Impact of different factors on the pollution-reduction and resource-saving effects of cleaner production

Peilei Zhang[†], *Ning Duan*[*,‡], *Zhigang Dan*[*,‡], *Jin Luo*[†], *Feifei Shi*[‡], *Huifeng Wang*[‡]

†School of Environment, Tsinghua University, No. 30, Shuangqing Road, Haidian district, Beijing, 100084, China.

‡Technology Center for Heavy Metal Cleaner Production Engineerings, Chinese Research Academy of Environmental Sciences, No.8, Dayangfang, Anwai Beiyuan, Chaoyang district, Beijing, 100012, China.

*Corresponding Authors

E-mail list:

493839818@qq.com (Peilei Zhang); ningduan@craes.org.cn (Ning Duan*, corresponding author); danzg@craes.org.cn (Zhigang Dan*, corresponding author); luojin765@gmail.com (Jin Luo); shifeifei789@163.com (Feifei Shi); 13311307830@163.com (Huifeng Wang).

---

9 pages including 8 tables, 3 figures and supplementary introduction of different statistical test methods.

SUPPORTING INFORMATION

Supplementary Introduction of Different Statistical Test Methods

In this manuscript, five statistical test methods were used in total. Following part will introduce their uses, null hypotheses, and the way to make statistical decision were introduced respectively.

Kolmogorov-Smirnov Test* and *Levine's Test introduction. *Kolmogorov-Smirnov Test* and *Levine's Test* were used to decide whether the *Analysis of Variance* was applicable. Two important requirements of *Analysis of Variance*'s application were that the data of every sub-factor complies with normal distribution and their variances were statistically same. The null hypothesis of *Kolmogorov-Smirnov Test* is the data of every sub-factor comply with normal distribution. If the *p*-value calculated by *rr* sample was greater than significance level (α), we should retain the null hypothesis, one requirement of *Analysis of Variance*'s application was meet. The null hypothesis of *Levine's Test* is the data of every sub-factor had the same variances. If the *p*-value calculated by *rr* sample was greater than *α*, we should retain the null hypothesis, and another requirement of *Analysis of Variance*'s application was met.

***Analysis of Variance* introduction.** If we retain the null hypotheses of *Kolmogorov-Smirnov Test* and *Levine's Test*, one-way (only analyzed one factor) *Analysis of Variance* was used to determine whether one factor impacts *rr* significantly. The null hypothesis of *Analysis of Variance* is that the mean *rr* of every sub-factor is the same. If the *p*-value calculated by the *rr* sample was less than *α*, we should reject the null hypothesis, the factor impacts *rr* significantly.

***Kruskal-Wallis Test* introduction.** If one-way *Analysis of Variance* indicated that the analyzed factor impacted *rr* significantly, *Kruskal-Wallis Test* was used to further verify the statistical conclusion of *Analysis of Variance* (analyzed factor impacted *rr* significantly). The null hypothesis of *Kruskal-Wallis Test* is that the *rr* of different sub-factors had the same distribution. If the *p*-value calculated by *rr* sample was greater than *α*, we should retain the null hypothesis, further indicating that the analyzed factor impacts *rr* significantly.

***Least Significant Difference* introduction.** When both *Analysis of Variance* and the *Kruskal-Wallis Test* showed that the factor significantly impacted *rr*, *Least Significant Difference* was used to identify the sub-factor pairs (such as to industry sectors) in which the *rr* of one sub-factor was significantly higher or lower than that of other sub-factor. The null hypothesis of *Least Significant Difference* is that the *rr* of two factors is the same. If the *p*-value calculated by *rr* sample was less than *α*, we should reject the null hypothesis, and the mean *rr* of two sub-factors were different.

Another use of *Analysis of Variance*. Moreover, except one-way *Analysis of Variance*, two-way *Analysis of Variance* was also used. The intention of two-way *Analysis of Variance* was to judge whether the significant influence of industry factor was caused by the difference of resource and pollutant types in different industry sector. In this study, three types of statistic, including *mean square*, *F*-value, and *p*-value, were used to analyze the reciprocal effect between pollutant and resource type and industry sector. *Mean square* measures the average variation of *rr* caused only by one factor, and is used to calculate *F*-value. Based on *F*-value we can obtain the *p*-value of one analyzed factor. Note that for two analyzed factors there are two *p*-values. If the *p*-vale of one analyzed factor was less than *α*, the analyzed factor still had a significant effect on *rr*, given that other factors remained unchanged. Specific results were shown in Table 4.

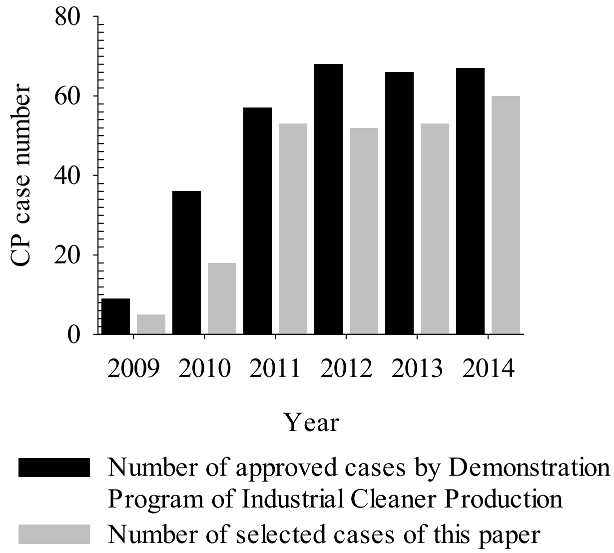

Figure S1. CP cases over different years

	2009	2010	2011	2012	2013	2014
Anhui	61.15		49.43	49.04	29.33	52.62
Bejing					41.48	
Fujian			34.00	50.00	40.00	
Gansu					100.00	
Guangdong		12.74	50.00	48.86		
Guangxi		75.50	20.00	30.00		
Hebei	55.00	60.14	32.18	50.55	100.00	68.59
Henan	15.50		61.09	52.48	57.14	58.49
Heilongjiang				31.97	41.47	78.26
Hubei			35.31	59.00	59.08	78.51
Hunan		32.50	60.57	56.71	100.00	60.14
Jilin		40.00	45.85		81.19	
Jiangsu			57.85	53.33	48.30	54.56
Jiangxi			19.72	59.65	64.03	66.54
Liaoning		100.00	73.87		78.03	95.00
Inner Mongolia			30.99	43.17		53.00
Ningxia	72.12		28.97	95.00	68.09	37.21
Qinghai					30.00	100.00
Shuandong			47.21	51.92	50.46	49.49
Shanxi			63.00	26.50		
Shaanxi		84.62		78.67		80.00
Sichuan			40.10	48.79	100.00	13.30
Tibet			45.49	70.00		
Sinkiang		50.00	49.66	100.00		43.39
Yunnan		66.59	55.97			
Zhejiang			55.91	32.19	50.00	77.24
Chongqing			51.00	50.85		

Mean *rr* of combinatioins (%): 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100

Figure S2. The mean *rr* of different region and year combinations

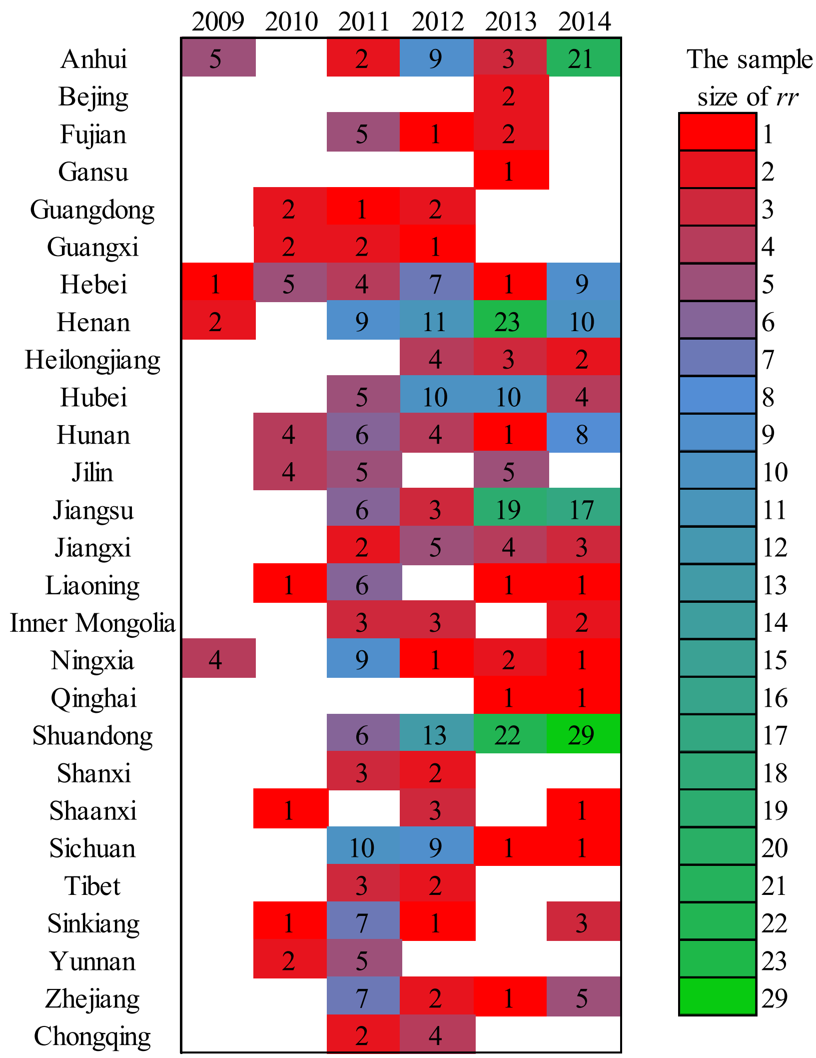

Figure S3. The sample size of the *rr* in different region and year combinations

Table S1. The normality tests of the *rr* of different years based on

Kolmogorov-Smirnov Test

Year	*p*	Statistical conclusion
2011	0.09	The rr of 2011 followed N~(47.41%, 27.41%)
2012	0.15	The rr of 2012 followed N~(52.46%, 28.30%)
2013	0.23	The rr of 2013followed N~(55.53%, 30.24%)
2014	0.27	The rr of 2014 followed N~(57.37%, 28.25%)

Table S2. The homogeneity test of the *rr* variances over different years based on

Levene's Test

Levene Statistic*	df_1 **	df_2	*p*	Statistical conclusion
0.905	3	421	0.44	The variances of rr in every year of 2011~2014 were the same.

**Levene statistic* is the tested statistic of *Levene's Test*; **The "*df*" is the acronym of the degree of freedom, which was used to decide the *p* value for *Levene's Test*. The "*df1*" and "*df2*" represented two degrees of freedom of *Levene's test*.

Table S3. The normality tests of the *rr* over different provincial regions based on *Kolmogorov-Smirnov Test*

Regions	*p*	Statistical conclusion
Jiangsu Province	0.60	The distribution of the rr of this region was normal with mean 52.27% and S.D 31.28%.
Shandong Province	0.39	The distribution of the rr of this region was normal with mean50.05% and S.D 26.14%.
Henan Province	0.32	The distribution of the rr of this region was normal with mean55.58% and S.D 30.20%.
Hebei Province	0.43	The distribution of the rr of this region was normal with mean57.61% and S.D 33.33%.
Sichuan Province	0.29	The distribution of the rr of this region was normal with mean45.40% and S.D 27.15%.
Hubei Province	0.83	The distribution of the rr of this region was normal with mean 57.63% and S.D 27.81%.
Hunan Province	0.62	The distribution of the rr of this region was normal with mean 56.58% and S.D 32.22%.
Anhui Province	0.21	The distribution of the rr of this region was normal with mean 50.98% and S.D 28.81%.
Ningxia Hui Autonomous Region	0.45	The distribution of the rr of this region was normal with mean 48.09% and S.D 34.56%.

Table S4. The homogeneity test of the *rr* variances over different provincial regions based on *Levene's Test*

Levene Statistic	df_1	df_2	*p*	Statistical conclusion
1.565	8	318	0.13	The variances of *rr* in different provincial regions were the same.

Table S5.The normality tests of the *rr* over different pollutants and resources based on *Kolmogorov-Smirnov Test*

Resoruce or pollutant	*p*	Statistical conclusion
Fresh water	0.81	The distribution of the *rr* of fresh water was normal with mean 50.29% and S.D 26.31%.
Wastewater	0.90	The distribution of the *rr* of wastewater was normal with mean56.91% and S.D 25.95%.
COD	0.47	The distribution of the *rr* of COD was normal with mean58.41% and S.D 25.90%.
Heavy metal	0.10	The distribution of the *rr* of heavy metal was normal with mean 78.93% and S.D 22.23%.
Exhaust gas	0.74	The distribution of the *rr* of exhaust gas was normal with mean 58.03% and S.D 26.76%.

Table S6. The homogeneity test of the *rr* variances over different pollutants and resources based on *Levene's Test*

Levene Statistic	df_1	df_2	*p*	Statistical conclusion
0.498	4	248	0.74	The variances of *rr* in different provincial regions were the same.

Table S7. The normality tests of the *rr* over different industry sectors based on *Kolmogorov-Smirnov Test*

Industries	p	Statistical conclusion
Primary chemical industry	0.26	The distribution of the *rr* of primary chemical industry was normal with mean 55.45% and S.D 29.96%.
Food industry	0.36	The distribution of the *rr* of food industry was normal with mean 45.04% and S.D 26.83%.
Pesticide industry	0.21	The distribution of the *rr* of pesticide industry was normal with mean 60.20% and S.D 32.42%.
Metallurgical industry	0.88	The distribution of the *rr* of metallurgical industry was normal with mean 60.02% and S.D 25.97%.
Textile industry	0.15	The distribution of the *rr* of textile industry was normal with mean 50.05% and S.D 28.41%.
Battery industry	0.35	The distribution of the *rr* of battery industry was normal with mean 54.98% and S.D 28.88%.

Table S8. The homogeneity test of the *rr* variances over different industry sectors based on *Levene's Test*

Levene Statistic	df_1	df_2	p	Statistical conclusion
2.093	5	353	0.07	The variances of *rr* in different industries were the same.