Mining Industry in China: A closer look at the Rare Earth Mining Industry and its Environmental Impact

Nils Lang

October 16, 2025

Note: This is a work-in-progress draft, please do not cite or distribute without permission.

1 Introduction

Rare Earths (RE) or Rare Earth Elements (REE) are essential in the high tech industry, renewable energy and the defense industry. This group of 17 chemical elements exhibit unique properties as they are magnetic, phosphorescent (light-emitting), and have catalytic properties. These attributes allow them to perform functions that are difficult to achieve with other materials, especially at the same weight. In High-Tech Devices (smartphones, computers, camera lenses etc.) REEs enable smaller sizes, increased efficiency, and better performance of these devices. In the field of green-tech they are the basis used to create strong, lightweight permanent magnets found in electric vehicle (EV) motors and wind turbine generators. These essential metals have been exported by China since 1976 and China has thereafter successively become the biggest miner, refiner, exported and user of REE worldwide. The main reasons for this is not only the fact that China has the largest - and more importantly easily accessible - deposits of REE, but also that the Chinese government has been very lenient with environmental regulations, which has allowed the industry to grow rapidly. This has led to a situation where China is now the dominant player in the global REE market, controlling more than 90% of the world's production and refining capacity. [3]

2 Rare Earth Supply Chain: An Overview

2.1 Definition

Upstream or primary products are RE ores and concentrates. Midstream products are defined as separated Rare Earth oxides, metals, and alloys. The downstream sector can generally be understood as the sum of users of midstream products for the manufacturing of Rare Earth-containing products. For this essay I will limit the scope to only include the mining and processing e.g. Upstream and Midstream.

2.2 Geology of Rare Earths

Rare Earths are not truly "rare" in geological abundance, but they are rarely found in economically viable concentrations. While they exist in the crust of the earth all around the world, finding deposits that are concentrated enough to be mined profitably is challenging. Ore deposits tightly bonded together and in varying proportions and mixes of REEs. Economic viability has also been shifting over recent decades, since the market is under the influence of highly volatile prices, which shift based on geopolitical and economic developments.

2.3 Upstream Extraction

The first step in the supply chain is the extraction of raw unprocessed REEs from the earth. They are extracted from the earth using acid in an environmentally damaging extraction process. This lead to many western nations not wanting to this anymore in the 1980s and China welcoming in the mining operations. Today more then 90% of yearly global output of refined metals used for rare earth magnets

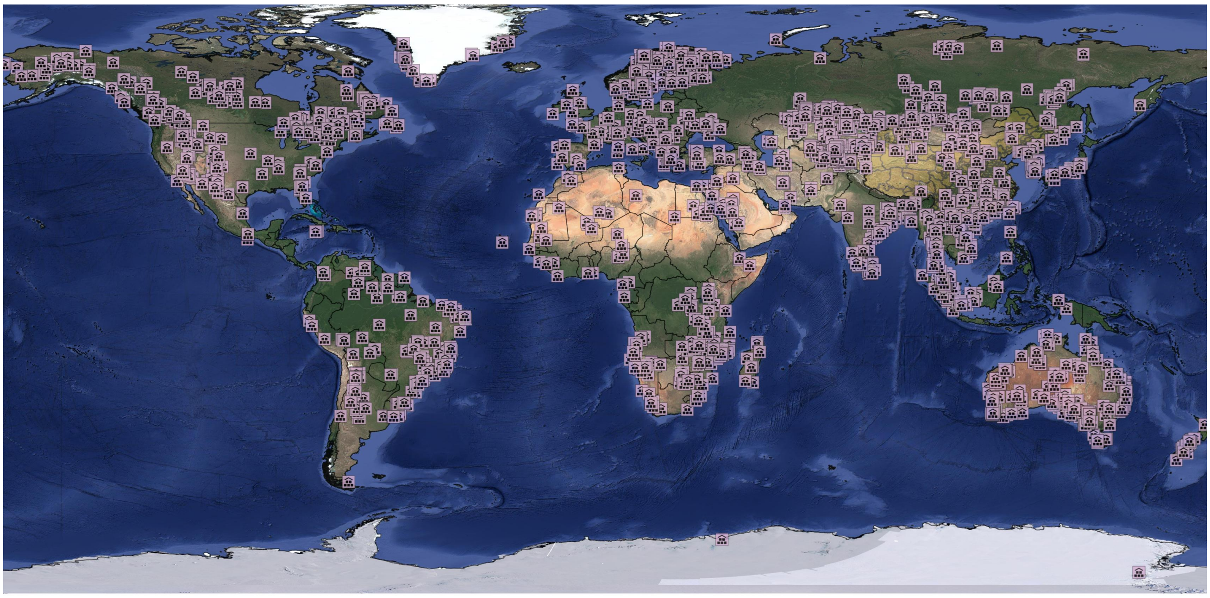

Figure 1: Known Deposits of Rare Earths worldwide [8]

come from China. [3] Deposits in the North of China have a relatively large proportion of the lower-valued lanthanum and cerium elements. These two products are frequently stockpiled at the mines there, because prices are very low. In the South the concentration of heavy rare earth elements, used in permanent magnets, is higher, while the North has more light rare earth element. Historically illegal mining has been a big problem in China, especially in the South, where the concentration of the more expansive heavy rare earth elements is higher.

2.4 Midstream Processing and Separation

Separation of the different REEs is a highly energy intensive process and require a lot of know-how, since the REEs are quite similar in their chemical structure. [5] While there are different fundamental approaches on how to arrive at $> 99\%$ pure Rare Earth Oxides, they all generate a lot of waste water and need to cheap energy to be profitable. After separation the resulting Oxides and Salts are then converted into metals or alloys. High-Purity metals are commonly produced by Metallothermic Reduction. For most industry applications in the downstream an alloy, a mixture of REEs and other metals (aluminum, steel, iron etc.), is needed.

3 Data

The United States Geological Survey (USGS) publishes two datasets related to the Rare Earth Industry in China. One is focused on the entire mineral industry in China and was published in 2023. Besides data on the location of a multitude of different types of mines, it also includes data for processing plant as well as major ports. This is also the data used for Fig.2 and Fig.3 [7] The other relevant dataset has a worldwide focus but only looks at Rare Earth occurrences as well as past and current producers. This was published in 2018, but is more detailed, nonetheless I will use both for different analysis of the water pollution with conductive materials. This is also the data used to visualize the occurrences in Fig.1 [8] Luan et al. recently published a dataset on a multitude of different metrics on water quality and made it publicly available. Using machine learning they backtracked some of the missing data from measurements downstream. The main focus for this essay is gonna lie with the electric conductivity. While the original data set has a **monthly granularity** to save my computer and not overload my RAM I had to work with yearly averages. The data set includes both data on lakes as well as reservoirs, but I will only use reservoirs data, since lakes tend to have water flowing in and out of them and therefore could lead to bad predictions about the pollution with conductivity. [4] In order to be able incorporate fixed-effects the administrative data by USGS was joined with the corresponding dataframes using the geopandas sjoin-function. Taiwan will be filtered out of the data, since there is no data on mines and

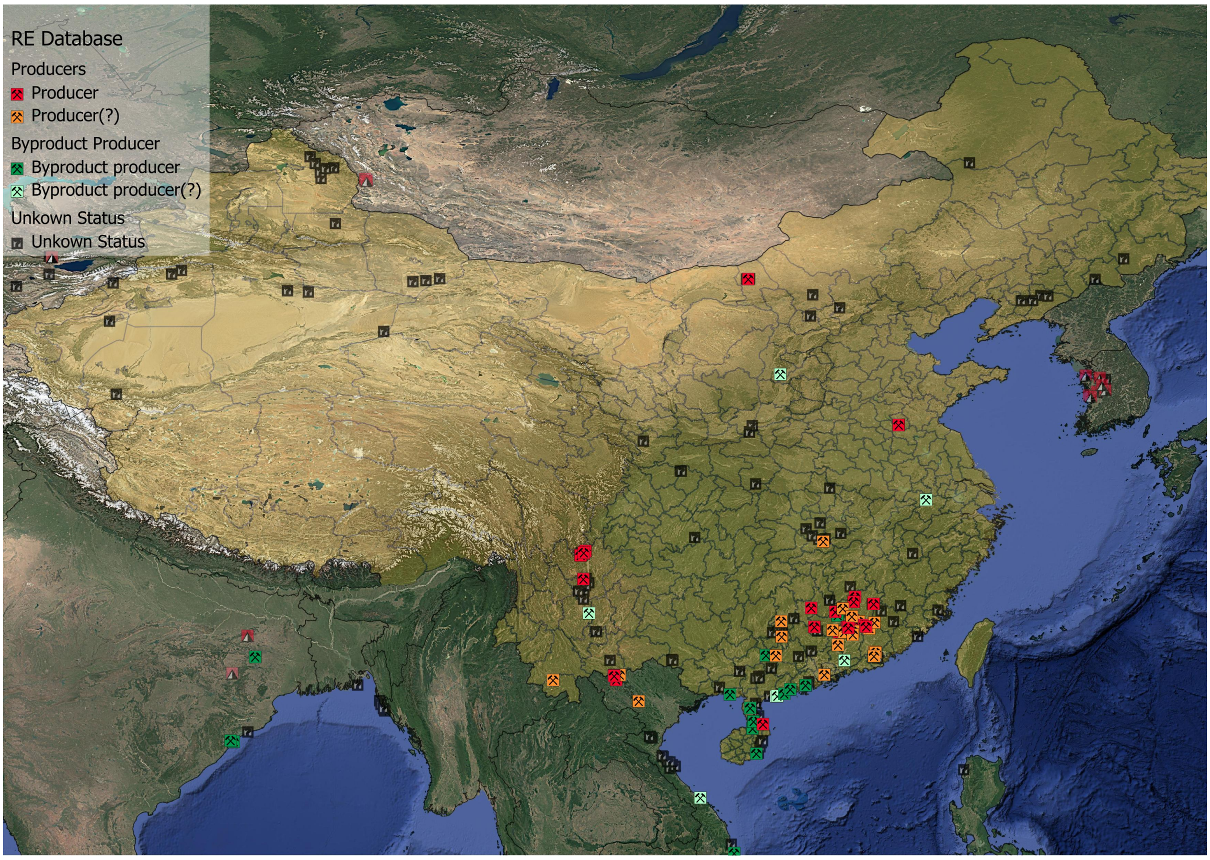

Figure 2: Active REE producing and as byproduct producing mines in China [8]

processing there and the ocean distance between mainland China and Taiwan is too large to be relevant for the analysis. For a quick overview of the conductivity data, I will include a table with the descriptive statistics of the conductivity variable (taken from Attempt 2).

Statistic	Cond
mean	3.429776×10^{-2}
std	2.071525×10^{-2}
min	8.996097×10^{-3}
25%	1.949954×10^{-2}
50%	2.671436×10^{-2}
75%	4.306361×10^{-2}
max	1.476583×10^{-1}

4 Empirical Strategy

In order to stay transparent about the process of how I arrived at my results, I wil go into detail about the three different approaches I took and hope to be able to showcase my understanding of Geo-Data Processing in the process as well. The result from the first one is counterintuitive and I believe to be the least relevant, but I will still include it for transparency. The second one is the most relevant and I will go into detail about the results. The third one is a more exploratory approach, which I will also include for transparency.

4.1 Attempt 1 - OLS with Buffers

My initial approach was to create a buffer around the mines and then run an OLS regression on the conductivity of the reservoirs within that buffer. The idea was to see if there is a significant difference in conductivity between reservoirs that are within the buffer of a mine and those that are not. I used QGIS to create a buffer around the mines with a degree parameter of 0.75, which is approximately 83 kilometers

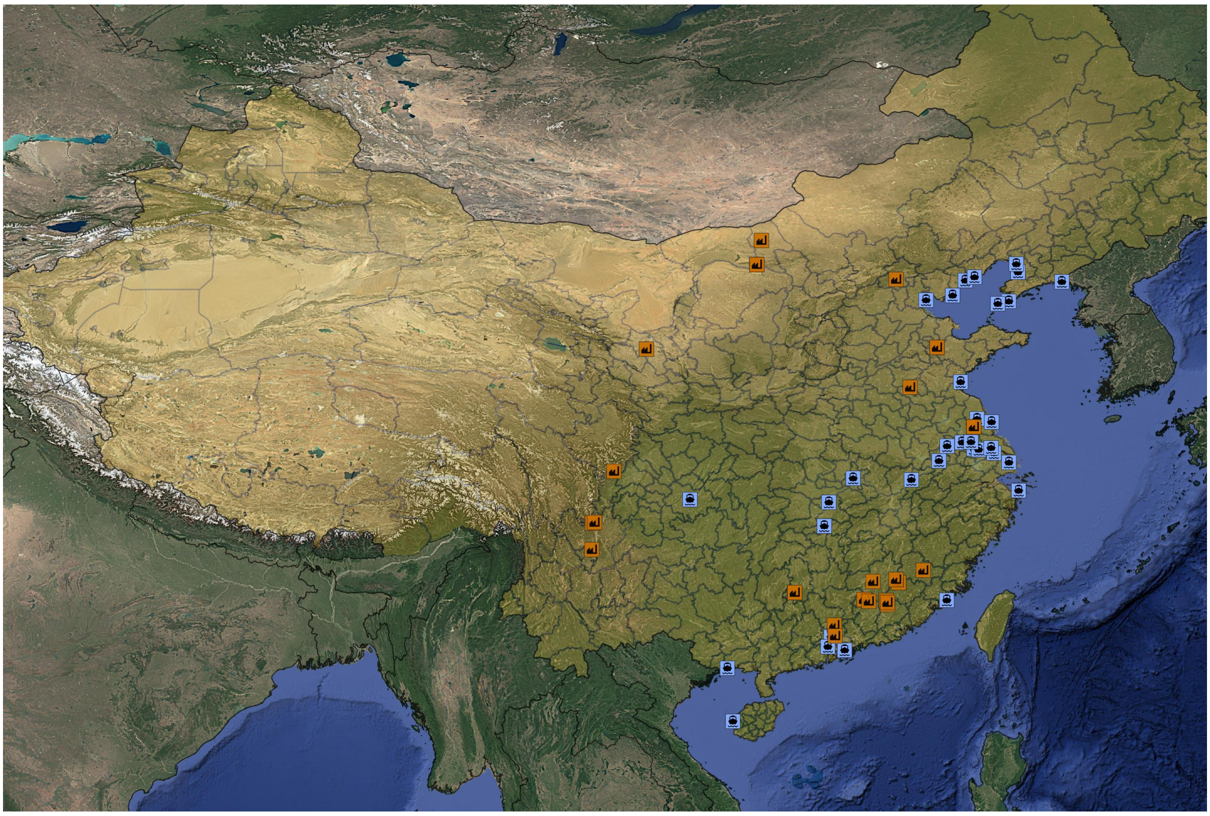

Figure 3: Active midstream RE Processing locations and major ports exporting metals in China [6]

at the equator, but varies significantly depending on the latitude. The buffer was then used to create a binary variable indicating whether a lake or reservoir is within the buffer of an active mine producer. Additionally, I included fixed effects for each administrative province in China to control for regional differences in conductivity. Fig.4 shows the mines with the buffer around them as they were defined in QGIS. The red points are active producers, while the green points are active byproduct producers. Active producers are defined as mines that are currently producing REEs, while byproduct producers are mines that produce REEs as a byproduct of other mining activities. The buffer was created using the EPSG:4326 coordinate reference system (CRS), which uses degrees as units. In the OLS the two types of mines were combined and resulted in the buffers seen in Fig.5 being used for fitting the model.

OLS Formula

The dependent variable in this first model is Cond (Conductivity), and the independent variables include a binary indicator is_inside_producer, which is a buffer done with QGIS and degree parameter of 0.75, and fixed effects for each administrative province admin_province. The regression model estimated is of the form:

$$\text{Cond}_i = \beta_0 + \beta_1 \cdot \text{is\_inside\_producer}_i + \sum_{j=1}^{N-1} \gamma_j \cdot \mathbb{I}(\text{admin\_province}_i = \text{Province}_j) + \epsilon_i$$

Where:

- Cond_i: The conductivity for observation i.
- $\text{is\_inside\_producer}_i$: A binary variable, which is 1 if observation i is within a (main and active) producer's geometry, and 0 otherwise.
- β_0: The intercept.
- β_1: The coefficient for is_inside_producer.
- γ_j: The coefficient for the j-th administrative province's fixed effect.

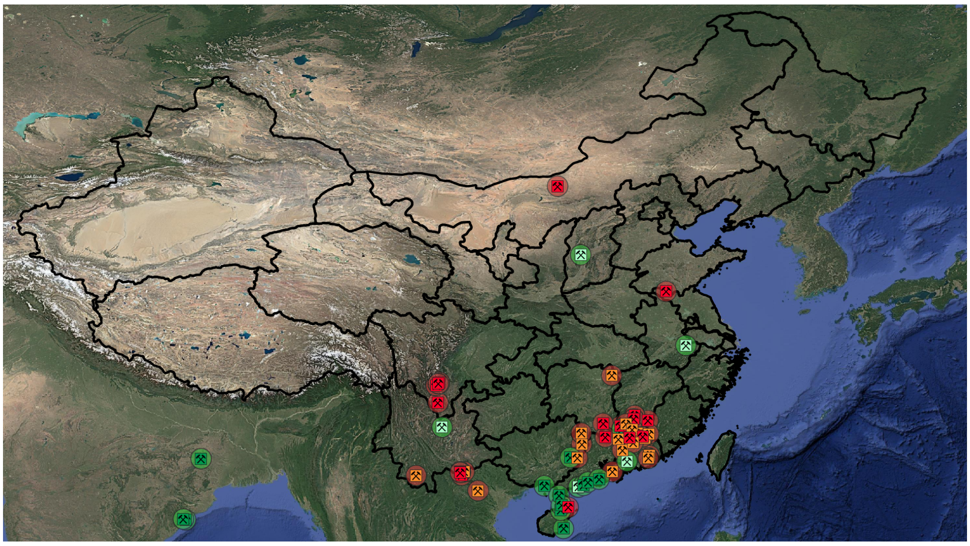

Figure 4: Mines with Buffer around them. Red - active producers. Green - active Byproduct Producers.

- $\mathbb{I}(\cdot)$: The indicator function.
- ϵ_i: The error term for observation i.
- N: The total number of unique administrative provinces in the dataset.

OLS Attempt 1 Results

Table 1: Summary of OLS Regression Results

Metric	**Value**
Dependent Variable	Cond
R-squared	0.657
Adj. R-squared	0.657
F-statistic	1.918e+05
Prob (F-statistic)	0.000
No. Observations	3,402,408

Table 2: OLS Regression Coefficients

Variable	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0338	2.75e-05	1231.474	0.000	0.034	0.034
is_inside_producer[T.True]	-0.0015	1.61e-05	-92.709	0.000	-0.002	-0.001

The OLS regression results show that the model explains approximately **65.7%** of the variance in Cond, as indicated by the **R-squared** value of **0.657**. The high **F-statistic** of **1.918e+05** and a **p-value** of **0.00** suggest that the model as a whole is highly statistically significant.

- The coefficient for is_inside_producer is **-0.0015** ($p < 0.001$), indicating a very small but statistically significant negative association between being inside a producer buffer and conductivity.
- Most administrative province fixed effects are highly statistically significant, with both positive and negative coefficients, reflecting substantial regional variation in conductivity.
- The model explains a substantial portion of the variance in conductivity ($R^2 = 0.657$), but the practical effect of proximity to mines, as measured by the buffer, appears minimal.

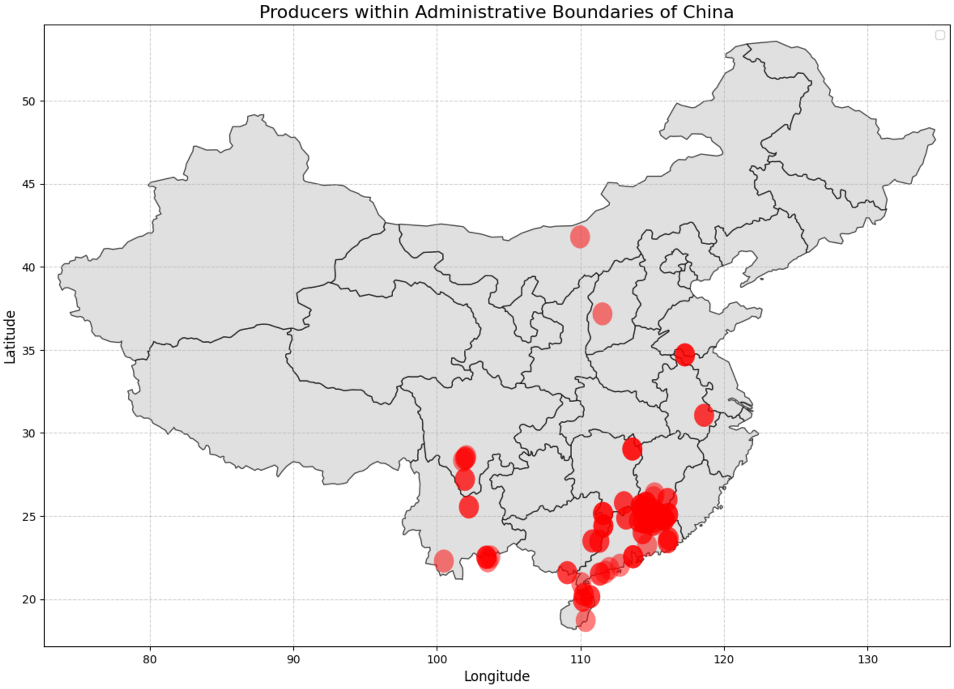

Figure 5: Mines included in the Buffer OLS Attempt 1

- There is also one administrative province, which is listed as T.Unkown, which is not a real province, but rather a placeholder for the data that could not be matched to any of the administrative provinces. This is due to the mines in the South of China, where the buffer goes into the country, but there is no association with any administrative province.

Figure 6 shows the fixed effects for the OLS regression attempt 1. The figure illustrates the coefficients for each administrative province, with Anhui Province being the reference Province. All the Fixed Effects have high significance and a **p-value** of **0.00**, but since the table is super large I hope the figure is enough. If there is a need for the specific number, please refer to Table 2 in the appendix.

Known Limitations

Since the CRS for was EPSG:4326 the buffers were done in degrees, the physical distance represented by one degree of longitude changes significantly depending on the latitude. Near the equator, one degree of longitude is approximately 111 kilometers, but it shrinks to almost zero at the poles. One degree of latitude is roughly constant at about 111 kilometers. This distortion means a "buffer of 1 degree" will be a different physical size at different locations.[2] [1] The buffers in QGIS were created using the EPSG:4326 coordinate reference system (CRS), which uses degrees as units. This means that the buffer size is not constant across different latitudes, which could lead to some issues in the analysis.

4.2 Attempt 2 - OLS with Distance to Closest Mine

OLS Formula

The dependent variable in this model is once again Cond (Conductivity), and the independent variables is the distance to the nearest mines (producer) as distance_to_nearest_producer, which is measured in **CRS-32649** and meters, as well as fixed effects for each administrative province admin_province. The mines are the same ones as used in the first attempt (including the mines in Vietnam), just the centroid this time. The regression model estimated is of the form:

$$\text{Cond}_i = \beta_0 + \beta_1 \cdot \text{distance\_to\_nearest\_producer}_i + \sum_{j=1}^{N-1} \gamma_j \cdot \mathbb{I}(\text{admin\_province}_i = \text{Province}_j) + \epsilon_i$$

Where:

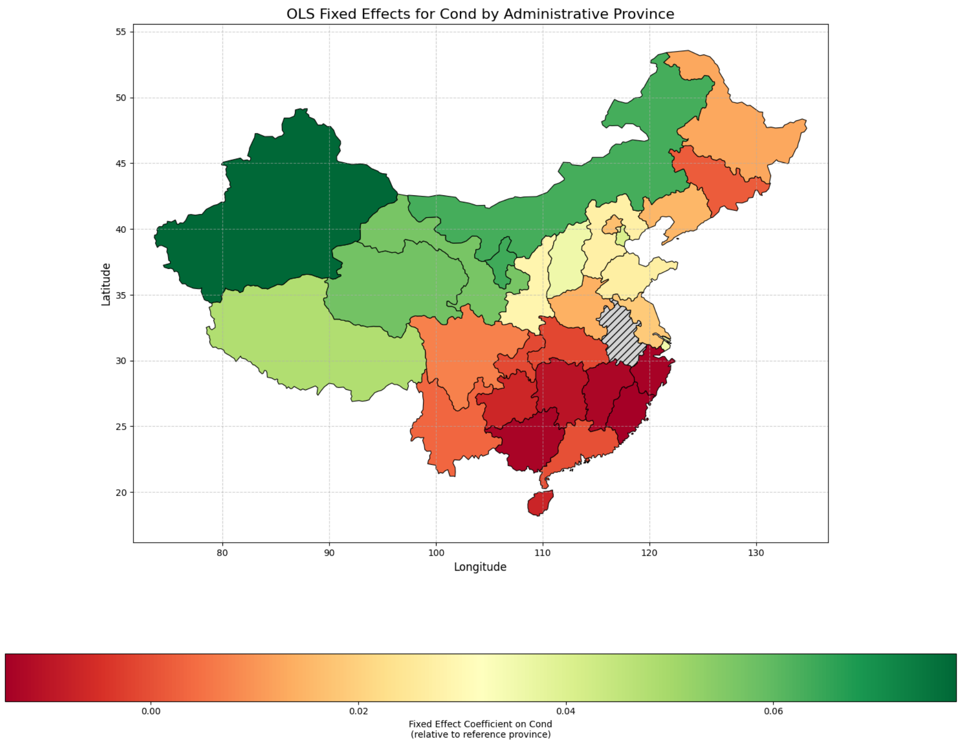

Figure 6: Fixed Effects for OLS Attempt 1

- Cond_i: The conductivity for observation i.
- $\text{distance\_to\_nearest\_producer}_i$: The distance from observation i to the nearest producer.
- β_0: The intercept.
- β_1: The coefficient for distance_to_nearest_producer.
- γ_j: The coefficient for the j-th administrative province's fixed effect.
- $\mathbb{I}(\cdot)$: The indicator function, which takes a value of 1 if the condition inside the parenthesis is true, and 0 otherwise.
- ϵ_i: The error term for observation i.
- N: The total number of unique administrative provinces in the dataset.

OLS Regression Results Summary

Table 3: Summary of OLS Regression Results

Metric	**Value**
Dependent Variable	Cond
R-squared	0.659
Adj. R-squared	0.659
F-statistic	1.326×10^5
Prob (F-statistic)	0.000
No. Observations	2,337,936

The OLS regression results show that the model explains approximately **65.9%** of the variance in Cond, as indicated by the **R-squared** value of **0.659**. The high **F-statistic** of **1.326e+05** and a **p-value** of **0.00** suggest that the model as a whole is highly statistically significant.

Table 4: OLS Regression Coefficients

Variable	**Coef.**	**Std. Err.**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0355	3.28×10^{-5}	1083.538	0.000	0.035	0.036
distance_to_nearest_producer	-8.624×10^{-9}	6.21×10^{-11}	-138.917	0.000	-8.75×10^{-9}	-8.50×10^{-9}

- The coefficient for distance_to_nearest_producer is **-8.624e-09** ($p < 0.001$), indicating a very small but statistically significant negative association between distance to the nearest producer and conductivity. This is contrary to attempt 1, but this time it is what we would expect. **Closer to the mine means more conductivity**
- Most administrative province fixed effects are highly statistically significant, with both positive and negative coefficients, reflecting substantial regional variation in conductivity.
- The model explains a substantial portion of the variance in conductivity ($R^2 = 0.659$), but the practical effect of distance to mines appears once again minimal.

Figure 7 shows once again the fixed effects for the OLS regression for Attempt 2. The figure illustrates the coefficients for each administrative province, with Anhui Province being the reference Province. All the Fixed Effects have high significance and a **p-value** of **0.00**, but since the table is super large I hope the figure is enough. If there is a need for the specific number, please refer to the full result in the Appendix.

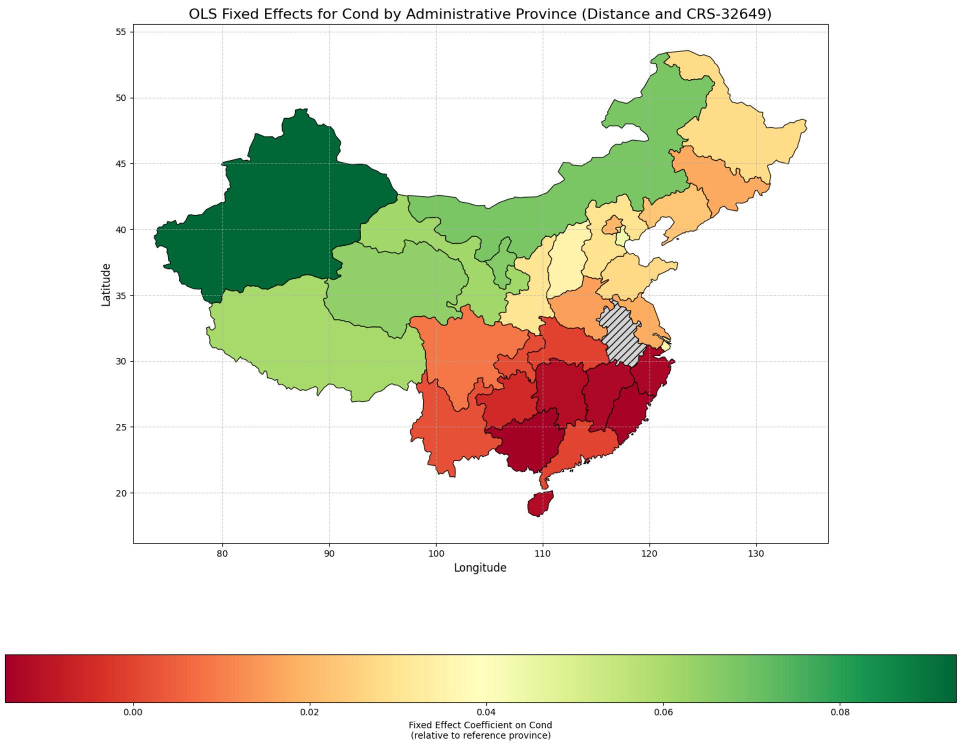

Figure 7: Fixed Effects for OLS Attempt 2

Known Limitations

The distance is measured in meters, which means that the coefficient is very small. **For 100 kilometers, the value would be a reduction of** -0.0008624 **in conductivity.** The distance to the nearest mine is also not a very good measure of the impact of mining on water quality, since it does not take into account the size of the mine or the amount of mining activity.

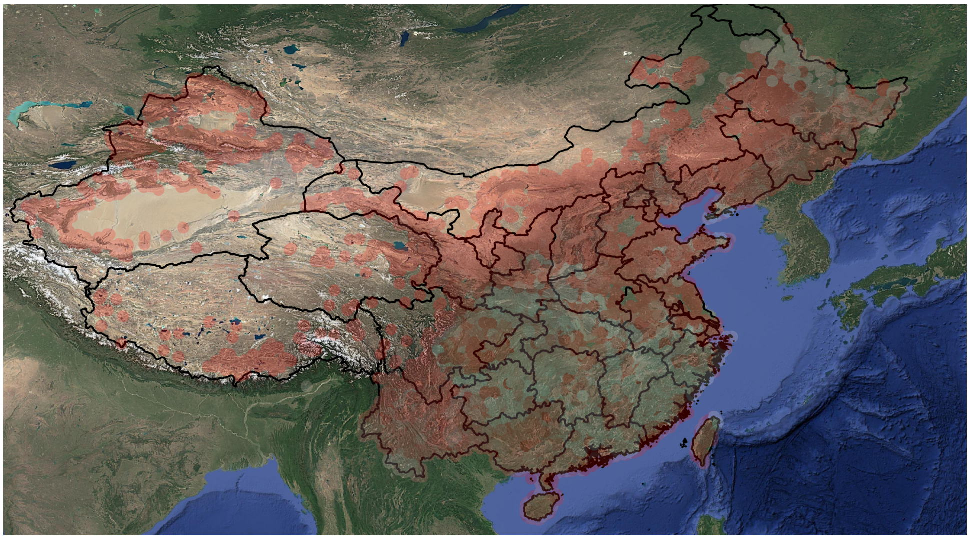

Figure 8: Reservoir Measurements in 2001

4.3 Attempt 3 - Fixed Effects with Distance to Closest Mine on Reservoirs

I will now run the OLS **on the whole data of the mining industry in China**, which can be seen in Fig. 9. Once again we will measure to closest distance to either a mine or a processing location in**CRS-32649**. The dependent variable in this model is once again Cond (Conductivity), and the independent variables is the distance to the nearest mining industry as distance_to_nearest_mining_industry_in_meters, which is measured in and meters, as well as fixed effects for each administrative province admin_province.

$$\text{Cond}_i = \beta_0 + \beta_1 \cdot \text{distance\_to\_nearest\_mining industry}_i + \sum_{j=1}^{N-1} \gamma_j \cdot \mathbb{I}(\text{admin\_province}_i = \text{Province}_j) + \epsilon_i$$

Where:

- Cond_i: The conductivity for observation i.
- $\text{distance\_to\_nearest\_mining industry}_i$: The distance from observation i to the nearest producer or processer.
- β_0: The intercept.
- β_1: The coefficient for distance_to_nearest_mining industry.
- γ_j: The coefficient for the j-th administrative province's fixed effect.
- $\mathbb{I}(\cdot)$: The indicator function, which takes a value of 1 if the condition inside the parenthesis is true, and 0 otherwise.
- ϵ_i: The error term for observation i.
- N: The total number of unique administrative provinces in the dataset.

The OLS regression results show that the model explains approximately **65.8%** of the variance in Cond, as indicated by the **R-squared** value of **0.658**. The high **F-statistic** of **1.330e+05** and a **p-value** of **0.00** suggest that the model as a whole is highly statistically significant.

- The coefficient for distance_to_nearest_producer is **-1.084e-09** ($p < 0.003$), indicating a very small but statistically significant negative association between distance to the nearest producer and conductivity. This is contrary to attempt 1, but this time it is what we would expect. **Closer to the mine means more conductivity**
- Most administrative province fixed effects are highly statistically significant, with both positive and negative coefficients, reflecting substantial regional variation in conductivity.

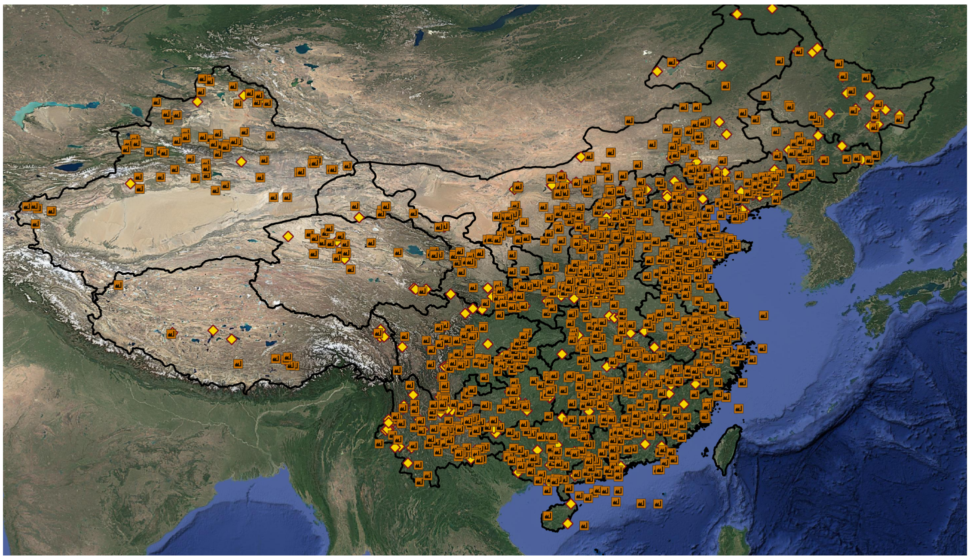

Figure 9: Processing and Mining Locations in China. Orange - Processing Locations. Yellow - Mining Locations.

Table 5: Summary of OLS Regression Results

Metric	**Value**
Dependent Variable	Cond
R-squared	0.658
Adj. R-squared	0.658
F-statistic	1.330×10^5
Prob (F-statistic)	0.000
No. Observations	2,352,768
Df Residuals	2,352,733
Df Model	34
Log-Likelihood	7.0353×10^6
AIC	-1.407×10^7
BIC	-1.407×10^7
Covariance Type	nonrobust

Table 6: OLS Regression Coefficients

Variable	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0343	3.34×10^{-5}	1027.130	0.000	0.034	0.034
distance_to_nearest_producer	-1.084×10^{-9}	3.53×10^{-10}	-3.076	0.002	-1.78×10^{-9}	-3.93×10^{-10}

- The model explains a substantial portion of the variance in conductivity ($R^2 = 0.658$), but the practical effect of distance to mines appears once again minimal.

Figure 10 shows once again the fixed effects for the OLS regression for Attempt 3. The figure illustrates the coefficients for each administrative province, with Anhui Province being the reference Province. All the Fixed Effects have high significance and a **p-value** of **0.00**. If there is a need for the specific number, please refer to the full result in the Appendix.

5 Reflections

The conductivity is not the only variable that is interesting, but it is the one that is most closely related to the mining industry. The other variables could also be used to look at the impact of mining on water quality, but they are not as directly related to the mining industry as conductivity. For example the pH

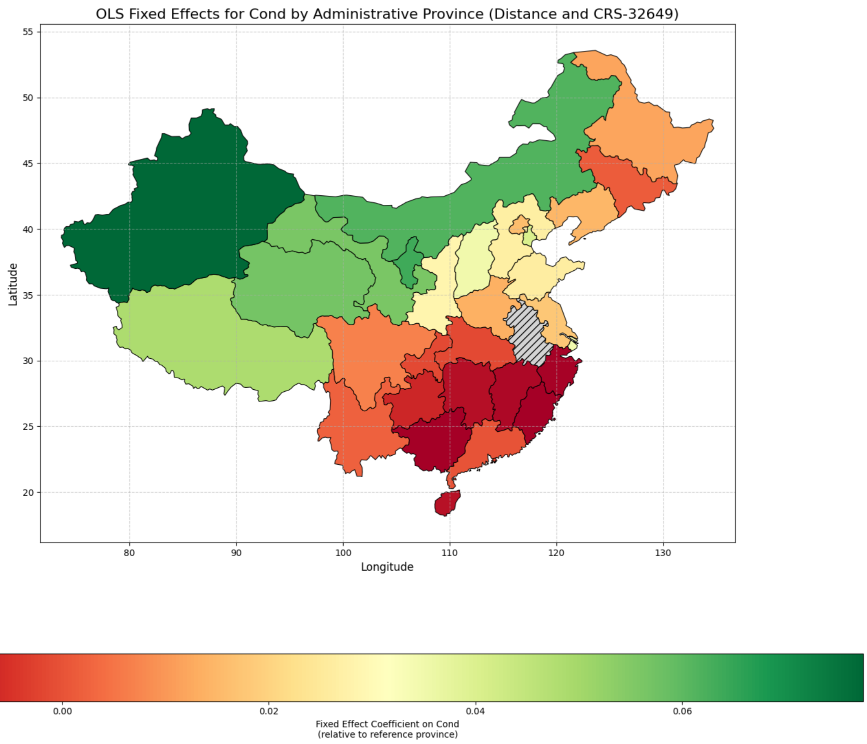

Figure 10: Fixed Effects for OLS Attempt 3

level is also heavily influenced by the agricultural industry, which is not the focus of this essay.
The results also show that there are significant differences in conductivity between different administrative provinces, which could be due to different mining practices. The type of rare earths being mined or other factors. Interestingly effect of decreasing conductivity with being further away from mines is seemingly stronger with the only rare earth mines data. This makes somewhat sense, since there are a lot of reports of illegal mining in the South of China, which is not explicitly modeled in the data, which the environmental agencies ignored until 2011. The makeup of the earth could also be interesting to look at, but fixed effects for the administrative provinces should control for that to some extent. Population and industrialization levels could furthermore also be interesting to look at, but they are not included in the data set.

6 Conclusion

In this essay I have shown that the mining industry in China has a significant impact on the water quality of reservoirs. The conductivity of the water is significantly higher in reservoirs that are closer to mines. The results of the OLS regression show that the distance to the nearest mine is a significant predictor of conductivity, even when controlling for fixed effects for each administrative province. The practical effect of proximity to mines appears minimal for $-8.624e-09$ per meter increase in distance to the nearest producer. I think this is a good first step in understanding the impact of mining on water quality, but more research is needed to understand the full extent of the impact. Future research could look at other variables, such as pH level, population levels, to get a more complete picture of the impact of mining on water quality. Flow simulations could also be used to look at the impact of mining on water quality in a more dynamic way. NOTE: If the python code is wanted, I can of course send it per mail.

References

[1] *How Far Is a Degree Longitude and Latitude.* https://www.johndcook.com/how_big_is_a_degree.html. (Visited on 07/29/2025).

[2] *How Much Distance Does a Degree, Minute, and Second Cover on Your Maps? — U.S. Geological Survey.* https://www.usgs.gov/faqs/how-much-distance-does-a-degree-minute-and-second-cover-your-maps. June 2017. (Visited on 07/29/2025).

[3] IEA. "Global Critical Minerals Outlook 2025". In: (2025).

[4] Shilong Luan et al. *High Resolution Water Quality Dataset of Chinese Lakes and Reservoirs from 2000 to 2023.* 2025. DOI: 10.6084/M9.FIGSHARE.27626286.V2. (Visited on 07/14/2025).

[5] Mahmut MAT. *Bastnäsite : Properties, Uses, Occurrence ¿¿ Geology Science.* Sept. 2024. (Visited on 07/14/2025).

[6] Elizabeth R Neustaedter et al. *Compilation of Geospatial Data (GIS) for the Mineral Industries and Related Infrastructure of the People's Republic of China.* 2023. DOI: 10.5066/P9HK2K8I. (Visited on 07/17/2025).

[7] Elizabeth R Neustaedter et al. *Compilation of Geospatial Data (GIS) for the Mineral Industries and Related Infrastructure of the People's Republic of China.* 2023. DOI: 10.5066/P9HK2K8I. (Visited on 07/29/2025).

[8] Greta J. Orris et al. *Global Rare Earth Element Occurrence Database.* 2018. DOI: 10.5066/F7DR2TN4. (Visited on 05/26/2025).

7 Appendix

Attempt 1 - OLS with Buffers

Dep. Variable:	Cond	**R-squared:**	0.657
Model:	OLS	**Adj. R-squared:**	0.657
Method:	Least Squares	**F-statistic:**	1.918e+05
Date:	Wed, 30 Jul 2025	**Prob (F-statistic):**	0.00
Time:	22:27:28	**Log-Likelihood:**	1.0466e+07
No. Observations:	3402408	**AIC:**	-2.093e+07
Df Residuals:	3402373	**BIC:**	-2.093e+07
Df Model:	34		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0338	2.75e-05	1231.474	0.000	0.034	0.034
is_inside_producer[T.True]	-0.0015	1.61e-05	-92.709	0.000	-0.002	-0.001
C(admin_province)[T.Beijing Municipality]	0.0165	0.000	77.874	0.000	0.016	0.017
C(admin_province)[T.Chongqing Municipality]	-0.0011	5.26e-05	-20.348	0.000	-0.001	-0.001
C(admin_province)[T.Fujian Province]	-0.0141	4.43e-05	-317.715	0.000	-0.014	-0.014
C(admin_province)[T.Gansu Province]	0.0567	9.17e-05	618.843	0.000	0.057	0.057
C(admin_province)[T.Guangxi Zhuang Autonomous Region]	-0.0132	3.93e-05	-335.016	0.000	-0.013	-0.013
C(admin_province)[T.Guangzhou Province]	-0.0114	3.25e-05	-351.104	0.000	-0.011	-0.011
C(admin_province)[T.Guizhou Province]	-0.0069	5.19e-05	-132.115	0.000	-0.007	-0.007
C(admin_province)[T.Hainan Province]	-0.0071	5.76e-05	-124.109	0.000	-0.007	-0.007
C(admin_province)[T.Hebei Province]	0.0273	8.53e-05	320.115	0.000	0.027	0.027
C(admin_province)[T.Heilongjiang Province]	0.0125	6.65e-05	187.747	0.000	0.012	0.013
C(admin_province)[T.Henan Province]	0.0141	5.62e-05	250.491	0.000	0.014	0.014
C(admin_province)[T.Hong Kong]	-0.0001	0.000	-0.461	0.645	-0.001	0.000
C(admin_province)[T.Hubei Province]	-0.0012	3.91e-05	-30.113	0.000	-0.001	-0.001
C(admin_province)[T.Hunan Province]	-0.0102	3.19e-05	-319.655	0.000	-0.010	-0.010
C(admin_province)[T.Inner Mongolia Autonomous Region]	0.0630	9.69e-05	649.641	0.000	0.063	0.063
C(admin_province)[T.Jiangsu Province]	0.0186	6.14e-05	303.129	0.000	0.018	0.019
C(admin_province)[T.Jiangxi Province]	-0.0129	3.12e-05	-414.170	0.000	-0.013	-0.013
C(admin_province)[T.Jilin Province]	0.0020	6.84e-05	28.533	0.000	0.002	0.002
C(admin_province)[T.Liaoning Province]	0.0153	6.98e-05	218.876	0.000	0.015	0.015
C(admin_province)[T.Macau]	0.0171	0.001	22.511	0.000	0.016	0.019
C(admin_province)[T.Ningxia Ningxia Hui]	0.0640	0.000	535.546	0.000	0.064	0.064
C(admin_province)[T.Qinghai Province]	0.0574	0.000	371.531	0.000	0.057	0.058
C(admin_province)[T.Shaanxi Province]	0.0289	7.2e-05	401.378	0.000	0.029	0.029
C(admin_province)[T.Shandong Province]	0.0271	3.71e-05	729.447	0.000	0.027	0.027
C(admin_province)[T.Shanghai Municipality]	0.0352	0.000	85.890	0.000	0.034	0.036
C(admin_province)[T.Shanxi Province]	0.0357	0.000	315.882	0.000	0.036	0.036
C(admin_province)[T.Sichuan Province]	0.0074	3.84e-05	191.752	0.000	0.007	0.007
C(admin_province)[T.Tianjin Municipality]	0.0407	0.000	125.905	0.000	0.040	0.041
C(admin_province)[T.Tibet Autonomous Region]	0.0480	0.000	199.685	0.000	0.048	0.048
C(admin_province)[T.Unknown]	0.0044	7.82e-05	56.784	0.000	0.004	0.005
C(admin_province)[T.Xinjiang Uyghur Autonomous Region]	0.0777	9.22e-05	842.025	0.000	0.077	0.078
C(admin_province)[T.Yunnan Province]	0.0032	3.71e-05	87.214	0.000	0.003	0.003
C(admin_province)[T.Zhejiang Province]	-0.0134	4.61e-05	-290.957	0.000	-0.013	-0.013

Omnibus:	700828.763	**Durbin-Watson:**	1.271
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2507153.572
Skew:	1.015	**Prob(JB):**	0.00
Kurtosis:	6.683	**Cond. No.**	145.

Notes:
1. Standard Errors assume that the covariance matrix of the errors is correctly specified.

Attempt 2 - OLS with Distance to Closest Mine

Dep. Variable:	Cond	**R-squared:**	0.659
Model:	OLS	**Adj. R-squared:**	0.659
Method:	Least Squares	**F-statistic:**	1.326e+05
Date:	Wed, 30 Jul 2025	**Prob (F-statistic):**	0.00
Time:	23:09:53	**Log-Likelihood:**	7.0026e+06
No. Observations:	2337936	**AIC:**	-1.401e+07
Df Residuals:	2337901	**BIC:**	-1.400e+07
Df Model:	34		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0355	3.28e-05	1083.538	0.000	0.035	0.036
C(admin_province)[T.Beijing Municipality]	0.0197	0.000	84.723	0.000	0.019	0.020
C(admin_province)[T.Chongqing Municipality]	0.0014	6.17e-05	21.961	0.000	0.001	0.001
C(admin_province)[T.Fujian Province]	-0.0139	5.53e-05	-250.754	0.000	-0.014	-0.014
C(admin_province)[T.Gansu Province]	0.0614	0.000	575.570	0.000	0.061	0.062
C(admin_province)[T.Guangxi Zhuang]	-0.0145	4.85e-05	-298.023	0.000	-0.015	-0.014
C(admin_province)[T.Guangzhou Province]	-0.0131	4.38e-05	-298.207	0.000	-0.013	-0.013
C(admin_province)[T.Guizhou Province]	-0.0051	5.97e-05	-84.682	0.000	-0.005	-0.005
C(admin_province)[T.Hainan Province]	-0.0119	8.84e-05	-134.638	0.000	-0.012	-0.012
C(admin_province)[T.Hebei Province]	0.0299	9.58e-05	312.478	0.000	0.030	0.030
C(admin_province)[T.Heilongjiang Province]	0.0280	0.000	205.209	0.000	0.028	0.028
C(admin_province)[T.Henan Province]	0.0154	6.32e-05	243.579	0.000	0.015	0.016
C(admin_province)[T.Hong Kong]	-0.0027	0.000	-5.864	0.000	-0.004	-0.002
C(admin_province)[T.Hubei Province]	-0.0002	4.49e-05	-5.445	0.000	-0.000	-0.000
C(admin_province)[T.Hunan Province]	-0.0113	3.79e-05	-298.440	0.000	-0.011	-0.011
C(admin_province)[T.Inner Mongolia]	0.0681	0.000	601.609	0.000	0.068	0.068
C(admin_province)[T.Jiangsu Province]	0.0176	7.14e-05	246.053	0.000	0.017	0.018
C(admin_province)[T.Jiangxi Province]	-0.0127	3.91e-05	-325.720	0.000	-0.013	-0.013
C(admin_province)[T.Jilin Province]	0.0175	0.000	126.954	0.000	0.017	0.018
C(admin_province)[T.Liaoning Province]	0.0229	9.63e-05	237.880	0.000	0.023	0.023
C(admin_province)[T.Macau]	0.0143	0.001	10.011	0.000	0.011	0.017
C(admin_province)[T.Ningxia Ningxia Hui]	0.0665	0.000	504.707	0.000	0.066	0.067
C(admin_province)[T.Qinghai Province]	0.0643	0.000	365.123	0.000	0.064	0.065
C(admin_province)[T.Shaanxi Province]	0.0304	8.01e-05	379.266	0.000	0.030	0.031
C(admin_province)[T.Shandong Province]	0.0272	4.32e-05	628.938	0.000	0.027	0.027
C(admin_province)[T.Shanghai Municipality]	0.0359	0.000	80.704	0.000	0.035	0.037
C(admin_province)[T.Shanxi Province]	0.0355	0.000	288.588	0.000	0.035	0.036
C(admin_province)[T.Sichuan Province]	0.0090	4.63e-05	194.481	0.000	0.009	0.009
C(admin_province)[T.Tianjin Municipality]	0.0435	0.000	123.633	0.000	0.043	0.044
C(admin_province)[T.Tibet]	0.0606	0.000	218.502	0.000	0.060	0.061
C(admin_province)[T.Unknown]	0.0051	9.53e-05	53.447	0.000	0.005	0.005
C(admin_province)[T.Xinjiang]	0.0932	0.000	608.871	0.000	0.093	0.093
C(admin_province)[T.Yunnan Province]	0.0022	4.32e-05	51.848	0.000	0.002	0.002
C(admin_province)[T.Zhejiang Province]	-0.0130	5.15e-05	-251.315	0.000	-0.013	-0.013
distance_to_nearest_producer	-8.624e-09	6.21e-11	-138.917	0.000	-8.75e-09	-8.5e-09

Omnibus:	342200.156	**Durbin-Watson:**	1.553
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	893089.179
Skew:	0.818	**Prob(JB):**	0.00
Kurtosis:	5.548	**Cond. No.**	8.84e+07

Notes:
1. Standard Errors assume that the covariance matrix of the errors is correctly specified.
2. The condition number is large, 8.84e+07. This might indicate that there are
strong multicollinearity or other numerical problems.

Attempt 3 - OLS with Distance to Closest Mine and Processers

Dep. Variable:	Cond	**R-squared:**	0.658
Model:	OLS	**Adj. R-squared:**	0.658
Method:	Least Squares	**F-statistic:**	1.330e+05
Date:	Thu, 31 Jul 2025	**Prob (F-statistic):**	0.00
Time:	00:27:57	**Log-Likelihood:**	7.0353e+06
No. Observations:	2352768	**AIC:**	-1.407e+07
Df Residuals:	2352733	**BIC:**	-1.407e+07
Df Model:	34		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0343	3.34e-05	1027.130	0.000	0.034	0.034
C(admin_province)[T.Beijing Municipality]	0.0160	0.000	69.169	0.000	0.016	0.016
C(admin_province)[T.Chongqing Municipality]	-0.0016	5.83e-05	-27.102	0.000	-0.002	-0.001
C(admin_province)[T.Fujian Province]	-0.0145	5.53e-05	-262.707	0.000	-0.015	-0.014
C(admin_province)[T.Gansu Province]	0.0562	0.000	558.174	0.000	0.056	0.056
C(admin_province)[T.Guangxi Zhuang Autonomous Region]	-0.0145	4.86e-05	-297.842	0.000	-0.015	-0.014
C(admin_province)[T.Guangzhou Province]	-0.0124	4.36e-05	-283.712	0.000	-0.012	-0.012
C(admin_province)[T.Guizhou Province]	-0.0073	5.74e-05	-126.993	0.000	-0.007	-0.007
C(admin_province)[T.Hainan Province]	-0.0113	8.92e-05	-126.627	0.000	-0.011	-0.011
C(admin_province)[T.Hebei Province]	0.0268	9.35e-05	286.223	0.000	0.027	0.027
C(admin_province)[T.Heilongjiang Province]	0.0120	7.59e-05	158.502	0.000	0.012	0.012
C(admin_province)[T.Henan Province]	0.0136	6.21e-05	218.702	0.000	0.013	0.014
C(admin_province)[T.Hong Kong Special Administrative Region]	-0.0019	0.000	-4.169	0.000	-0.003	-0.001
C(admin_province)[T.Hubei Province]	-0.0014	4.44e-05	-30.751	0.000	-0.001	-0.001
C(admin_province)[T.Hunan Province]	-0.0114	3.81e-05	-300.108	0.000	-0.012	-0.011
C(admin_province)[T.Inner Mongolia Autonomous Region]	0.0616	0.000	583.641	0.000	0.061	0.062
C(admin_province)[T.Jiangsu Province]	0.0178	7.17e-05	247.580	0.000	0.018	0.018
C(admin_province)[T.Jiangxi Province]	-0.0131	3.92e-05	-333.596	0.000	-0.013	-0.013
C(admin_province)[T.Jilin Province]	0.0014	7.52e-05	18.419	0.000	0.001	0.002
C(admin_province)[T.Liaoning Province]	0.0148	7.69e-05	192.308	0.000	0.015	0.015
C(admin_province)[T.Macau Special Administrative Region]	0.0151	0.001	10.535	0.000	0.012	0.018
C(admin_province)[T.Ningxia Ningxia Hui Autonomous Region]	0.0635	0.000	484.958	0.000	0.063	0.064
C(admin_province)[T.Qinghai Province]	0.0569	0.000	337.029	0.000	0.057	0.057
C(admin_province)[T.Shaanxi Province]	0.0284	7.92e-05	358.558	0.000	0.028	0.029
C(admin_province)[T.Shandong Province]	0.0261	4.29e-05	609.425	0.000	0.026	0.026
C(admin_province)[T.Shanghai Municipality]	0.0347	0.000	77.630	0.000	0.034	0.036
C(admin_province)[T.Shanxi Province]	0.0349	0.000	284.014	0.000	0.035	0.035
C(admin_province)[T.Sichuan Province]	0.0068	4.39e-05	155.618	0.000	0.007	0.007
C(admin_province)[T.Tianjin Municipality]	0.0402	0.000	113.997	0.000	0.040	0.041
C(admin_province)[T.Tibet Autonomous Region]	0.0484	0.000	203.668	0.000	0.048	0.049
C(admin_province)[T.Unknown]	0.0036	9.66e-05	37.342	0.000	0.003	0.004
C(admin_province)[T.Xinjiang Uyghur Autonomous Region]	0.0775	9.9e-05	783.074	0.000	0.077	0.078
C(admin_province)[T.Yunnan Province]	0.0024	4.32e-05	54.536	0.000	0.002	0.002
C(admin_province)[T.Zhejiang Province]	-0.0139	5.13e-05	-271.102	0.000	-0.014	-0.014
distance_to_nearest_producer	-1.084e-09	3.53e-10	-3.076	0.002	-1.78e-09	-3.93e-10

Omnibus:	324241.513	**Durbin-Watson:**	1.545
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	917277.044
Skew:	0.750	**Prob(JB):**	0.00
Kurtosis:	5.666	**Cond. No.**	7.28e+06

Notes:
1. Standard Errors assume that the covariance matrix of the errors is correctly specified.
2. The condition number is large, $7.28e+06$. This might indicate that there are strong multicollinearity or other numerical problems.