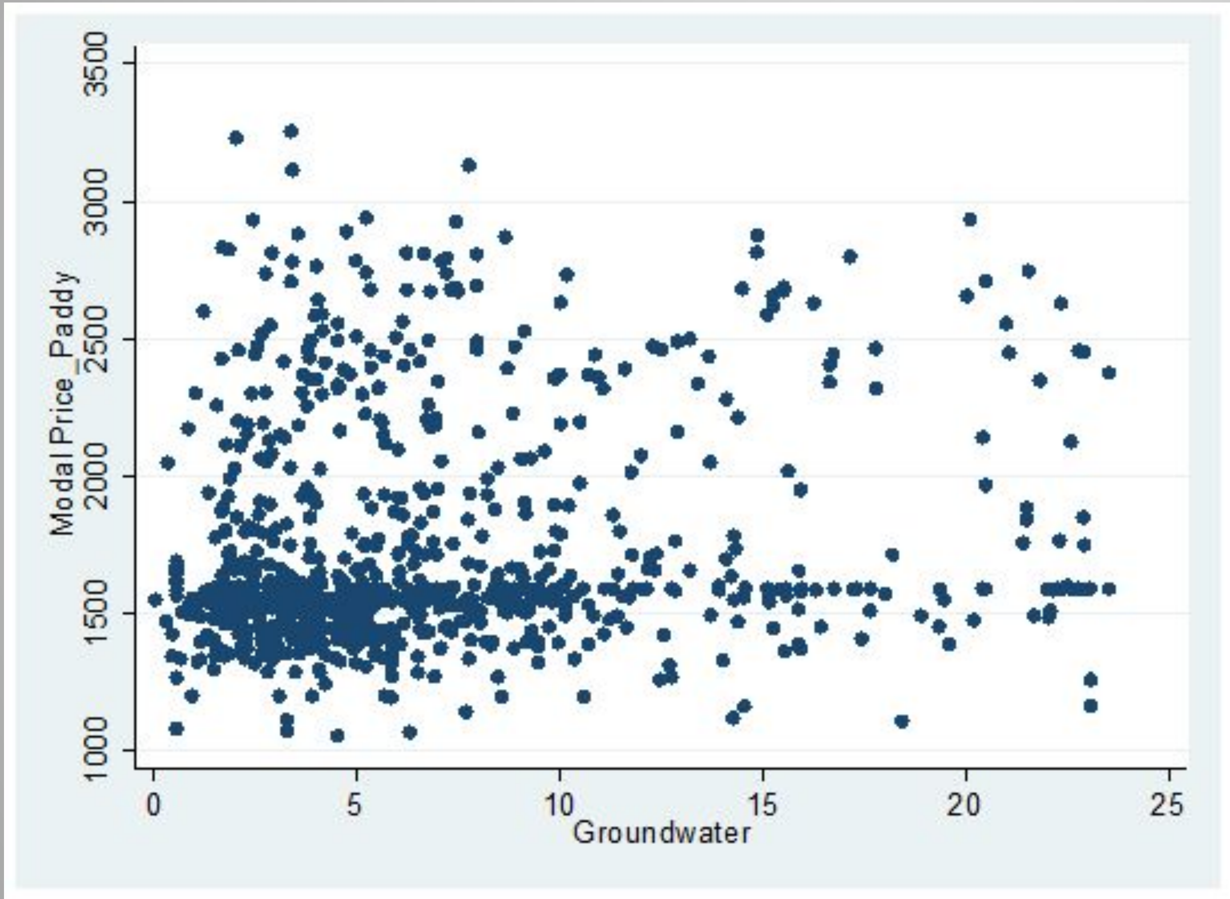
Humid regions have higher density for lower price values as compared to arid regions.

Zone Wise Histogram: Groundwater



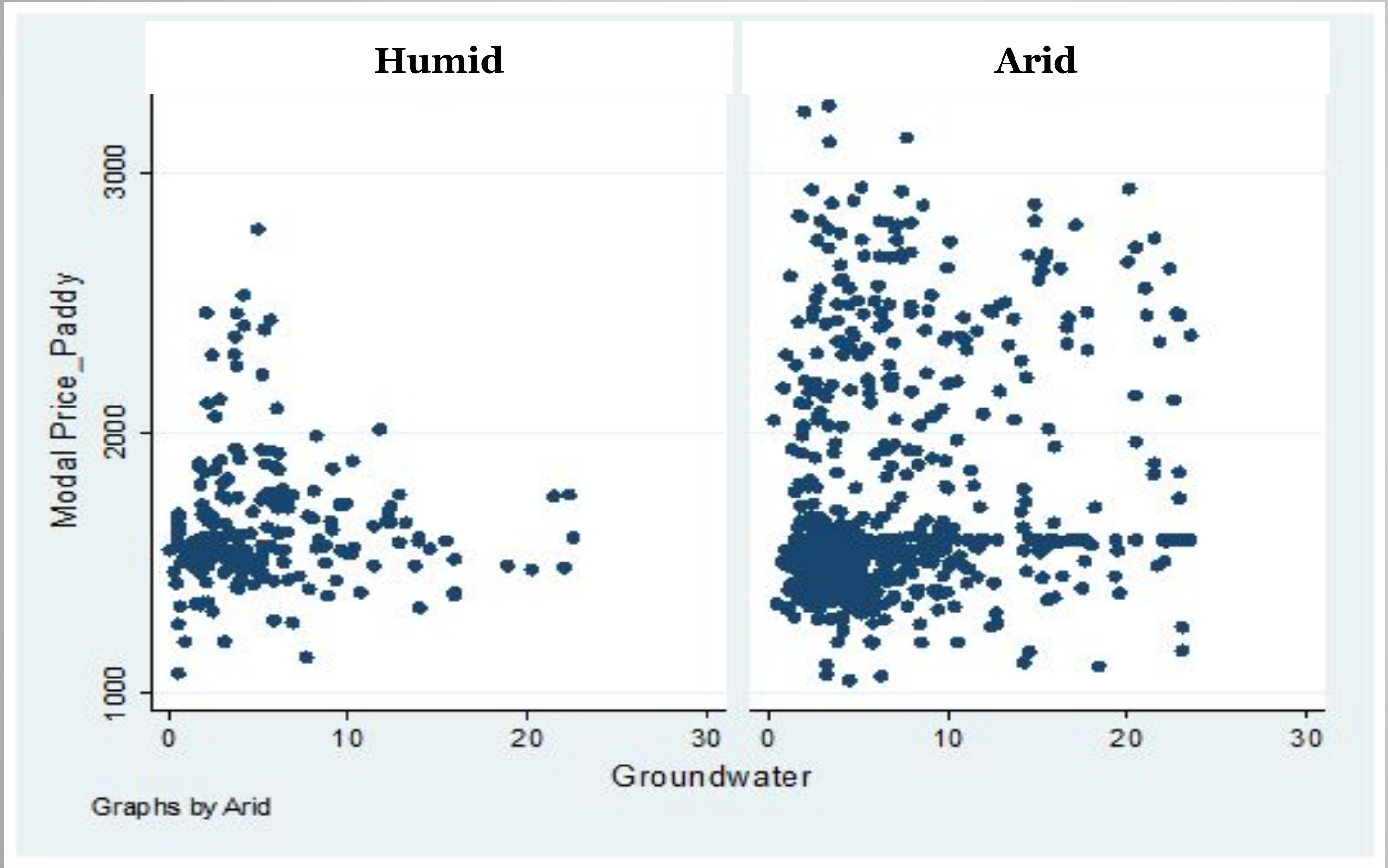
Humid regions have lower depth to groundwater as compared to arid regions.

Correlation between groundwater levels and modal price for Paddy



Positive correlation between GW levels and prices.

Correlation between groundwater levels and modal price for Paddy (Raw Data)



Visibly flatter curve for humid regions.

OLS: Ordinary Least Squared (Assuming Zero Spatial Correlation)

Modal Price = $\beta_0+\beta_1(GW)+u$ ...(1)

Variable (Dependent: Modal Price)	Coefficient (SE)
Intercept	1659.84*** (17.01)
Groundwater	7.95*** (1.36)
N = 1025 | R^2 = 0.032	

***p≤0.01 | **p≤0.05 | *p≤0.10

Areas with higher groundwater stress, have higher prices.

OLS Model (Including Rainfall):

Modal Price = $\beta_0 + \beta_1(GW) + \beta_2(Rainfall) + u$ ...(1)

Variable (Dependent: Modal Price)	Coefficient (SE)
Intercept	1997.15*** (37.06)
Groundwater	3.42* (1.38)
Rainfall	-0.41*** (0.04)
N = 1025 | R^2 = 0.13	

***p≤0.01 | **p≤0.05 | *p≤0.10

Areas with higher groundwater stress, have higher prices but the coefficient is now lower. Further, areas with higher rainfall have lower prices.

Zone-wise OLS Regressions:

Region	Variable (Dependent: Modal Price)	Coefficient (SE)
ARID	**Intercept**	1684.46 *** (19.81)
	Groundwater	7.90*** (1.46)
HUMID	**Intercept**	1630.32*** (34.40)
	Groundwater	-0.03 (4.22)
N = 1025 | R^2= 0.95		

***p≤0.01 | **p≤0.05 | *p≤0.10

Slope and intercepts vary across all the regions (regimes).

Modal Price = $\beta_0 + \beta_1(GW) + u$ ...(2)

In humid areas, the effect of groundwater is not significant because rainfall is in abundance. So, the effect of GW translates to price discovery only in places where it really matters, i.e. areas of GW shortage.

Further, the value of slope coefficient itself shows that all the effect of GW coming from the arid regions.

Zone-wise OLS Regressions (Including Rainfall)

Region	Variable (Dependent: Modal Price)	Coefficient (SE)
ARID	Intercept	2229.04 *** (48.35)
	Groundwater	1.51 (1.45)
	Rainfall	-0.70*** (0.06)
HUMID	Intercept	1742.78*** (62.98)
	Groundwater	-1.89 (4.05)
	Rainfall	-0.12*** (0.06)
N = 1025 | R^2= 0.96		

Modal Price = $\beta_0 + \beta_1(GW) + \beta_2(Rainfall) + u$ ...(2)

Upon inclusion of rainfall, we see similar results. However, these results have not yet accounted for the spatial aspect.

***p≤0.01 | **p≤0.05 | *p≤0.10

Chow test: Comparing model with and without AEZ dummies

Without rainfall:

- The value of the test statistic obtained for the model was 21.33***. This indicates that the model coefficients are not constant across regimes, indicating spatial heterogeneity. And thus, there is need to introduce spatial regimes.

With rainfall:

- Now, test statistic value came out to be 20.80***, which is significant as well but the value is lower as compared to the model without considering the rainfall, indicating that including rainfall is a better option. In either case, considering spatial zones is important.

***$p \leq 0.01$ | **$p \leq 0.05$ | *$p \leq 0.10$

Moran’s I for Regression Residuals (OLS):

Model Specification	Global Moran’s I for regression residuals
Modal ~ GW	0.59*** (35.59)
Spatial regime regression (GW)	0.58*** (35.54)
Modal ~ GW + Rainfall	0.56*** (34.37)
Spatial regime regression (GW + Rainfall)	0.54*** (33.04)

- The results for all OLS regressions show high Moran’s I values which are highly significant. This rejects the null hypothesis of non-spatially correlated error terms. Thus, there is a need to use regression model that accounts for this.

Spatial Lag Model: Global Correlation

- MLE of the spatial simultaneous autoregressive lag model of the form:

$$Y = \rho WY + X\beta + \varepsilon$$

Variable (Dependent: Modal Price)	Coefficient (SE)
Intercept	537.80*** (53.23)
GW	0.20 (0.96)
Rainfall	-0.11*** (0.03)
N: 1025 | Rho: 0.74 Log likelihood: -7463.84 LM test for residual autocorrelation test value: 23.78***	

On accounting for the spatial correlation in Y, we see that the impact of GW on prices is now insignificant. This indicates that the impact of GW on prices is completely translated by the neighbouring prices at a point. To explore this further, we run the spatial durbin model.

***p≤0.01 | **p≤0.05 | *p≤0.10

Spatial Lag Model: Zone-wise

Zone	Variable (Dependent: Modal Price)	Coefficient (SE)
ARID	**Intercept**	627.95*** (62.78)
	GW	0.05 (1.05)
	Rainfall	-0.21*** (0.04)
HUMID	**Intercept**	520.70*** (62.73)
	GW	-3.37 (2.91)
	Rainfall	-0.05 (0.04)
N: 1025 | Rho: 0.72 Log likelihood: -7458.6 LM test for residual autocorrelation test value: 20.22***		

***p≤0.01 | **p≤0.05 | *p≤0.10

Spatial Durbin Model: Local and Global Correlation

Variable (Dependent: Modal Price)	Coefficient (SE)
Intercept	529.94*** (57.67)
GW	-2.72** (1.09)
Rainfall	0.08 (0.12)
Lagged GW	2.81** (1.08)
Lagged Rainfall	-0.19 (0.12)
N: 1025 | Rho: 0.73 Log likelihood: -7461.12 LM test for residual autocorrelation test value: 22.56***	

MLE of the spatial durbin model of the form:

$$\mathbf{Y = \rho WY + X\beta + WX\theta + \varepsilon}$$

***p≤0.01 | **p≤0.05 | *p≤0.10

Spatial Durbin Model Inferences_Part 1:

- Higher groundwater stress in neighbouring regions drives the prices in a region upwards. (High-High regions of GW stress coincided with High-High regions of prices).

- Interestingly, for a region, its own groundwater shows an inverse relation with prices and rainfall levels show a positive relation with prices. But, for the neighbouring regions, both these signs are flipped.

- So, in a sense, the own and neighbouring effects of groundwater as an input on price are non-complementary in nature.

Spatial Durbin Model: Zone-wise

Zone	Variable (Dependent: Modal Price)	Coefficient (SE)
HUMID	**Intercept**	560.19*** (144.34)
	GW	-0.23 (1.16)
	Rainfall	-0.01 (0.13)
	Lagged GW	0.14 (2.05)
	Lagged Rainfall	-0.21 (0.15)
ARID	**Intercept**	467.46*** (74.32)
	GW	-5.34* (3.09)
	Rainfall	0.18 (0.14)
	Lagged GW	9.62* (5.51)
	Lagged Rainfall	-0.21 (0.15)
N: 1025 | Rho: 0.72 Log likelihood: -7455.16 LM test for residual autocorrelation test value: 20.25***		

***p≤0.01 | **p≤0.05 | *p≤0.10

Spatial Durbin Model Inferences_Part 2:

- For arid regions, groundwater in both own and neighbouring locations has a significant role to play in price determination. While higher GW stress in a region leads to lower prices in that region, in general being surrounded by highly stressed regions, drives the price upwards. (overall shortage in supply, could drive the price upwards.)

- For humid regions, both the types of water inputs do not play a significant role in market price determination. (Only scarcity/ over-abundance of water has an impact on price.)

- Also, LM test value becomes less significant as compared to the spatial lag model, thus justifying the model preference of Spatial Durbin model.

ML Estimation of the Spatial Error Model

- MLE of the spatial simultaneous autoregressive error model of the form:

$$\mathbf{Y = X\beta + u, u = \lambda Wu + \varepsilon}$$

Variable (Dependent: Modal Price)	Coefficient (SE)
Intercept	1900.62*** (67.80)
GW	15.48*** (1.10)
Rainfall	-0.23*** (0.08)
N: 1025 | Lambda: 0.74 Log likelihood: -7468.63	

***p≤0.01 | **p≤0.05 | *p≤0.10

Spatial Error Model: Zone-wise

Zone	Variable (Dependent: Modal Price)	Coefficient (SE)
ARID	Intercept	1999.01*** (78.78)
	GW	10.14*** (1.16)
	Rainfall	-0.38*** (0.10)
HUMID	Intercept	1813.77*** (95.56)
	GW	5.58* (3.08)
	Rainfall	-0.1 (0.10)
N: 1025 | Lambda: 0.73 Log likelihood: -7464.535		

***$p \leq 0.01$ | **$p \leq 0.05$ | *$p \leq 0.10$

Lagrange Multiplier Test Statistics (Spatial regression by AEZ)

Model	GW	GW + Rainfall
Spatial Error Model	1248.8***	1156.30***
Robust Spatial Error Model	3.05*	0.30
Spatial Lag Model	1316.3***	1180.70***
Robust Spatial Lag Model	70.50***	24.73***

- Both LMerr and LMlag are highly significant, while RLMlag is still strongly significant, there is much weaker evidence given by RLMerr. Hence, it suggests lag model as better alternative.

References:

- Chatterjee, S. & Kapur, D., 2016. *Understanding Price Variation in Agricultural Commodities in India: MSP, Government Procurement, and Agriculture Markets,* New Delhi: National Council of Applied Economic Research.
- Government of India, 2016. *State of Indian Agriculture: 2015-16,* New Delhi: Ministry of Agriculture & Farmers Welfare, Department of Agriculture, Cooperation & Farmers Welfare, Government of India.
- Seckler, D. et al., 1998. *World Water Demand and Supply, 1990 to 2025: Scenarios and Issues,* Colombo: Colombo: International Water Management Institue.
- Sekhri, S., 2014. Wells, water, and welfare: the impact of access to groundwater on rural poverty and conflict. *American Economic Journal: Applied Economics,* 6(3), pp. 76-102.

- Ramankutty, N., T. Hertel, H.-L. Lee, and S. K. Rose, Global Agricultural Land Use Data for Integrated Assessment Modeling, in Human-Induced Climate Change: An Interdisciplinary Assessment. Accessed here: https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3184

Thank you.