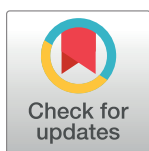


RESEARCH ARTICLE

Evidence of air quality data misreporting in China: An impulse indicator saturation model comparison of local government-reported and U.S. embassy-reported PM_{2.5} concentrations (2015–2017)

Jesse S. Turiel^{1*}, Robert K. Kaufmann²

1 Ash Center for Democratic Governance and Innovation, Harvard University John F. Kennedy School of Government, Cambridge, Massachusetts, United States of America, **2** Department of Earth and Environment, Boston University, Boston, Massachusetts, United States of America

* jsturiel@hks.harvard.edu**OPEN ACCESS**

Citation: Turiel JS, Kaufmann RK (2021) Evidence of air quality data misreporting in China: An impulse indicator saturation model comparison of local government-reported and U.S. embassy-reported PM_{2.5} concentrations (2015–2017). *PLoS ONE* 16(4): e0249063. <https://doi.org/10.1371/journal.pone.0249063>

Editor: James R. Lyons, Arizona State University, UNITED STATES

Received: June 5, 2020

Accepted: February 26, 2021

Published: April 21, 2021

Copyright: © 2021 Turiel, Kaufmann. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data has been published in the Harvard Dataverse repository under the title "Government vs Embassy-Reported Hourly PM_{2.5} Concentrations in 5 Mainland Chinese Cities (2015–2017)". The URL is: <https://doi.org/10.7910/DVN/SPSJNR>. The DOI is: [10.7910/DVN/SPSJNR](https://doi.org/10.7910/DVN/SPSJNR).

Funding: The authors received no specific funding for this work.

Abstract

This paper analyzes hourly PM_{2.5} measurements from government-controlled and U.S. embassy-controlled monitoring stations in five Chinese cities between January 2015 and June 2017. We compare the two datasets with an impulse indicator saturation technique that identifies hours when the relation between Chinese and U.S. reported data diverges in a statistically significant fashion. These temporary divergences, or impulses, are 1) More frequent than expected by random chance; 2) More positive than expected by random chance; and 3) More likely to occur during hours when air pollution concentrations are high. In other words, relative to U.S.-controlled monitoring stations, government-controlled stations systematically under-report pollution levels when local air quality is poor. These results contrast with the findings of other recent studies, which argue that Chinese air quality data misreporting ended after a series of policy reforms beginning in 2012. Our findings provide evidence that local government misreporting did not end after 2012, but instead continued in a different manner. These results suggest that Chinese air quality data, while still useful, should not be taken entirely at face value.

1. Introduction

For several decades, air quality in China has consistently ranked among the world's worst. Since the beginning of the reform era in 1978, Chinese air pollution has caused tens of millions of deaths [1] and reduced GDP by trillions of dollars [2]. Nonetheless, widespread reporting of real-time air quality data in China began only recently. Today, nearly all Chinese air quality monitoring stations are owned and operated by government officials, which raises questions about the reliability and integrity of the reported data. Because substandard environmental performance can cause local party leaders to be punished and denied promotion, they have a

Competing interests: The authors have declared that no competing interests exist.

strong incentive to alter the data reported to the public and the central government in a way that understates pollution.

The tendency of Chinese officials to misreport environmental data is widely documented. Government statistics misreport fish catch [3], coal use [4], GDP growth [5], and carbon emissions [6]. Government data also understates the rate at which agricultural land is converted to urban areas [7] and the rate at which burning coal emits carbon dioxide [8]. These misrepresentations make it difficult for the international community to assess Chinese compliance with international treaties [9–11] and prevent domestic policymakers from accurately gauging environmental impacts. For example, government data indicate that urban air quality improved significantly over the past several years [12–14]. While independent data sources confirm this positive trend [15, 16], the magnitude of improvements could be exaggerated if the Chinese government misreports the data.

Chinese central leaders increasingly recognize the dangers posed by inaccurate environmental information, and in recent years government reforms have attempted to improve oversight and increase penalties for officials accused of data fraud. However, government efforts to fix these institutional problems beg the question: are such efforts effective? In this paper, we analyze hourly data for PM_{2.5} from five Chinese cities to evaluate the success/failure of central government efforts to eliminate misreporting of local air quality data.

1.1 Bureaucratic incentives for local air quality data misreporting in China

In the absence of democratic elections, all local state and party leaders in China are appointed by government officials at higher levels of the political system. The placement and promotion of these officials is determined by the cadre evaluation system (CES), which ranks the performance of local leaders using a formula that weights various ‘hard’ and ‘soft’ performance metrics. While the CES has traditionally emphasized economic growth, family planning, and maintaining social order, since 2012 the central government has also affirmed environmental performance as an important ‘hard’ target [17]. Each year, environmental targets are passed from central to local party leaders, who then sign a ‘target responsibility document’ with the director of the local environmental protection bureau (EPB). The EPB enforces pollution control measures and collects local pollution data that are then submitted to the central environmental ministry and released to the public [18].

Due to the fragmented nature of China’s political system, city-level EPBs effectively serve two ‘masters,’ each with different policy goals [19, 20]. First, city-level EPBs must achieve the targets set by higher-level environmental bureaucrats, who tend to favor stricter pollution control policies. Second, they must also maintain the support of the city’s Communist Party standing committee, which typically has close ties with local business leaders and tends to favor more growth-oriented policies. In response to such conflicting demands, city-level EPBs use short-term coping strategies to achieve their bureaucratic goals [21]. With bottom-line objectives prioritized above all else, local bureaucrats face immense pressure to report the ‘correct’ numbers to their higher-ups, and some resort to colluding with other local officials or misreporting data [22, 23]. Given these institutional incentives to cheat, official air pollution data in China often is treated with a high degree of skepticism, by both outside observers and the general public.

1.2 Evidence of pre-2012 data tampering

In China, concerted attempts to control local-scale air pollution began in 1996 with the adoption of the National Ambient Air Quality Standards (NAAQS). NAAQS mandated the collection and publication of daily data for the atmospheric concentrations of three pollutants:

sulphur dioxide (SO₂), nitrogen dioxide (NO₂), and suspended particulates with a diameter of 10 microns or less (PM₁₀). These three values were aggregated to form a composite measure of general air quality called the air pollution index (API), which ranged from 0 to 500. By the early 2000s, 86 cities across China reported daily API values to the central environmental ministry (then called the State Environmental Protection Administration, or SEPA), which released the data to the public online [24].

At the end of each year, SEPA used API data to rank the performance of cities from best to worst, with the most important metric being the annual percentage of ‘blue sky days’ (days with an average API less than 100). Cities with at least 85% blue sky days were awarded full credit for air pollution control and could be designated as a ‘National Model City for Environmental Protection.’ This designation was intended to spur competition and improve local environmental quality, as city leaders could receive favorable publicity and possibly even increase their odds of promotion [25].

However, the arbitrary dividing line between ‘good’ and ‘bad’ air quality created a strong incentive for local leaders to misreport data when API was close to the blue sky day threshold of 100. Distortions around the blue sky threshold were first noted in 2008 by Andrews [26], who analyzed air quality data from Beijing and found a much higher-than-expected frequency of API values just below 100 (and a correspondingly lower-than-expected frequency just above 100). These distortions were confirmed in a 2012 study by Chen et al. [25], who used daily API data from 37 Chinese cities and found a statistically significant discontinuity at the blue sky threshold of 100. Expanding on this work, Ghanem and Zhang [27] tested air quality data from 86 Chinese cities between 2001 and 2010, finding sharp discontinuities at the blue sky threshold for roughly half the cities in their dataset. Notably, these discontinuities were more pronounced on days with low wind speed and high visibility, which suggests that city officials were more likely to misreport data on days when pollution was harder to visibly detect.

1.3 Post-2012 government reforms: The end of local air quality data misreporting?

Beginning in 2012, the central government restructured the country’s air quality monitoring system, in part to discourage local officials from misreporting data [28, 29]. The number of monitored cities was increased from 86 to 363, and API was replaced with the more sophisticated and comprehensive Air Quality Index (AQI), which added ground-level ozone (O₃), carbon monoxide (CO), and suspended particulates with a diameter of 2.5 microns or less (PM_{2.5}). Additionally, after a two-year transition period, all cities were required to report hourly pollution concentrations instead of daily averages, with all data relayed directly from local monitoring stations to the central environmental ministry without any handling by city EPBs. Finally, the blue sky day metric was officially discontinued in 2013. Under the revised Air Pollution Prevention and Control Action Plan (APPCAP), cities were instead judged by their ability to reduce average annual particulate concentrations between 2012 and 2017 [30].

Early reports suggested that these reforms reduced the misreporting of pollution data. Applying Benford’s Law (a statistical benchmark often used to detect financial fraud) to observations between 2008 and 2012, Stoerk [31] found that differences between concentrations of particulate matter reported by government-controlled monitoring stations and the U.S. embassy in downtown Beijing suggested manipulation. However, including observations for 2013 changed the results to suggest that manipulation stopped in that year. Similarly, using data collected between January 2013 and December 2015, Liang et al. [32] showed that hourly concentrations of PM_{2.5} reported by government-controlled monitoring were not lower than concentrations reported at U.S.-controlled stations in five Chinese cities. Together, these

results suggest that misreporting of air quality data no longer is an acute problem, at least in China's largest megacities.

Contrary to these recent findings, we postulate that some government measurements of local air quality still are being misreported. However, the form of misreporting has changed; instead of manipulating around a given threshold, Chinese officials are now more likely to understate pollution during periods when concentrations are high. We test this hypothesis by estimating the relation between hourly PM_{2.5} concentrations measured by U.S. embassies and consulates in five Chinese cities—which we assume to be reported accurately—and concentrations reported by municipal governments in those same five cities. We choose PM_{2.5} because it is the only pollutant measured by stations controlled by both the U.S. and Chinese governments.

2. Data and methods

2.1 Air quality data

Since January 1, 2015, China's central environmental ministry (now called the Ministry of Ecology and Environment, or MEE) has published continuous, real-time hourly measurements of PM_{2.5} in 363 Chinese cities. These data are available to the public for 48 hours, after which they are removed from the MEE website. To obtain these deleted data, we use scraped and archived data from Beijing Sinaapp (publicly available at <<https://beijingair.sinaapp.com>> [33]), which is the only website to preserve continuous hourly measurements of PM_{2.5} concentrations for the five cities in our sample. These data, and all other air quality data used in our paper, are available for public use, and our research complies with the websites' terms and conditions.

Each city has several local government-controlled monitoring stations. To create a single value for each city that can be compared to the single U.S. station, we create an average hourly value from values reported by individual Chinese stations in each city. This average weights measurements from individual stations based on the inverse of their distance from the station controlled by the U.S. embassy as follows:

$$GOVT_t = \frac{\sum_{i=1}^n \left(\frac{z_{it}}{d_i} \right)}{\sum_{i=1}^n \left(\frac{1}{d_i} \right)} \quad (1)$$

in which z_{it} represents the hourly PM_{2.5} concentration at a given local government-controlled monitoring station, and d_i represents that station's distance (in km) from the city's U.S. embassy or consulate. This inverse distance weighting (IDL) gives the largest weight to government-controlled stations closest to the U.S. embassy.

To test the degree to which our findings are robust to the weighting scheme used in Eq (1), we repeat the analysis using inverse quadratic distance weighting (i.e. $\frac{1}{d_i^2}$). The results of this alternative weighting specification, which are described in S1-S3 Tables in [S1 File](#), do not affect our conclusions in a substantive manner.

Observations for hourly PM_{2.5} concentrations at stations operated by U.S. embassies are obtained from the U.S. State Department, and are publicly available at <<https://china.usembassy-china.org.cn>> [34]. Measurements at U.S. embassies extend through June 30, 2017, which allows us to compare hourly PM_{2.5} concentrations during a 30-month period (January 2015-June 2017) when the two datasets overlap.

As described S4 Table in [S1 File](#), each city contains eight to twelve stations that generally are within 10 km of the station controlled by the U.S. government. The large number of

stations per city implies that conditions unique to a single station have relatively little effect on the city average. The importance of local weather conditions is reduced further by the proximity of stations and locations that ‘surround’ the station controlled by the U.S. government.

We also collect hourly PM_{2.5} concentration data from 74 government-controlled monitoring stations in Taiwan, publicly available at <<https://airtw.epa.gov.tw/ENG/default.aspx>> [35], which we use to test our methodology for Type I errors (Section 3.5).

2.2 Statistical methodology

Consistent with our hypothesis that officials report lower values when concentrations are high, we identify the hours when the relation between measurements at embassy-controlled and government-controlled monitoring stations changes in a statistically significant fashion using the following general equation:

$$EMB_t = \alpha_1 + \beta_1 GOVT_t + \beta_2 IMP_t + \varepsilon_t \quad (2)$$

in which EMB_t is the concentration of PM_{2.5} ($\mu\text{g}/\text{m}^3$) measured by U.S. embassy-controlled stations during hour t ; $GOVT_t$ is the corresponding inverse-distance-weighted average concentration measured by local-government controlled stations; IMP_t is an impulse that identifies a statistically significant change in this relation at hour t ; α , β_1 , and β_2 are regression coefficients; and ε_t is a random, heteroskedastic regression residual.

In the absence of misreporting, geographical, or meteorological differences between stations, we expect $\alpha = 0$ and $\beta_1 = 1.0$. These expectations are strongly rejected by estimating the relation across all five cities; $\hat{\alpha}_1 = 4.42$ ($t = 38.1, p < 0.000001$) and $\hat{\beta}_1 = 1.016$ ($t = -13.9, p < 0.000001$). Alone, these results do not indicate misreporting because measurements can be affected by the instruments used to measure concentrations as well as geographic and meteorological differences between stations. For example, U.S. embassies and consulates are typically located in the urban core of cities, whereas government-controlled monitoring stations are diffused throughout each city, and so embassy-controlled stations are likely to record slightly higher pollution values (e.g. $\alpha_1 > 0$ and/or $\beta_1 > 1.0$). As such, empirical estimates of $\hat{\alpha}$ and $\hat{\beta}$ alone cannot be used to detect misreporting.

Misreporting could be detected by testing the relation $EMB_t = \alpha_1 + \beta_1 GOVT_t + \varepsilon_t$ for one or more changes in α_1 and/or β_1 [36, 37]. However, this approach is not well suited for testing our hypothesis because change points identify systemic changes that continue over an extended period. For example, if one or both of the instruments used to measure PM_{2.5} is not maintained correctly or replaced, the relation between measurements could change for an extended period, and this change would likely alter the intercept α_1 . Furthermore, extended changes in α_1 and/or β_1 are inconsistent with our hypothesis because systematic changes would be easier to detect than changes in individual measurements.

To detect changes for short periods when the value of EMB_t is high, we focus on the quantity, sign, and timing of values of β_2 in Eq (1). Values of β_2 that are statistically different from zero identify hours when the relation between measurements reported by Chinese and U.S. controlled stations differs from the relation that prevails during most other hours. This approach is flexible because it contains no *a priori* assumptions about misreporting. Misreporting can occur at any time, be positive or negative, and have any magnitude. These hourly divergences are termed impulses and they can be used to detect misreporting by testing the following hypotheses:

- **Null Hypothesis #1:** The number of impulses is consistent with random chance. Rejecting this null hypothesis would indicate the data are being misreported.

- **Null Hypothesis #2:** The number of negative impulses equals the number of positive impulses. Rejecting this null hypothesis would indicate the data are being misreported in a specific direction.
- **Null Hypothesis #3:** The timing of the impulses is random. Rejecting the third null hypothesis would indicate that data are being misreported during strategically important hours.

To evaluate the degree to which our methodology is robust, we check for Type I and Type II errors by testing hypotheses #4 and #5:

- **Null Hypothesis #4:** Using similar hourly PM_{2.5} concentration data from Taiwan, where no misreporting is suspected, the number of impulses is consistent with random chance. Rejecting this null hypothesis would indicate that the methodology is prone to false positives (Type I errors) and is therefore not well suited to detecting whether the Chinese government misreports data.
- **Null Hypothesis #5:** Using data from two Chinese cities where local officials were caught manipulating data at specific times and locations, the number of impulses during this period is consistent with random chance. Failing to reject this null hypothesis would indicate that the IIS methodology is not well suited to detecting misreporting when and where it is already known to have occurred (Type II error).

The impulses in Eq (2) that are used to test these hypotheses are identified using an econometric technique, impulse indicator saturation (IIS) [38, 39]. IIS creates an impulse (a zero-one dummy variable) for every hourly observation in a sample. Impulses are dropped/retained by (1) dropping irrelevant variables (gauge) based on a specified nominal significance level, (2) retaining relevant variables (potency) near the theoretical average power, based on a significance level that is specified by the user. We specify a significance level of $p = .01$ because it reduces the likelihood that the impulses are chosen by random chance, but still identifies a sufficient number of observations via random chance that the results can be evaluated statistically (see Eq (3)). The asymptotics of IIS are thoroughly documented [40].

Although developed by statisticians to analyze econometric data, the IIS procedure is used well beyond economics. For example, the IIS procedure is used to identify the timing of the so-called hiatus in climate change [41], and also to identify periods when speculation and policy changes move oil prices away from market fundamentals [42].

To identify significant impulses in our dataset, we use the IIS procedure to estimate Eq (2) using the R-package *gets* (<https://cran.r-project.org/web/packages/gets/index.html>). Because this software cannot analyze the large number of hourly observations (~20,000) for each city, we break the thirty-month sample into fourteen subsamples, each containing roughly 1,500 hourly observations. Results generated by analyzing these subsamples are not affected by the number of splits, or by choosing unequal splits [43]. Similarly, the IIS methodology is able to analyze both stationary and non-stationary autoregressive processes without bias [44].

All of our tests are based on the null hypothesis that random chance generates the impulses. Using a significance level of $p = .01$ implies that random chance would identify one impulse for every 100 hourly observations. We test the null hypothesis that the number of impulses retained is not different than the number expected based on random chance with a test developed specifically for the IIS methodology [45]:

$$S_{\text{prop}} = n^{1/2}(\tilde{y}_c - \gamma_c) \quad (3)$$

in which \tilde{y}_c is the observed proportion of impulses and γ_c denotes the gauge (the expected proportion of detected impulses under the null hypothesis of no impulses) in the initial step of the

impulse indicator saturation algorithm. The normal approximation to the gauge is first established by selecting the fixed cut-off c to control the frequency of wrongly detected impulses as the sample size n increases. Under the null hypothesis of no impulses in the model, the expected and observed proportion of impulses should be equal and the S_{prop} statistic (with appropriate scaling) follows a standard normal distribution. Rejecting the null hypothesis using the S_{prop} statistic provides evidence that the observed proportion of impulses is significantly different from $\gamma_c = 0.01$, which is the proportion expected by random chance. Rejecting the null hypothesis implies two alternatives; large but infrequent shifts in measuring equipment and/or meteorological conditions, or data misreporting.

We choose between these two possibilities based in part on the signs of $\hat{\beta}_2$ (Eq (2)). Under the null hypothesis of no misreporting, there is no *a priori* reason to expect more positive or negative impulses. However, if government officials periodically underreport hourly $\text{PM}_{2.5}$ concentrations relative to the ‘true’ values measured by U.S. embassies and consulates, we would expect more positive impulses than negative impulses, because a positive impulse identifies an hour when the $\text{PM}_{2.5}$ concentration reported by the U.S. embassy is significantly greater ($p < 0.01$) than implied by the corresponding Chinese station, as given by $(\hat{\alpha} + \hat{\beta}_1 \text{GOVT}_t)$. We test the null hypothesis that the number of positive and negative impulses are equal with a z-statistic for the proportion of positive impulses ($\beta_{2t}^+ = \beta_{2t}^-$) and a t-statistic that the mean value of the impulses ($\bar{\beta}_2$) equals zero ($\frac{\bar{\beta}_2 - 0.0}{\text{se}_{\bar{\beta}_2}}$).

Finally, we evaluate the null hypothesis that random chance generates the impulses by testing whether impulses are distributed randomly throughout the sample. If local Chinese officials misreport measurements, they are more likely to understate pollution when true $\text{PM}_{2.5}$ concentrations are high. Under this alternative hypothesis, we would expect a positive relation between the observed impulses (β_2) and $\text{PM}_{2.5}$ concentrations measured at U.S. embassies. We test this third hypothesis by estimating a logistic regression given by Eq (4):

$$\text{Impulse}_t^+ = \alpha_2 + \beta_3 \text{EMB}_t + \varepsilon_t \quad (4)$$

in which Impulse_t^+ is a binary variable that equals one for hours when $\beta_2 > 0$ (positive impulses). We also estimate a second logistic regression in which the dependent variable equals one for hour(s) when any impulse—either positive or negative—is present ($\beta_2 \neq 0$). For Eq (4), positive values of β_3 indicate that high concentrations (as measured by U.S. embassies) increase the likelihood of a positive impulse. This would suggest that Chinese government officials underreport concentrations of $\text{PM}_{2.5}$ during periods of heavy pollution, a result which would be consistent with purposeful misreporting. We calculate the threshold (X) at which local officials are likely to understate concentrations as follows:

$$X = \exp\left(\frac{\ln\left(\frac{0.5}{1-0.5}\right) - \alpha_2}{\beta_3}\right) \quad (5)$$

where X is the $\text{PM}_{2.5}$ concentration at which the likelihood of a positive impulse reaches 50%, and α_2 and β_3 are regression coefficients from the logit model (Eq (4)).

3. Results

3.1 Null hypothesis #1: Differences between Chinese and U.S. station measures are generated by random chance

Strong evidence for data misreporting is provided by the higher-than-expected frequency of impulses in Eq (2). Using a significance level of 0.01 to identify impulses implies that random

Table 1. Test of null hypothesis #1 for the five Chinese cities and pooled model.

Location	Hourly Obs.	Expected Impulses ($p = 0.01$)	Observed Impulses	Mean Impulse Coefficient (β_2)	Null Hypothesis #1	
					S_{prop} Statistic	P-Value
Beijing	20,843	208	310	18.99	8.33	< 0.001**
Shenyang	20,139	201	293	21.23	7.65	< 0.001**
Shanghai	19,557	196	189	-2.01	-0.56	0.578
Guangzhou	20,292	203	273	16.41	5.83	< 0.001**
Chengdu	20,414	204	325	6.403	10.02	< 0.001**
Pooled	101,245	1,012	1,390	12.89	14.06	< 0.001**

The test statistic rejects the null hypothesis at $p = 0.05$ (*) and $p = 0.01$ (**).

<https://doi.org/10.1371/journal.pone.0249063.t001>

chance will generate roughly 1,012 impulses from the 101,245 hourly observations across all five sample cities. However, using this pooled sample to estimate Eq (2) identifies 1,390 impulses, roughly 40% more than expected by random chance. This greater-than-expected number of impulses also is present in four of the five cities (except Shanghai) when tested individually. These results are confirmed statistically by the positive and significant values for the S_{prop} statistic (Section 2.2), which indicate that random chance likely did not generate the large number of impulses estimated from Eq (2) (Table 1). This suggests that the higher-than-expected frequency of observed impulses in four of the five tested cities is caused by non-random weather patterns, changes/poor maintenance of monitoring equipment, and/or purposeful misreporting by local Chinese officials. Two graphical examples of specific 24 and 36-hour time periods with high concentrations of significant impulses (in Beijing and Shenyang respectively) are provided in S1 and S2 Figs in S1 File.

3.2 Null hypothesis #2: The number of positive and negative impulses are equal

While a higher-than-expected frequency of impulses can signify a non-random data generation process, the signs associated with β_2 also contain important information about how they are generated. If random chance generates the impulses, the number of positive and negative impulses should be roughly equal. The specification of Eq (2) makes $\beta_2 > 0$ when measurements at stations controlled by the Chinese government are significantly lower than the value implied by the corresponding measurement at the U.S. embassy-controlled station. In the pooled sample (and in four of the five individual cities), more than 63% of the observed impulses are positive. A z-statistic (Table 2) indicates that, for the pooled model and all individual cities except for Shanghai, it is highly unlikely ($p < 0.01$) that random chance generates the preponderance of positive impulses. This result is confirmed by a two-tailed, one sample t-statistic (Table 2) which evaluates the null hypothesis that the mean coefficient value of significant impulses is

Table 2. Test of null hypothesis #2 for the five Chinese cities and pooled model.

Location	Observed Impulses	% Positive Impulses	Null Hypothesis #2			
			Z-Statistic: $IMPt(+) = IMPt(-)$	P-Value	T-Statistic: $\beta_2 = 0$	P-Value
Beijing	310	63.50%	4.71**	< 0.001**	3.06	0.002**
Shenyang	293	78.80%	9.81**	< 0.001**	3.76	< 0.001**
Shanghai	189	48.70%	-0.29	0.772	-0.77	0.442
Guangzhou	273	67.80%	5.81**	< 0.001**	5.28	< 0.001**
Chengdu	325	73.20%	8.32**	< 0.001**	2.72	0.007**
Pooled	1,390	67.80%	13.28**	< 0.001**	5.99	< 0.001**

<https://doi.org/10.1371/journal.pone.0249063.t002>

zero. Thus, not only are impulses more frequent than expected but, when they do occur, government-controlled stations are far more likely than U.S. embassy-controlled stations to report concentrations lower than the values implied by $EMB_t = \alpha_1 + \beta_1 GOVT_t + \varepsilon_t$. Although, by itself, this result does not reject the possibility that the high frequency of impulses is generated by weather conditions and/or changes in/poor maintenance of monitoring equipment, the preponderance of positive impulses suggests that the observed divergences are *directional*, which is clearly consistent with the hypothesis that local Chinese officials misreport data in ways that understate population exposure to high concentrations of $PM_{2.5}$.

3.3 Null hypothesis #3: The timing of positive impulses is random

If random chance generates the excessive number of positive impulses, they would likely occur randomly throughout the sample. If impulses are generated by non-random weather patterns or changes/poor maintenance of monitoring equipment, the excessive number of positive impulses would likely cluster during periods throughout the sample. By contrast, if the excessive number of positive impulses is generated by Chinese officials purposefully understating measurements, positive impulses are likely to be positively correlated with concentrations measured at U.S. embassies. That is, because officials want to lower annual average pollution concentrations, the greatest incentive to understate pollution occurs during periods when 'true' $PM_{2.5}$ concentrations are unusually high. Consistent with this notion, we find a strong positive correlation between observed impulses and hourly $PM_{2.5}$ concentrations measured at U.S. embassies. This positive correlation is evident for the pooled sample (Fig 1), and for each of the five individual cities (Table 3).

The positive correlation suggested by Fig 1 and Table 3 are confirmed by the results of the logit model (Eq (4)). The positive coefficients (β_3) associated with $PM_{2.5}$ concentrations measured at U.S. stations indicate that higher concentrations of $PM_{2.5}$ increase the likelihood of positive impulses (Table 3). These results are confirmed by a second logit model in which the dependent variable assumes a value of one for every significant impulse, regardless of whether it is positive or negative (S5 Table in S1 File). The $PM_{2.5}$ concentration at which the likelihood of a positive impulse surpasses 50% (which we term the 50% misreporting threshold) is $502 \mu\text{g}/\text{m}^3$ in the pooled sample, and ranges from $180 \mu\text{g}/\text{m}^3$ to $625 \mu\text{g}/\text{m}^3$ for individual cities.

To approximate the magnitude of misreporting, we compare average reported $PM_{2.5}$ values between Chinese and U.S.-controlled stations for all positive impulses above the 50% misreporting threshold (Fig 2). The results show that, during hours with significant positive impulses, Chinese-controlled stations underreport U.S.-controlled stations by between $63 \mu\text{g}/\text{m}^3$ and $304 \mu\text{g}/\text{m}^3$. In percentage terms, Chinese stations underreport U.S. stations by between 18% (Shanghai) and 41% (Guangzhou) during these hours.

Together, the results of hypothesis #3 are consistent with strategic efforts to hide hours with unusually high pollution levels. Conversely, this finding is inconsistent with the 'clustered pattern' that would result if the impulses are generated by the poor performance of the instruments used to measure $PM_{2.5}$ and/or local weather.

3.4 Robustness checks: Testing the IIS methodology

The impulse indicator saturation technique identifies periods when government-controlled monitoring stations underreport $PM_{2.5}$ concentrations in a systematic, non-random fashion. These results could be caused by local Chinese officials misreporting high concentrations of $PM_{2.5}$ and/or biases in our methodology. To assess the degree which the IIS technique is prone to false positives (Type I errors) or false negatives (Type II errors), we apply the same methodology to samples in which we 'know' the 'correct' results.

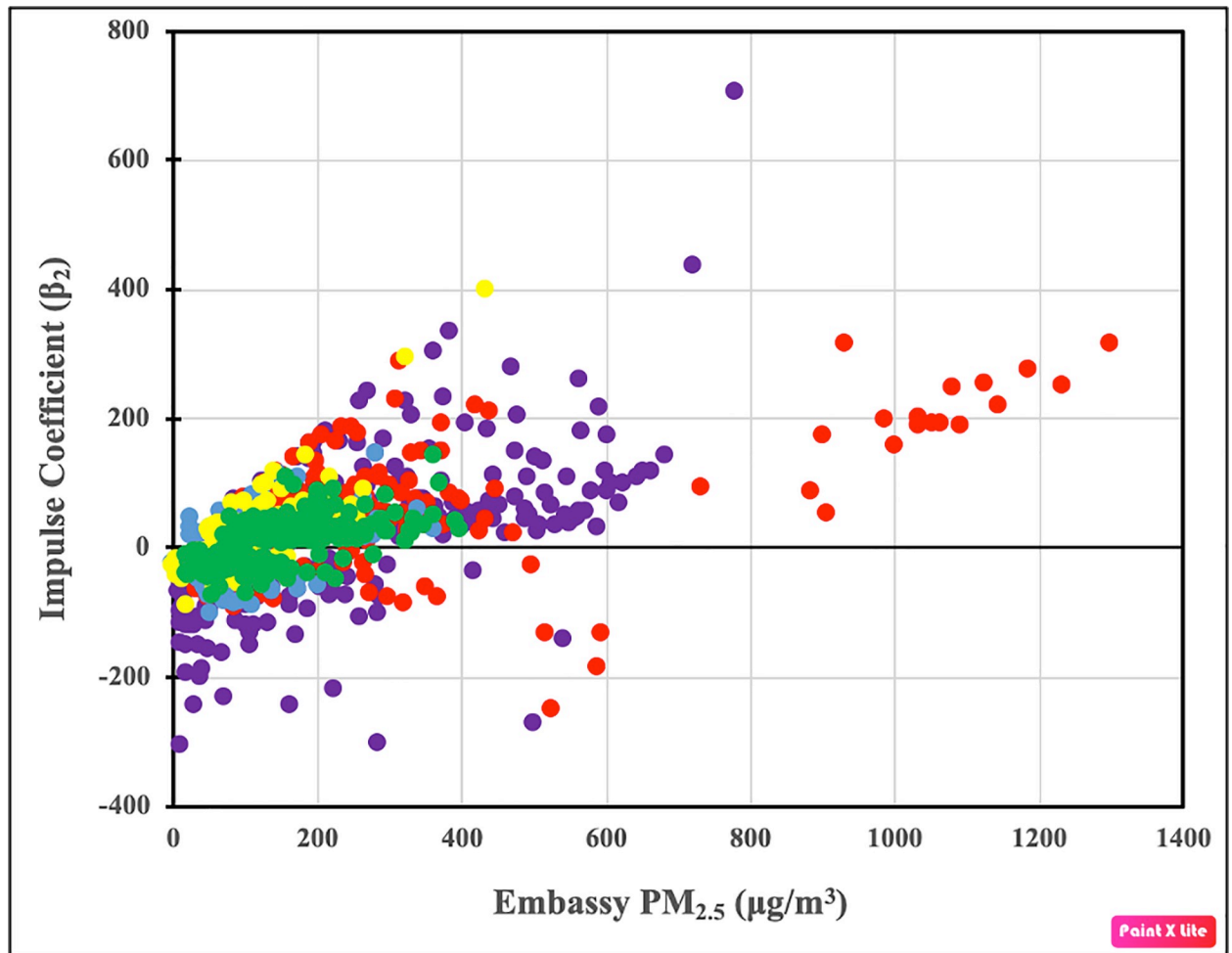


Fig 1. Relation between U.S. embassy-controlled PM_{2.5} concentrations and significant impulses. The coefficient values of all significant impulses were estimated using Eq (2) for Beijing (purple), Shenyang (red), Shanghai (blue), Guangzhou (yellow), and Chengdu (green). Statistical information about these relations is given in Table 3.

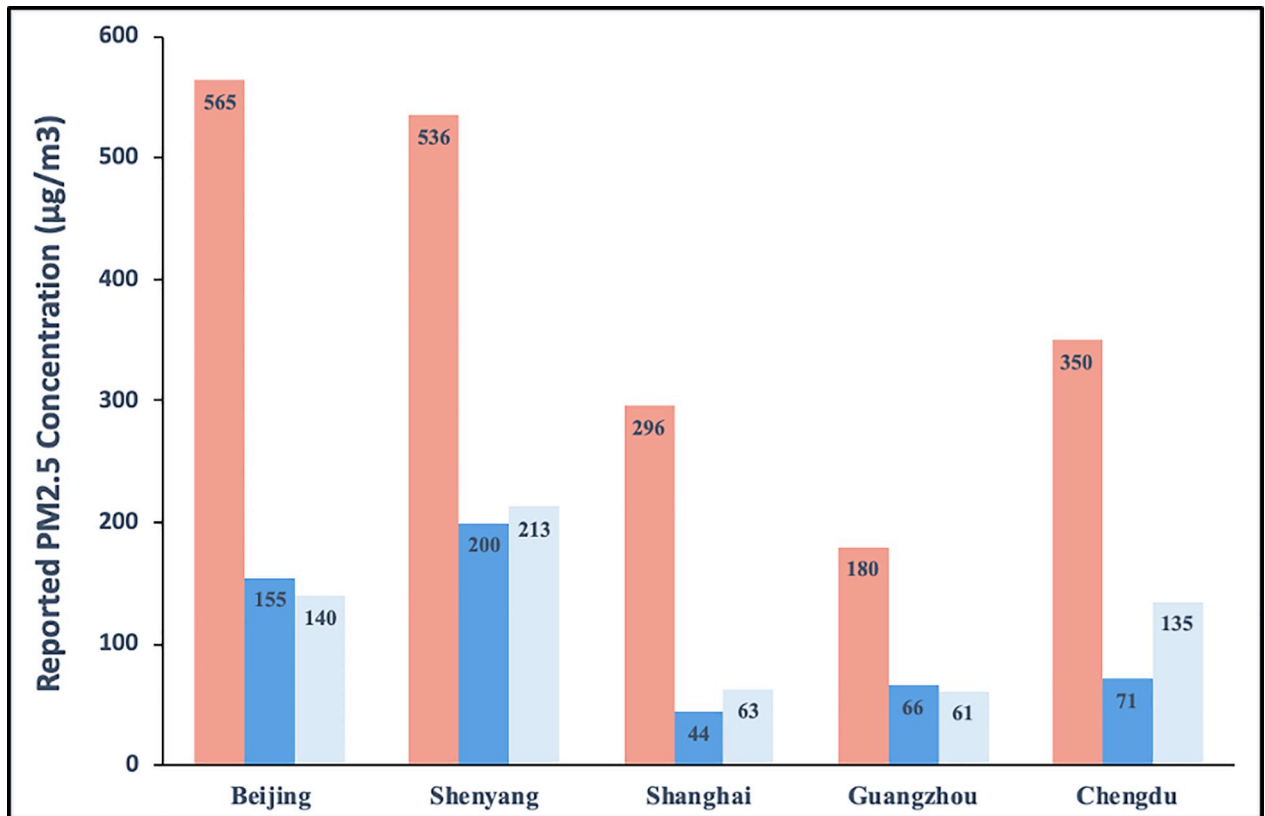
<https://doi.org/10.1371/journal.pone.0249063.g001>

To test for Type I errors, we analyze hourly measurements of air pollution from Taiwan. We assume that Taiwanese officials do not misreport pollution data because Taiwan exhibits greater levels of environmental transparency and because all monitoring stations are controlled directly by the central government. Under these conditions, our methodology should

Table 3. Test of null hypothesis #3 for the five Chinese cities and pooled model.

Location	Null Hypothesis #3					
	Fig 1 r ²	Logit Model (β ₃)	P-Value (β ₃ = 0)	50% Misreporting Threshold (µg/m ³)	# Hours > 50% Threshold	# Positive Impulses > 50% Threshold
Beijing	0.335	0.011	< 0.001**	565	41	19
Shenyang	0.142	0.01	< 0.001**	536	25	18
Shanghai	0.218	0.026	< 0.001**	296	2	2
Guangzhou	0.551	0.043	< 0.001**	180	59	33
Chengdu	0.332	0.021	< 0.001**	350	6	5
Pooled	0.289	0.012	< 0.001**	502	133	74

<https://doi.org/10.1371/journal.pone.0249063.t003>



Point X lite

Fig 2. The effect size of positive impulses above the 50% misreporting threshold for each of the five Chinese cities. For each of the five mainland Chinese cities in the dataset, Fig 2 shows the PM_{2.5} threshold at which the likelihood of misreporting exceeds 50% (red). It also shows the average coefficient values for impulses when concentrations at U.S. stations are above this threshold (dark blue), as well as the average difference between Chinese and U.S.-reported PM_{2.5} levels during these impulses (light blue).

<https://doi.org/10.1371/journal.pone.0249063.g002>

not detect large numbers of positive impulses when concentrations are high. If our methodology detects widespread evidence of data misreporting in Taiwan, this would suggest that our statistical methodology is prone to false positives.

Conversely, we test for Type II errors by analyzing data in cities where the Chinese MEE admits that air quality data were manipulated by local officials at known times and at known locations. For these times and locations, our methodology should detect large numbers of positive impulses when embassy-reported PM_{2.5} concentrations are high. If our methodology does not detect evidence of data misreporting when it is known to have occurred, this would suggest that our statistical methodology is prone to false negatives.

3.5 Null hypothesis #4: Measurement errors in Taiwan are caused by random chance

Because U.S. diplomatic outposts in Taiwan do not report air quality data, we test our methodology by choosing the five pairs of Taiwanese monitoring stations (out of 74 total stations) that are separated by the shortest geographical distance (average distance \approx 4.3km). Although we

Table 4. Analysis of stations in Taiwan with results generated by applying the IIS methodology to five station pairs.

Station Pairs	Location	Distance (Km)	Hourly Obs.	Expected Impulses ($p = 0.01$)	Observed Impulses	S_{prop} Statistic	P-Value
Renwu + Nanzi Districts	Kaohsiung City	4.8	20,180	202	180	-1.82	0.069
Qianjin + Zuoying Districts	Kaohsiung City	4.7	19,652	197	211	1.22	0.221
Qianzhen + Fuxing Districts	Kaohsiung City	0.7	20,165	202	196	-0.47	0.637
Taixi + Mailiao Townships	Yunlin County	8.3	20,023	200	219	1.57	0.116
Zhongming + Xitun Districts	Taichung City	3.2	21,888	219	231	0.97	0.332

<https://doi.org/10.1371/journal.pone.0249063.t004>

cannot assume that either of the two government-controlled monitoring stations represent ‘true’ pollution values, as long as both stations don’t misreport air quality data simultaneously and by the same magnitude, any misreporting at either station will still cause a significant impulse.

The null hypothesis of no misreporting implies that our methodology should identify impulses at a frequency approximately equal to the significance level of the criterion used to identify impulses ($p = 0.01$). Consistent with this notion, our results fail to reject the null hypothesis that the frequency of observed impulses is generated by random chance. None of the S_{prop} statistics in Table 4 are significant at a level ($p < 0.05$), which suggests that the number of impulses retained by the IIS technique is not significantly greater than the number that would be generated by random chance. This null hypothesis is nearly rejected by Pair 1 ($p = 0.069$), but in this case there were *fewer* impulses than expected by random chance, rather than more as we would expect with purposeful misreporting.

These results suggest that the methodology used to analyze concentrations of $PM_{2.5}$ in Chinese cities is not prone to Type 1 errors. As such, false positives probably do not cause us to (incorrectly) conclude that the Chinese government understates concentrations of $PM_{2.5}$ during periods of high concentrations. Furthermore, the failure to reject the null hypothesis using Taiwan data also suggests that errors/changes in instrumentation and/or local variations in weather are not responsible for the preponderance of impulses found in four of the five mainland Chinese cities analyzed.

3.6 Null hypothesis #5: Government misreporting causes the large number of positive impulses in Chinese cities

To test whether our methodology correctly identifies instances of data falsification in Chinese cities, we test for Type II errors by analyzing data from cities where local officials admit to underreporting measurements for $PM_{2.5}$. In both Xinyu City of Jiangxi Province and Xinyang City of Henan Province, the local EPB director hired individuals to falsify air quality measurements during September and October of 2017 (the exact dates and hours of tampering were not released by the MEE). According to the MEE, both instances of manipulation involved physical tampering with air quality monitors, including stuffing cotton yarn into sensors and spraying them with mist from ‘fog gun cars’ [46]. More importantly for the test of our methodology, officials manipulated only one monitoring station in each city, leaving all other stations unaffected.

For both Xinyu and Xinyang, we use the IIS methodology to compare hourly $PM_{2.5}$ measurements from the corrupted monitoring station to an average of hourly measurements from the surrounding, non-corrupted monitoring stations. To identify periods when data may be misreported, we break the full sample (January 2016 to December 2018) into eighteen two-month subsamples. If our methodology is not prone to Type II errors, the period when local officials misreported data (the September 2017 –October 2017 subsample) should contain

Table 5. Results for stations in China where local officials were caught misreporting data.

Subsample Time Period	Xinyu City (Jiangxi Province)						Xinyang City (Henan Province)					
	Hr. Obs.	Obs. Imp.	S _{prop} Stat.	Mean Imp. Coef. (β_2)	T-Stat: $\beta_2 = 0$	P-Val	Hr. Obs.	Obs. Imp.	S _{prop} Stat.	Mean Imp. Coef. (β_2)	T-Stat: $\beta_2 = 0$	P-Val
Jan 2016—Feb 2016	7,018	84	1.95	0.202	1.60	0.110	5,626	88	5.01**	0.433	4.84	< 0.001**
Mar 2016—Apr 2016	7,187	101	4.07**	0.313	3.76	0.001**	5,768	63	0.83	0.212	0.64	0.523
May 2016—Jun 2016	7,226	59	-1.85	-0.045	1.87	0.062	5,644	60	0.56	0.015	0.41	0.682
Jul 2016—Aug 2016	7,215	72	-0.02	0.084	0.01	0.994	5,803	67	1.39	0.133	1.12	0.264
Sep 2016—Oct 2016	7,170	65	-0.94	0.044	0.74	0.462	5,760	62	0.69	0.059	0.52	0.604
Nov 2016—Dec 2016	7,155	83	1.6	0.151	1.30	0.193	5,714	50	-1.12	-0.034	0.88	0.377
Jan 2017—Feb 2017	7,020	95	3.51**	0.292	2.88	0.004**	5,616	71	2.35*	0.248	1.98	0.048*
Mar 2017—Apr 2017	7,280	91	2.53*	0.214	2.07	0.038*	5,824	75	2.60**	0.316	2.15	0.032*
May 2017—Jun 2017	7,150	64	-1.05	-0.057	0.84	0.405	5,720	60	0.44	-0.111	0.51	0.610
Jul 2017—Aug 2017	7,235	87	2.04*	0.122	1.67	0.095	5,788	52	-0.92	-0.089	0.72	0.473
Sep 2017—Oct 2017	7,110	106	4.90**	0.276	4.31	< 0.001**	5,688	93	5.67**	0.49	5.21	< 0.001**
Nov 2017—Dec 2017	7,208	81	1.25	-0.119	1.00	0.319	5,624	68	1.86 ⁺	0.188	1.77	0.076
Jan 2018—Feb 2018	7,076	75	0.6	-0.061	0.45	0.655	5,585	55	-0.13	-0.12	0.05	0.962
Mar 2018—Apr 2018	7,250	68	-0.63	0.111	0.47	0.636	5,704	51	-0.95	-0.02	0.74	0.460
May 2018—Jun 2018	7,111	66	-0.72	0.09	0.55	0.583	5,716	48	-1.44	-0.106	1.15	0.249
Jul 2018—Aug 2018	7,198	63	-1.25	-0.025	1.01	0.315	5,775	62	0.66	0.189	0.50	0.620
Sep 2018—Oct 2018	7,220	86	1.92	0.137	1.61	0.108	5,604	61	0.78	0.105	0.63	0.529
Nov 2018—Dec 2018	7,153	75	0.49	-0.05	0.35	0.724	5,648	64	1.19	-0.083	0.94	0.348

In Table 5, the row containing the subsample period (September 2017 –October 2017) where data misreporting was known to have occurred is highlighted in yellow.

<https://doi.org/10.1371/journal.pone.0249063.t005>

more impulses than expected by random chance. It should also contain more positive than negative impulses.

Consistent with these predictions, test statistics reject the null hypothesis that the number of impulses is consistent with random chance for both cities in the September 2017 –October 2017 subsample (Table 5). Similarly, test statistics reject the null hypothesis that the mean value of impulses is equal to zero. Instead, the number of positive impulses is significantly higher than expected in both cities.

Furthermore, the test of our methodology expands our understanding of what happened in Xinyu and Xinyang. None of the subsamples after October 2017 reject either of the two null hypotheses at a $p < 0.05$ significance level, which suggests that local officials stopped misreporting data after being caught and punished by the central government. Conversely, both null

hypotheses are rejected in some of the subsamples before September 2017 (such as January 2015–February 2015 in Xinyang city and January 2016–February 2016 in Xinyu City). Together, these results suggest that local officials in Xinyu and Xinyang were misreporting air quality data well before the period in which they were caught. Finally, the small number of impulses observed in the periods after misreporting was detected also suggests that errors/changes in instrumentation and/or local variations in weather are not responsible for causing the high frequency of impulses present in our main dataset.

4. Discussion

Our results strongly suggest that some local Chinese officials continued to misreport measurements of $PM_{2.5}$ concentrations in many of the country's largest megacities, even after the government's post-2012 policy reforms. Consistent with our findings, in early 2018 the MEE announced that it had caught officials in seven cities manipulating data during the previous year [46]. Moreover, in April 2018, the central Ministry of Public Security (MPS) charged sixteen local officials in Linyi, Shanxi Province, with tampering with air quality monitors fifty-three times between April 2017 and March 2018 [47].

Our findings of ongoing air quality data misreporting in China are not surprising, because the government's post-2012 reforms did not eliminate incentives for local officials to cheat. Although requiring hourly, real-time measurements and abolishing the blue sky day metric eliminated manipulation around a given API threshold, local EPBs still face enormous pressure to report pollutant concentrations that decline continuously year-over-year. This pressure is compounded by the fact that the central government has increased penalties for local cadres in failing cities without also increasing the flow of centrally-backed resources or financial support [48]. Thus, faced with increasingly difficult attainment targets and a persistent lack of resources, some local officials have taken the path of least resistance by continuing to misreport air quality data.

Nonetheless, the persistence of local data misreporting does not invalidate results which suggest urban air quality in China has improved in recent years [12–16]. Even measurements from U.S. embassies and consulates show that annual concentrations of $PM_{2.5}$ fell by more than 25% between 2013 and 2017. Although these broader trends are clear, day-to-day air quality numbers remain highly suspect, especially on high pollution days. The fact that air pollution data is less likely to be accurate on highly polluted days is of particular importance, because acute health effects appear to be more strongly related to hourly peak concentrations than daily averages [49]. Moreover, even though nationwide concentrations of certain pollutants are likely decreasing, local data misreporting makes it difficult for central officials to determine which cities are driving these improvements and which are free-riding.

Central leaders are aware of this problem, but until recently their policy responses have been slow and ineffectual. Starting in late 2016, the central government instituted a series of new reforms aimed at improving environmental governance. Specifically, Beijing announced a transition towards a more centralized environmental bureaucracy that seeks to reduce the negative impacts of 'local protectionism' in environmental management [50]. Also, in September 2016, the power to nominate city-level EPB directors was transferred from city governments to provincial EPBs (although city approval is still required to confirm nominees). Also, at the end of 2020, provincial EPBs (rather than city EPBs) assumed full responsibility for funding and personnel decisions at local monitoring stations [51]. The central government also is increasing its supervision and oversight of local officials, and in July 2017, the MEE began surprise environmental inspections in the Beijing-Tianjin-Hebei region [52]. The party-controlled *People's Daily* announced that these inspections would become the 'new normal,' and by the end

of 2017, central agencies had disciplined nearly 12,000 local officials and assessed more than \$130 million in fines [53]. These MEE inspections have been carried out in tandem with the Central Organization Department (COD) and the Central Commission for Discipline Inspection (CCDI), which are the two most important groups in determining the promotion prospects of local officials. Together, these most recent reforms may represent the beginning of an effective strategy to combat misreporting of local air quality. However, we cannot fully evaluate these reforms because they do not begin until late-2016 and 2017, and our sample only runs through June 2017. Under these conditions, further research is needed to determine whether these latest reforms have been more successful in curtailing the misreporting of local pollution data.

Overall, China's post-2012 environmental policy reforms improved the availability of air quality data and helped to reduce pollutant concentrations in urban areas, but they did not dissuade some local officials from misreporting data. Although previous studies erroneously suggest that data misreporting ended after 2012, this paper indicates that it continued; just in different, more difficult-to-detect forms. As the central government continues to implement new efforts to address data fraud, it remains to be seen whether these efforts will be effective or whether they will simply induce local officials to find newer and ever-more-innovative ways to falsify official statistics.

Supporting information

S1 File.
(DOCX)

Acknowledgments

We acknowledge Felix Prentis for his help in designing and implementing the impulse indicator saturation model used in this paper and Jeff Geddes for suggesting the need for tests of Type I and II errors.

Author Contributions

Conceptualization: Jesse S. Turiel, Robert K. Kaufmann.

Data curation: Jesse S. Turiel.

Formal analysis: Jesse S. Turiel, Robert K. Kaufmann.

Methodology: Jesse S. Turiel.

Supervision: Robert K. Kaufmann.

Validation: Jesse S. Turiel.

Writing – original draft: Jesse S. Turiel.

Writing – review & editing: Jesse S. Turiel, Robert K. Kaufmann.

References

1. Rohde RA, Muller RA. Air pollution in China: Mapping of concentrations and sources. *PLoS One*. 2015; 10: 1–20. <https://doi.org/10.1371/journal.pone.0135749> PMID: 26291610
2. Guo X. Cost of pollution in China: economic estimates of physical damages. World Bank. 2007; 39236: 1–151. Available from: <http://edgar.jrc.ec.europa.eu/overview.php?v=CO2ts1990-2015>
3. Watson R, Pauly D. Systematic distortions in world fisheries catch trends. *Nature*. 2001; 414: 534–536 (2001). <https://doi.org/10.1038/35107050> PMID: 11734851

4. Nielsen CP, Ho MS. Clearer skies over China: Reconciling air quality, climate, and economic goals. Cambridge, MA: MIT Press; 2013.
5. Wallace JL. Juking the stats? Authoritarian information problems in China. *Br J Polit Sci.* 2016; 46: 11–29.
6. Guan D, Liu Z, Geng Y, Lindner S, Hubacek K. The gigatonne gap in China's carbon dioxide inventories. *Nat Clim Chang.* 2012; 2: 672–675.
7. Seto KC, Kaufmann RK, Woodcock CE. Agricultural land conversion in southern China. *Nature.* 2000; 406: 121–125.
8. Buckley C. China's carbon dioxide emissions may have been overstated by more than 10%. *The New York Times.* 2015 Aug 20 [cited 2019 Feb 12]. Available from: <https://www.nytimes.com/2015/08/20/world/asia/china-carbon-dioxide-emissions-may-be-overstated-nature-study.html>
9. Montzka SA, Dutton GS, Yu P, Ray E, Portmann RW, Daniel JS, et al. An unexpected and persistent increase in global emissions of ozone-depleting CFC-11. *Nature.* 2018; 557: 413–417. <https://doi.org/10.1038/s41586-018-0106-2> PMID: 29769666
10. Lunt MF, Park S, Li S, Henne S, Manning AJ, Ganesan AL, et al. Continued emissions of the ozone-depleting substance carbon tetrachloride from eastern Asia. *Geophys Res Lett.* 2018; 45: 11–423. <https://doi.org/10.1029/2018gl079500> PMID: 33005064
11. Rigby MS, Saito T, Western LM, Redington AL, Fang X, Henne S. Increase in CFC-11 emissions from eastern China based on atmospheric observations. *Nature.* 2019; 569: 546–550. <https://doi.org/10.1038/s41586-019-1193-4> PMID: 31118523
12. Zhang S, Guo B, Dong A, He J, Xu Z, Chen SX. Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences.* 2017; 473: 1–17. <https://doi.org/10.1098/rspa.2017.0457> PMID: 28989318
13. Silver B, Reddington CL, Arnold SR, Spracklen DV. Substantial changes in air pollution across China during 2015–2017. *Environ Res Lett.* 2018; 13: 114012.
14. 2017 report on the state of the ecology and environment in China. Ministry of Ecology and Environment. 2018 Aug 1 [Cited 2019 Feb 12]. Available from: <http://english.mee.gov.cn/Resources/Reports/soe/SOEE2017/201808/P020180801597738742758.pdf>
15. Ma Z, Liu R, Liu Y, Bi J. Effects of air pollution control policies on PM_{2.5} pollution improvement in China from 2005 to 2017: A satellite-based perspective. *Atmos Chem and Physics.* 2019; 19: 6861–6877.
16. Zhang L, Zhang M, Yao Y. Multi-time scale analysis of regional aerosol optical depth changes in national-level urban agglomerations in China using modis collection 6.1 datasets from 2001 to 2017. *Remote Sensing.* 2019; 11: 201–218.
17. Zhang P, Wu J. Impact of mandatory targets on PM_{2.5} concentration control in Chinese cities. *J Cleaner Prod.* 2018; 197: 323–331.
18. Turiel JS, Ding I, Liu J. Environmental governance in China. *Brill Research Perspectives: Governance and Public Policy in China.* 2017; 1: 1–76.
19. Lieberthal K. The fragmented authoritarianism model and its limitations. In: Lieberthal K, Lampton D, editors. *Bureaucracy, politics, and decision making in post Mao China.* Berkeley: Berkeley University Press; 1992. pp. 1–30. <https://doi.org/10.3109/10731199209119698> PMID: 1391492
20. Sinkule B, Ortolano L. *Implementing environmental policy in China.* Westport: Praeger; 1995.
21. Zhou X, Lian H, Ortolano L, Ye Y. A behavioral model of 'muddling through' in the Chinese bureaucracy: The case of environmental protection. *China Journal.* 2013; 70: 120–147.
22. Minzner CF. Riots and cover-ups: Counterproductive control of local agents in China. *U Pa J Int'l L.* 2009; 31: 53–123.
23. Tsai L. Understanding the falsification of village income statistics. *The China Quarterly.* 2008; 196: 805–826.
24. You M. Addition of PM_{2.5} into the National Ambient Air Quality Standards of China and the contribution to air pollution control: The case study of Wuhan, China. *The Sci World J.* 2014; 1: 1–10. <https://doi.org/10.1155/2014/768405> PMID: 24982994
25. Chen Y, Jin GZ, Kumar N, Shi G. Gaming in air pollution data? Lessons from China. *B E J Econ Analysis & Pol.* 2012; 12: 1–42.
26. Andrews SQ. Seeing through the smog: Understanding the limits of Chinese air pollution reporting. *China Environment Series.* 2008; 10: 5–29.
27. Ghanem D, Zhang J. Effortless perfection: Do Chinese cities manipulate air pollution data? *J Environ Econ Manage.* 2014; 68: 203–225.

28. Zhang Q, Qiang Y, Zheng Y, Tong D, Shao M, Wang S. Drivers of improved PM_{2.5} air quality in China from 2013 to 2017. *Proc Natl Acad Sci USA*. 2019; 116: 24463–24469. <https://doi.org/10.1073/pnas.1907956116> PMID: 31740599
29. Kostka G. Barriers to the implementation of environmental policies at the local level in China. *World Bank Policy Research Working Paper*. 2014; 7016: 1–45.
30. Hsu A, Yan C, Cheng Y. Addressing gaps in China's environmental data: The existing landscape. *Yale Data-Driven*. 2017; 1: 1–38.
31. Stoerk T. Statistical corruption in Beijing's air quality data has likely ended in 2012. *Atmos Environ*. 2016; 127: 365–371.
32. Liang X, Li S, Zhang S, Huang H, Chen SX. PM_{2.5} data reliability, consistency, and air quality assessment in five Chinese cities. *J Geophys Res Atmos*. 2016; 121: 10–22.
33. Beijing Sinapp. Hourly Chinese air quality; 2018 [Cited 2019 Jan 3]. Database: figshare [Internet]. Available from: <https://beijingair.sinaapp.com>
34. US Embassy and Consulates in China. Hourly air quality data; 2018 [Cited Jan 5]. Database: figshare [Internet]. Available from: <https://china.usembassy-china.org.cn>
35. Taiwan Air Quality Monitoring Network. Hourly air quality data; 2018 [Cited 17 Jan 2019]. Database: figshare [Internet]. Available from: <https://taqm.epa.gov.tw/taqm/en/>
36. McCarry J. Manipulation of the running variable in the regression discontinuity design: A density test. *J Econometrics*. 2008; 142: 698–714.
37. Bai J, Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica*. 1998; 1: 47–78.
38. Castle JL, Doornik JA, Hendry DF, Pretis F. Detecting location shifts during model selection by step-indicator saturation. *Econometrics*. 2015; 3: 240–264.
39. Marczak M, Proietti T. Outlier detection in structural time series models: The indicator saturation approach. *Intl J Forecasting*. 2016; 32: 180–202.
40. Johansen S., Nielsen B. Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics*. 2016; 43: 321–348.
41. Pretis F, Mann ML, Kaufmann RK. Testing competing models of the temperature hiatus: Assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change*. 2015; 131: 705–718.
42. Kaufmann RK, Connelly C. Oil price regimes and their role in price diversions from market fundamentals. *Nat Energy*. 2020; 5: 141–149.
43. Santos C, Hendry DF, Johansen S. Automatic selection of indicators in a fully saturated regression. *Comput Stat*. 2008; 23: 317–335.
44. Johansen S., Nielsen B. An analysis of the indicator saturation estimator as a robust regression estimator. *Castle, and Shephard (2009)* 1, 1–36 (2009).
45. Jiao X, Pretis F. Testing the presence of outliers in regression models. SSRN. 2018. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3217213
46. Ministry of the Environment: Linyi environmental monitoring fraud case has a bad impact. *Beijing News*. 2018 Jun 24 [Cited 2019 Jan 25]. Available from: <http://www.bjnews.com.cn/news/2018/06/24/492340.html>
47. How was the fraud of Linyi environmental monitoring data discovered? Ministry of Ecology and Environment. 2018 Jun 23 [Cited 2019 Jan 26]. Available from: http://www.mee.gov.cn/gkml/sthjbgw/qt/201806/t20180623_443636.htm
48. Economy E. Environmental governance in China: State control to crisis management. *Daedalus*. 2014; 143: 184–197.
49. Lin H, Ratnapradipa K, Wang X, Zhang Y, Xu Y, Yao Z, et al. Hourly peak concentration measuring the PM_{2.5}-mortality association: Results from six cities in the Pearl River Delta study. *Atmos Environ*. 2017; 16: 27–33.
50. Brombal D. Accuracy of environmental monitoring in China: Exploring the influence of institutional, political and ideological factors. *Sustainability*. 2017; 9: 324–341.
51. Vertical management reform of grassroots environmental protection organizations. *Xinhuanet*. 2016 Sep 23 [Cited 2018 Dec 29]. Available from: http://news.xinhuanet.com/legal/2016-09/23/c_129294953.htm
52. Fuchs D, Schmitt E. How China's environmental crackdown is affecting business owners and workers: The case of Chengdu. *Chinoiresie*. 2017 Nov 2 [Cited 2019 Apr 10]. Available from: <https://www.chinoiresie.info/how-china-environmental-crackdown-is-affecting-business-owners-and-workers-the-case-of-chengdu>

53. Gan N. 12,000 officials disciplined and 18,000 companies punished in China's sweeping crackdown against pollution. South China Morning Post. 2017 Sep 2 [Cited 2019 Apr 10]. Available from: <https://www.scmp.com/news/china/policies-politics/article/2109342/top-level-china-pollution-inspections-wrapping>