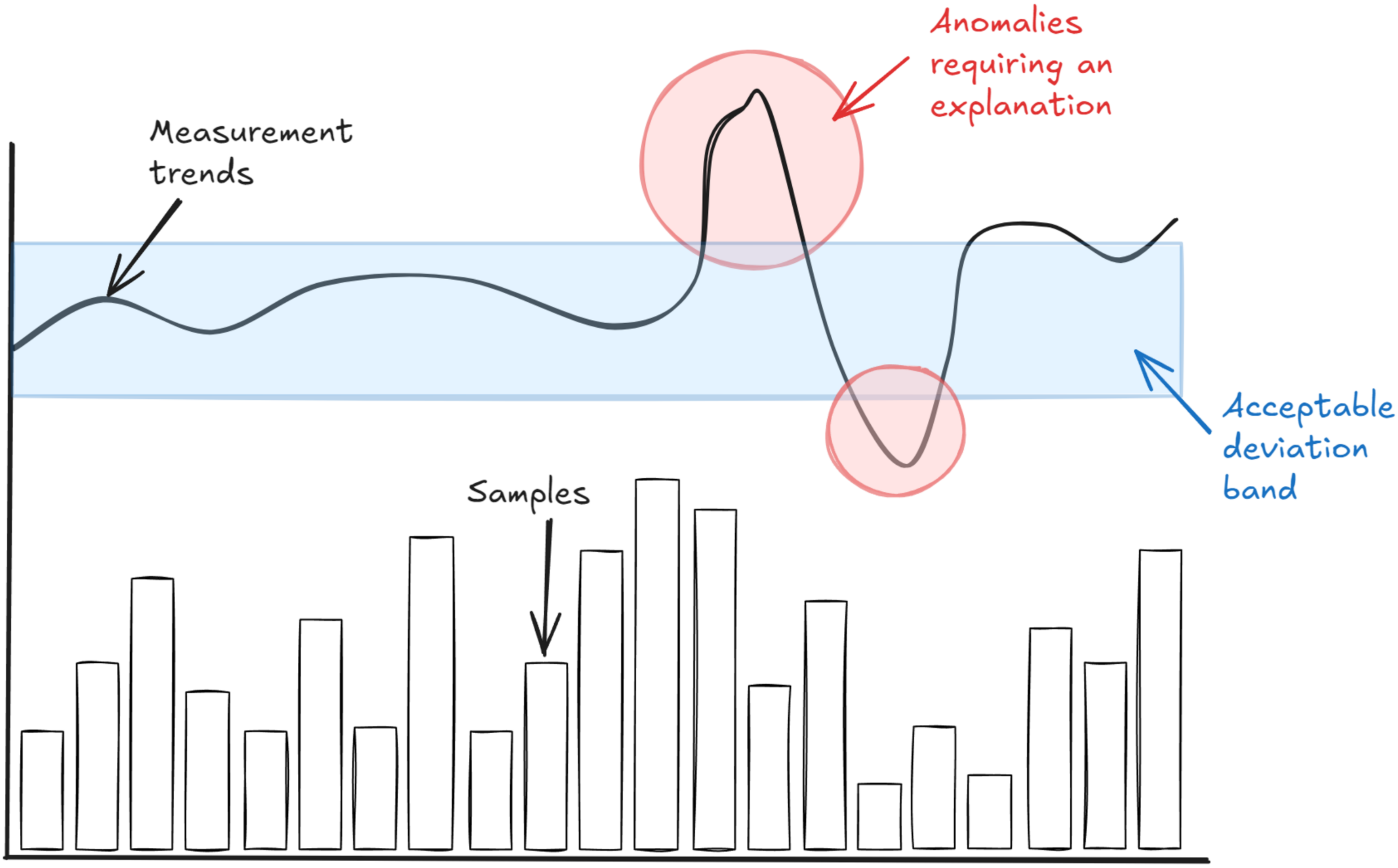
HOW TO SPOT ANAMOLIES IN DATA TRENDS



Evaluating AQI data from Indian cities

Sai Krishna Dammalapti
Sarath Guttikunda
Special credit: Karn Vohra

SIM-air Working Paper Series #49-2024

URBAN emissions .info

UrbanEmissions (UEinfo) was founded in 2007 with the vision to be a repository of information, research, and analysis related to air pollution. UEinfo has four objectives: (1) sharing knowledge on air pollution (2) science-based air quality analysis (3) advocacy and awareness raising on air quality management and (4) building partnerships among local, national, and international airheads.

All our publications are accessible @ www.urbanemissions.info/publications

Air Quality Index (AQI) data from Indian cities, utilized in this study, is available (open-access via github) as part of SIM-Series working paper #47-2024

Special credit to Dr Karn Vohra, University College London, for suggesting Benford's Law during our poster discussions at IGAC 2024.

Send your questions and comments to simair@urbanemissions.info

Short Story

Year-on-year and day-to-day deviations are natural in air quality data. The goal of the proposed 2-step statistical investigation into data trends is only to identify deviations and seek justification. We used Benford's law and year-on-year non-overlapping area calculations to identify deviations.

Key Messages

At all-India scale for years 2018 to 2023, we interpret that Benford's law provides no material to suspect misreporting, and the K-S test failed in all the instances because the sample size is different every year with 80% of the cities running only one station in 2023.

Delhi, with a consistent and large network of 35+ stations for all years, provided the only benchmarking case study for this exercise.

All other cities reported data from a mix of stations over the years. While year-on-year comparisons are presented in this paper, the inconsistency in the reporting patterns is a flaw in the individual city data.

At the city scale, deviations were evident. Deviation from Benford's law only means that additional information is necessary to explain the change in the city's data patterns. In Indian cities, we suspect "small sample sizes" and "heterogeneity among the locations" as the main reason to explain the deviations.

Cities with 15+ monitors never deviated from Benford's law. In other words, if there is a violation, it is often from a city with a small monitoring network (under 4 stations).

Cities with denser and consistent monitoring networks deviated less from Benford's law. Even if they deviated, their annual distributions are consistent and thus predictable. Chennai was a good example, where Benford's law failed, but the second step clears it of any deviations.

AQI data with good statistical confidence is more useful for policy work. In other words, for good policy support, all efforts to increase the city monitoring networks to operate 10-15+ stations must be encouraged at all costs.

An ambient monitoring network in a city requires a minimum of 4-5 stations to truly represent the spatial and temporal trends of emission intensities in an urban airshed. These locations must include representation from residential, commercial, industrial, traffic, and background activities.

Forward applications of These Methods

Any portal collating data can apply this 2-step method on day-to-day data at the station or at the city level (with reasonable network size) and flag any sudden changes. E.g., data from Friday the 13th looks different from the 12th, why?

A researcher (or any of the sensor network groups) can check on the data streams for deviations and flag it for further scrutiny. E.g., data from Sept-5th is looking odd, is the instrument working okay or is it because its Sunday?

Same can be applied for comparing seasons over years. If the deviations are not significant, it is possible that energy and emission patterns are business as usual, which can be used for baseline judgements.

The methods can be extended to other sectors. For example, electricity generation and transmission rates to flag the surges in supply and demand. A year-on-year comparison can reveal the changes in consumption patterns. Similar inferences can be made on data from fuel sales at a fuel station or fuel sales in a city – is the consumption distribution reflecting the push for electric vehicles.

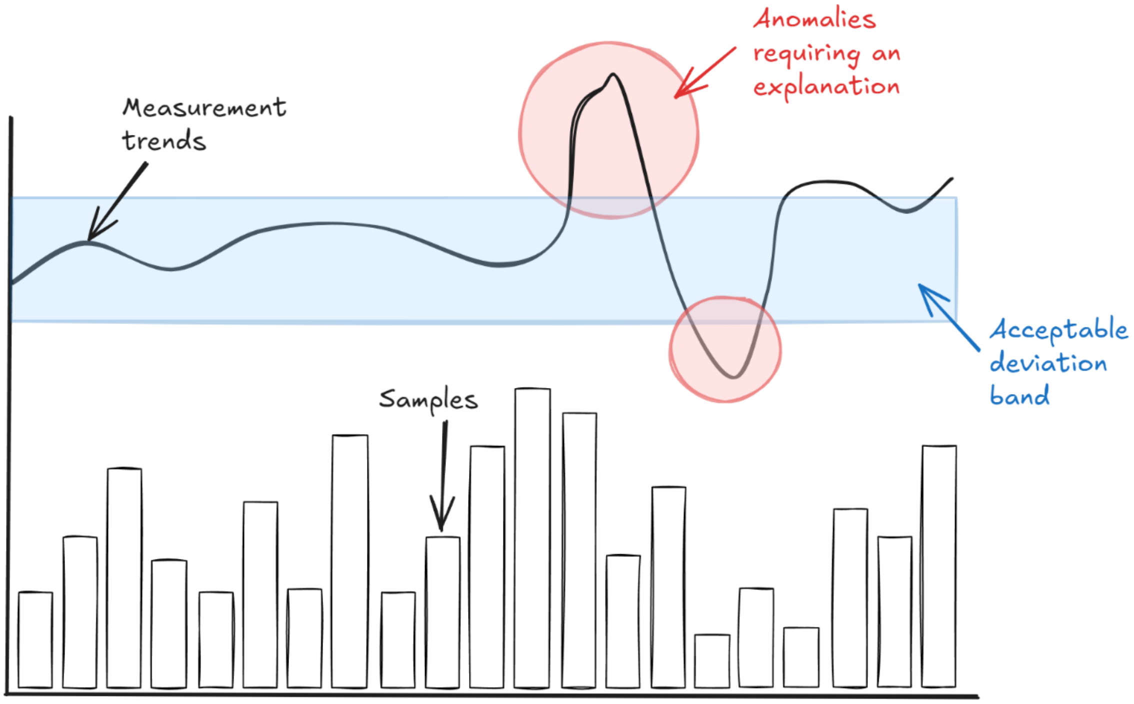
1. Problem Statement

Is it possible to detect anomalies (deviations) in the data patterns from average air quality values or from an instrument on a day-to-day basis or from a cluster of instruments operating in an area?

Sometimes, there is a suspicion that air quality numbers are misreported, or an instrument is misbehaving or one of the many instruments in a cluster is out of order. This can be because of technical reasons (monitoring instrument malfunctioning or operational error), statistical reasons (selection bias of monitoring locations), or political reasons (selection bias on data openness).

While a thorough investigation is needed to resolve these suspicions, a statistical investigation can be conducted to identify potential anomalies or at least highlight data sections where the trends are not followed, which require additional scrutiny.

Typically, most of the data analysis among the air quality monitoring community to understand the patterns, identify the deviations, and flag the extreme unknowns, happens after the field experiments are completed or after a certain volume of data is collected. If the identification can be conducted in advance or in real-time, these deviations can be used to flag instances for scrutiny, and help provide better explanations (in later presentations or in writing articles)



Two key inferences to be aware of:

1. It is natural for variations to occur, especially when it comes to dynamic air quality. No two days, no January over two years, no winter over two years, nor two sets of annual averages will ever show the same distribution of values.
2. Not all anomalies can be interpreted as misreporting or misbehaving. Some of them could be genuine and the statistical investigation is only a means to identify these sudden changes and allow us to seek the right answers for deviating from the normal. For example: comparing data from two days, the deviation can be due to meteorology (like more rain on one day) or a shutdown of an activity (like traffic or industries, as witnessed during the COVID19 lockdown periods).

In this working paper, we are presenting a 2-step method to identify anomalies in data trends – Benford's law with Euclidian distance and two sample Kolmogorov-Smirnov (K-S) test with non-overlapping area. The methods are applied over India's daily average air quality index (AQI) dataset for years 2018 to 2023.

The goal of these investigative methods is only to identify the unexplainable statistical deviations and seek justification, if there are any. The methods do not answer the question of why the deviations occurred.

2. Data Source

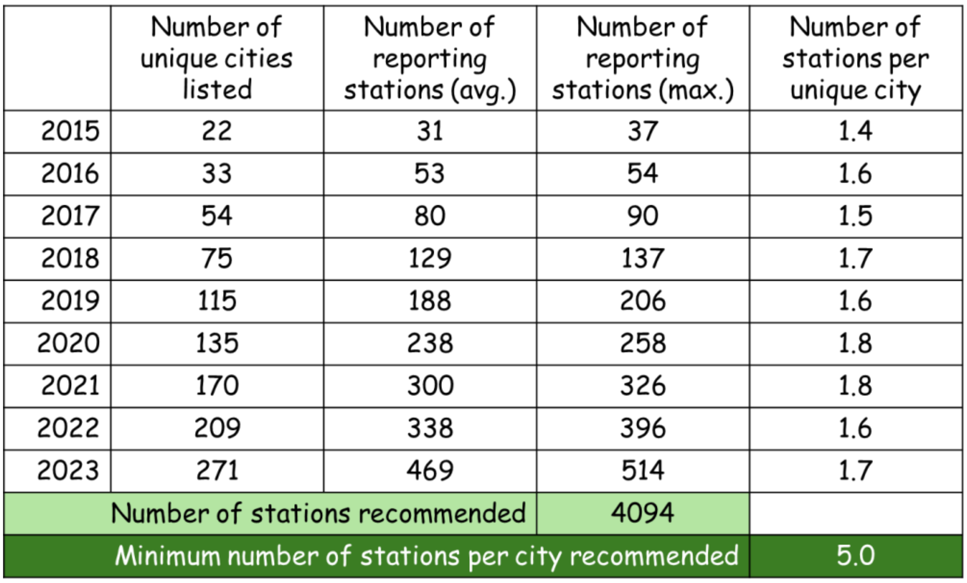
Statistical and uncertainty analysis presented in this working paper is based on air quality index (AQI) data extracted from the official daily AQI bulletins issued by the Central Pollution Control Board (CPCB), New Delhi, India, between 2015 and 2023[1].

Air Quality Index (AQI) is an important tool for communicating the quality of air pollution as health-related alerts. AQI unifies all this complicated science of pollution composition, exposure rates-based health severity, ambient standards, measurements, and standard protocols, into simple colour coded bins for everyone to see how good or bad or severe the pollution levels are[2].

AQI calculations is often based on the ambient monitoring data for 6 pollutants – particulates (as $PM_{2.5}$ and PM_{10} size fractions), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone.

Key messages from India's AQI bulletins

Between 2015 and 2023 (a) the number of unique cities increased 12-fold from 22 to 271 (b) the average number of reporting stations increased 15-fold from 31 to 469 (c) and the average number of stations per unique city increased from 1.4 to 1.7– an overall 20% increase.

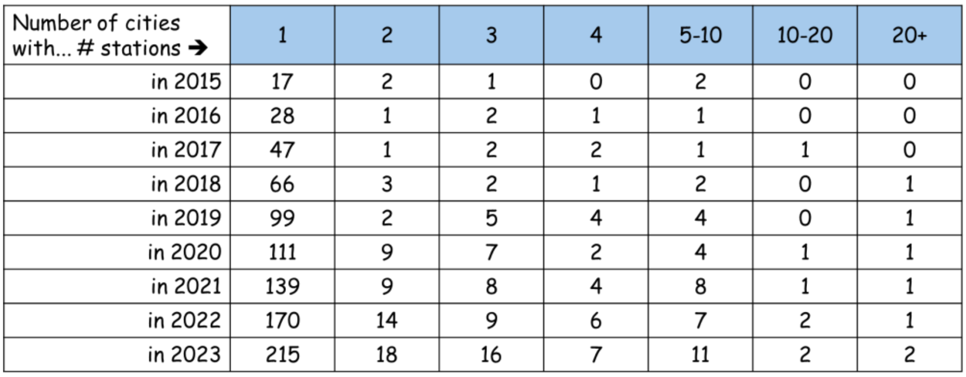


	Number of unique cities listed	Number of reporting stations (avg.)	Number of reporting stations (max.)	Number of stations per unique city
2015	22	31	37	1.4
2016	33	53	54	1.6
2017	54	80	90	1.5
2018	75	129	137	1.7
2019	115	188	206	1.6
2020	135	238	258	1.8
2021	170	300	326	1.8
2022	209	338	396	1.6
2023	271	469	514	1.7
Number of stations recommended			4094	
Minimum number of stations per city recommended				5.0

[1] A cleaned database of AQI data from all Indian cities, some statistical analysis, and visualizations were released as SIM-air Working Paper Series # 47-2024 @ https://urbanemissions.info and a library of python scripts used to tabulate the data from PDF bulletins is available @ www.github.com/urbanemissions

[2] An example AQI calculator comparing approved methodologies from six countries and two instructional videos is available @ https://urbanemissions.info/tools

While the number of cities and overall monitoring capacity increased between 2015 and 2023, 80% (215 out of 271) of the cities had only one monitoring station and 92% (249 out of 271) had three or less monitoring stations.



Number of cities with... # stations ➔	1	2	3	4	5-10	10-20	20+
in 2015	17	2	1	0	2	0	0
in 2016	28	1	2	1	1	0	0
in 2017	47	1	2	2	1	1	0
in 2018	66	3	2	1	2	0	1
in 2019	99	2	5	4	4	0	1
in 2020	111	9	7	2	4	1	1
in 2021	139	9	8	4	8	1	1
in 2022	170	14	9	6	7	2	1
in 2023	215	18	16	7	11	2	2

In 2023, only metropolitan and some Tier-1 cities, reported data from more than five (5) monitoring stations – which is a representative sample size for any city.

These 15 cities are – Agra (6), Ahmedabad (9), Bengaluru (13), Chennai (8), Delhi (39), Hyderabad (14), Jaipur (6), Jodhpur (5), Kolkata (7), Lucknow (6), Moradabad (6), Mumbai (28), Navi Mumbai (7), Patna (6), and Pune (8).

CPCB guidelines suggests a minimum of four (4)

CPCB approved the following guidelines[3] to calculate the minimum number of monitoring stations required to operate in an airshed, based on airshed's population and commercial density. The guideline for particulate pollution monitoring start with a minimum of four (4) stations for any airshed. Similar guidelines exist for gaseous pollutants – SO_2, NO_2, CO and Ozone.

```
Thumb rule: India defined the following
based on total population (TP) for PM monitoring.

For TP under 100,000              - 4 units
For TP between 100k and 1 million - 4 + 0.6 per 100,000
For TP between 1 to 5 million      - 7.5 + 0.25 per 100,000
For TP above 5 million            - 12 + 0.16 per 100,000

Finally, the city's financial, personnel, and operational
capacity decides "how many" monitors to install and
"where to" monitor in the city.
```

[3] "Guidelines for ambient air quality monitoring", by the Central Pollution Control Board (CPCB), New Delhi, India, April-2003. Full document is available @ https://urbanemissions.info (under resources)

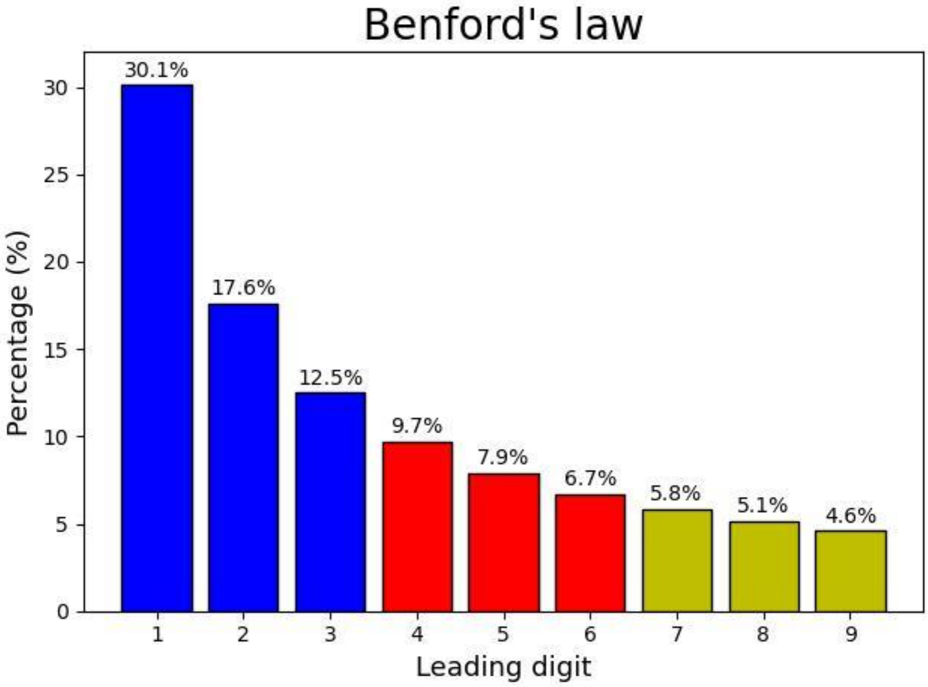
3. Methods

Benford’s Law

Researchers have leveraged Benford’s law in identifying anomalies. Relevant to this study, one such application was for the quality of official air quality numbers reported in Beijing, China[4]. Other applications include an instance flagging likely misreporting in the fields of accounting[5], economics[6]. and with CO2 emissions data[7].

According to Benford’s law, also called First Digit Law, the distribution of the leading digit of any naturally occurring data is a logarithmic distribution, governed by the equation (see the illustration below):

$$Benford\ Frequency(i) \ = \ log_{10}(1 \ + \ \frac{1}{i})\ where\ i \ = \ 1,2,3...9$$



[4] “Statistical corruption in Beijing's air quality data has likely ended in 2012” (2016) @ https://www.sciencedirect.com/science/article/abs/pii/S1352231015306336
“An Investigation of the Quality of Air Data in Beijing” University of Berkeley (2014) @ https://are.berkeley.edu/~sberto/BeijingJuly16.pdf

[5] “I've Got Your Number - How a mathematical phenomenon can help CPAs uncover fraud and other irregularities” (1999) @ https://www.journalofaccountancy.com/issues/1999/may/nigrini.html

[6] “Do countries falsify economic data strategically? Some evidence that they might” (2013) @ https://shs.hal.science/halshs-00482106v3

[7] “Testing CO2 Emissions Data During Covid-19 Pandemic Using Benford’s Law” (2023) @ https://erl.scholasticahq.com/article/38783-testing-co-2-emissions-data-during-covid-19-pandemic-using-benford-s-law

As per this distribution, the smaller digits would have higher probability of occurring than the larger digits. Deviation from Benford's law is considered as an indication of misreporting, even fraud in some cases.

To quantify if the observed distribution of first digits conforms with Benford's law, we use Euclidean Distance (ED) as a metric[8] defined in the following way:

$$ED = \sqrt{\sum_{i=1}^{9} (Observed\ Frequency_i - Benford\ Frequency_i)^2}$$

ED is not a formal statistical test like χ^2 (chi-square) or Kolmogorov-Smirnov (K-S) which are used most often to check if the observed frequency matches with an expected frequency[9]. While these two formal tests are sample size sensitive, ED is not sensitive to the sample size as it only works with the proportions.

ED is a relative metric bound between 0 (when the observed frequency is exactly the Benford frequency) and 1.036 (when all leading digits observed are 9). Hence, if ED is closer to 0.0, there is closer confirmation to the Benford's frequency. According to Goodman's rule of thumb, ED < 0.25 would be considered as confirmation to Benford's frequency and not otherwise.

However, there are reported issues about the use of Goodman's rule of thumb, especially when a small frequency is observed for any one digit[10]. Hence, we also present the confidence intervals of the observed frequencies calculated by non-parametric bootstrapping.

Deviation from Benford's law is only an indicator and not a foolproof law. It requires that data has a wide range. Datasets like student's marks out of 100 will always deviate from this law. To further investigate the deviations, we implement a normalized form of two sample Kolmogorov-Smirnov (K-S) test and calculate non-overlapping areas of probability distribution functions (PDFs) as a follow-up test.

[8] "Benford's law in the Gaia universe" (2020)
@ https://www.aanda.org/articles/aa/full_html/2020/10/aa37256-19/aa37256-19.html

[9] Chi-square and K-S tests are powerful at high sample sizes and as a result would reject the Null Hypothesis (Ho) most of the time, if there is a mismatch between the observed and the expected frequencies.

[10] "Testing Benford's law: from small to very large datasets" (2022)
@ https://doi.org/10.37830/SJS.2022.1.03

Two sample Kolmogorov-Smirnov (K-S) test

Given the counter-intuitive nature of Benford's law, we also investigate the raw data, instead of limiting ourselves to the first digit. This is because Indian air quality data may not always fit well with Benford's law, given the low instances of small AQI values in Indian cities[11].

The two sample K-S test is generally used to investigate if two samples come from the same distribution of data. Empirical cumulative distribution functions (CDFs) are generated for both the samples and the maximum vertical distance between them is calculated. If this distance is above a critical value, then it is considered that the two samples come from different distributions. It is a non-parametric test and hence it is applicable for any population distribution, where the sample is randomly picked from.

However, the K-S test is generally sensitive to both location and shape of the distribution. For example, consider two samples of size 1000 each picked from two different normal distributions of different means and standard deviations. The K-S test would indicate that these samples belong to different population distributions.

Sample 1	**Sample 2**	**K-S statistic (p-value)**	**Remark** Fail to reject = Both samples may not be from different distributions. Reject = Both samples are from different distributions
~N (100,10)	~N (100,10)	0.044 (0.29)	Fail to reject the null hypothesis.
~N (100,10)	~N (10,10)	1.00 (0.00)	Reject the null hypothesis.
~N (100,10)	~N (100,100)	0.121 (0.01)	Reject the null hypothesis
~N (100,10)	~N (10,100)	0.725 (0.00)	Reject the null hypothesis
~logN(0,0.1)	~logN(0,0.1)	0.026 (0.89)	Fail to reject the null hypothesis
~logN(0,0.1)	~logN(0.3,0.1)	0.862 (0.00)	Reject the null hypothesis
~logN(0,0.1)	~logN(0,0.3)	0.248 (0.00)	Reject the null hypothesis

This sensitivity of the K-S test does not interest our investigation. We expect that air quality data will change in its central tendencies and shape over years, and we don’t want to flag these natural (expected) deviations as anomalies. However, we want to flag if there is a drastic change in these parameters or if the nature of the distribution itself changes and requires an explanation. For this, we normalized the sample data in such a way that the K-S test would become insensitive to central tendency and shape.

[11] Indian ambient air quality standard for $PM_{2.5}$ is 40 $\mu g/m^3$, 8-times the World Health Organization (WHO) guideline of 5 $\mu g/m^3$ (as of September 2024). The higher standard is because of the background (nature) pollution in the region will never allow the levels to close to the guideline. Which means, small or extremely small air quality and AQI values are rare. Reanalysed $PM_{2.5}$ concentrations database by year and by month for the period 1998 to 2022 is available @ https://urbanemissions.info

The sample data was normalized the following way:

$$x_i = \frac{x_i - \underline{x}}{x_{max} - x_{min}} \; ; \; \forall \; i \epsilon (1, N)$$

Where x_i is an individual data point in the sample, $\underline{x}$ is the mean of the sample, x_{max} is the maximum value of the sample, x_{min} is the minimum value of the sample and N is the sample size (1000). After this normalization, we find that the K-S test becomes less sensitive to central tendencies and shape. This K-S test application will work when the samples are picked from normal and log-normal distributions.

Sample 1	**Sample 2**	**K-S statistic (p-value)**	**Remark** Fail to reject = Both samples may not be from different distributions. Reject = Both samples are from different distributions
~N (100,10)	~N (100,10)	0.024 (0.94)	Fail to reject the null hypothesis.
~N (100,10)	~N (10,10)	0.024 (0.94)	Fail to reject the null hypothesis.
~N (100,10)	~N (100,100)	0.024 (0.94)	Fail to reject the null hypothesis.
~N (100,10)	~N (10,100)	0.024 (0.94)	Fail to reject the null hypothesis.
~logN(0,0.1)	~logN(0,0.1)	0.032 (0.68)	Fail to reject the null hypothesis
~logN(0,0.1)	~logN(0.3,0.1)	0.032 (0.68)	Fail to reject the null hypothesis.
~logN(0,0.1)	~logN(0,0.3)	0.054 (0.11)	Fail to reject the null hypothesis.
~logN(0,0.1)	~logN(0.3,0.3)	0.054 (0.11)	Fail to reject the null hypothesis.

We assume that the air quality data of a year is distributed in either way[12]. Hence, if the normalized air quality data from two different years fails the K-S test, we find it as a potential case of deviation requiring additional explanation.

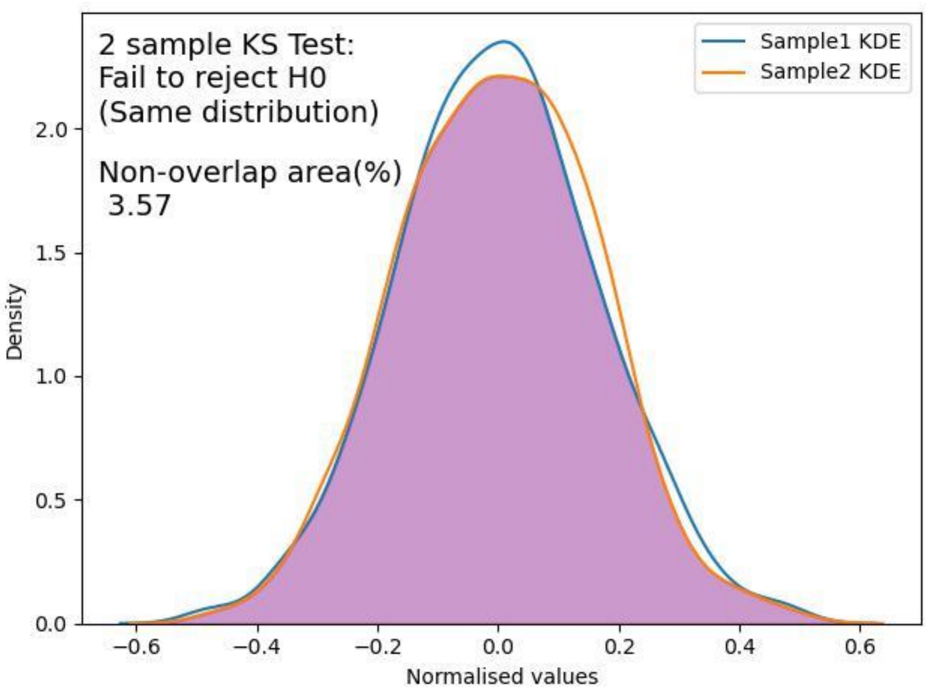
[12] "A Physical Explanation of the Lognormality of Pollutant Concentrations" (1990) @ https://www.tandfonline.com/doi/abs/10.1080/10473289.1990.10466789

PDF non-overlapping area

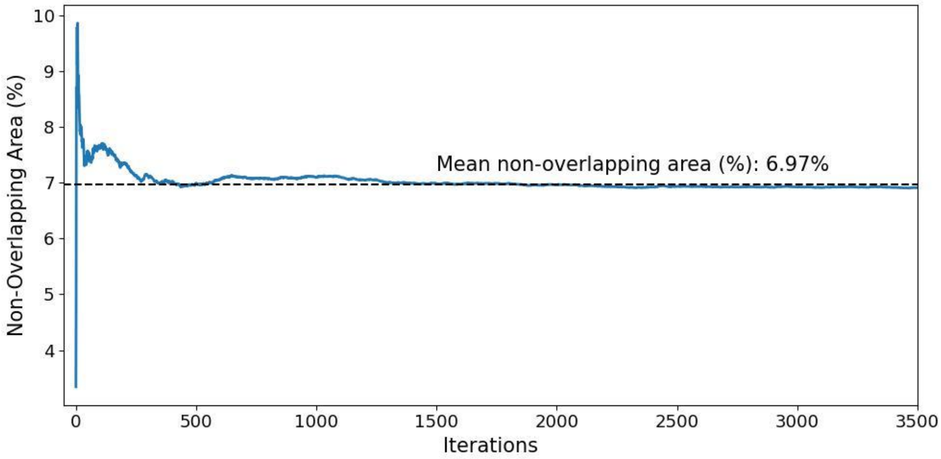
To better appreciate the deviation in the air quality data between years, we also present it visually by comparing the probability distribution functions (PDFs) of the years. Assumption is that there will be some non-overlap in distribution of data (whether comparing day-to-day, season-to-season, year-to-year, or instrument-to-instrument) and in which case, **what is an acceptable level of non-overlap area as deviation?**

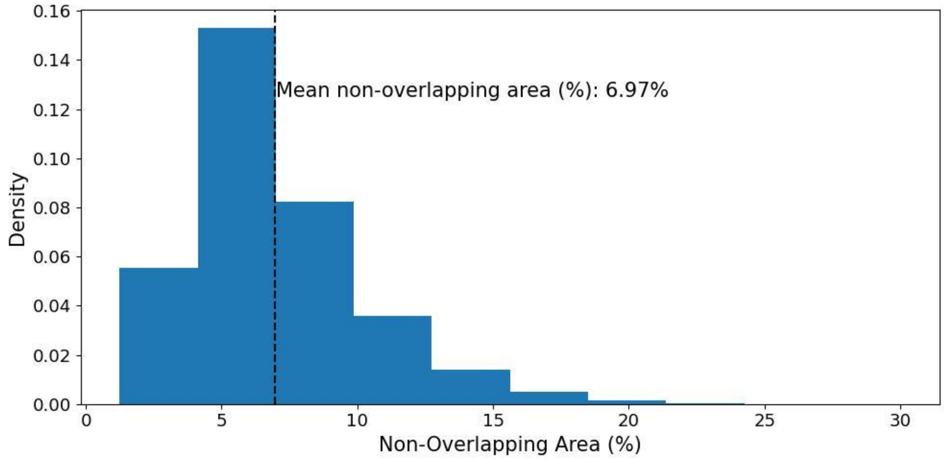
To benchmark this number, this investigation is performed on normalised sample data, like the K-S test.

Consider two random samples of size 365 (days in a year) taken from the standard normal distribution ~ N (0,1). We plot the kernel density estimates (KDE; non-parametric estimation of PDFs) of both the samples and calculate the area not overlapping between them. We can use this non-overlapping area to estimate deviation between both the samples.



In the figure, KDEs of two large samples picked randomly from the same distribution contain some non-overlapping area. This area will change with different samples. We calculated the expected value of this non-overlapping area by performing a Monte-Carlo simulation (using 3500 simulations, as illustrated in the following figures). This expected value could serve as a benchmark to understand what percentage of non-overlapping area can be considered as normal and at what scale is it considered an anomaly requiring further investigation.





The Monte-Carlo simulation converged at a non-overlapping area percentage of 6.97%. The histogram presents the distribution of possible non-overlapping areas. This area percentage can go beyond 15% in a few cases, even when both samples are from the same standard normal distribution. We use these numbers as benchmarks to identify potential deviations.

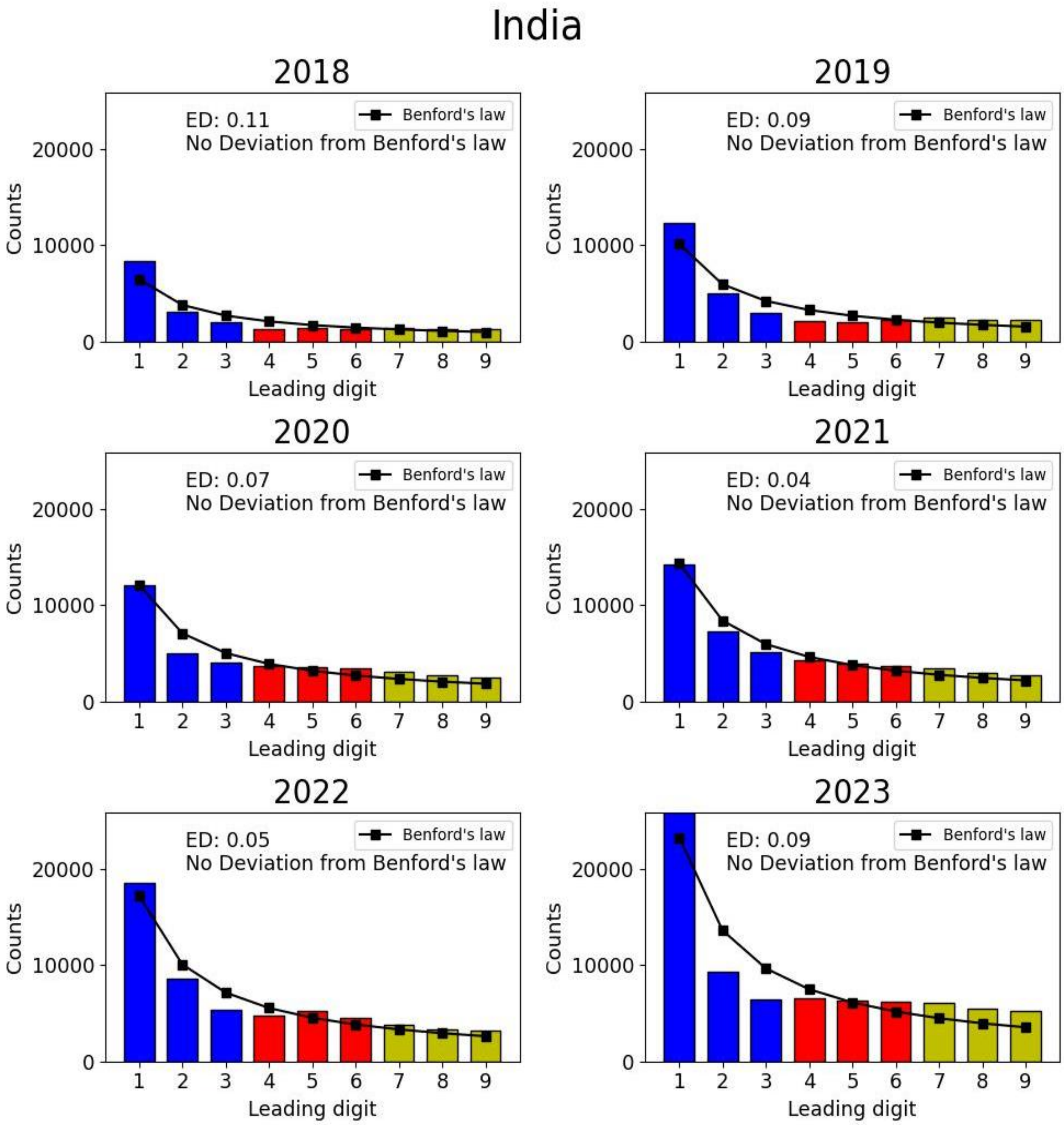
For year-to-year comparisons, we considered 20% as an acceptable deviation, assigning it to the impacts of meteorological fields, growing emission rates from all anthropogenic sources, and reducing emission rates from some sectoral management programs.

4. All India Analysis

According to the Benford’s law

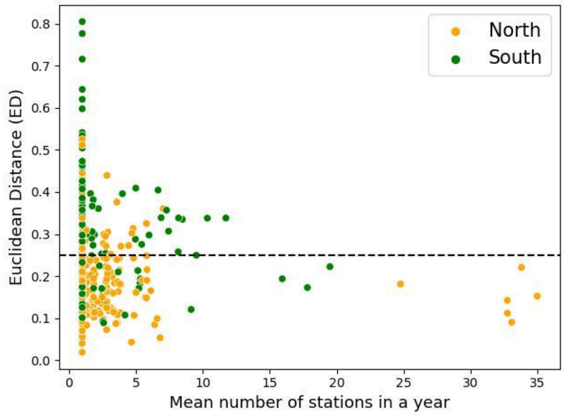
- **Is there any deviation in air quality data at all-India scale?**
- **Is there any deviation in air quality data at sub-national scale?**

Firstly, we observed that the daily average AQI values reported by all Indian cities closely follow Benford's law. From the figure and the overall ED value (of under 0.1) for years 2018 to 2023, we interpret that Benford’s law provides no material to suspect misreporting at the national level.

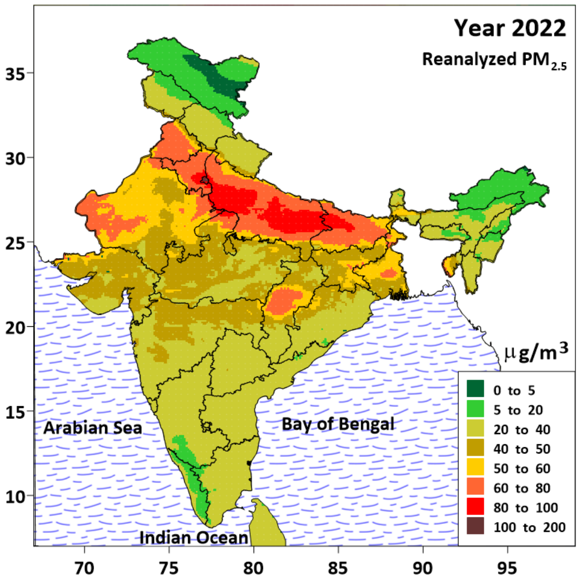


Secondly, we calculated the ED's from Benford's law for every city and every year, if the daily average AQI data is reported for at least 300 days in that year. We collected 610 data points, as not every city could maintain the minimum 300 days. We observed that cities deviated from Benford's law 194 times (all points with ED > 0.25).

Left panel: Euclidean distance calculated for the cities with at least 300 days of data in a year. Colours indicate the city location as North-India or South-India



Right panel: Reanalysed $PM_{2.5}$ concentrations from WUSTL-GEOS-chem chemical transport modelling system[13]



CPCB recommends that a city should have a minimum of four (4) ambient monitor stations to spatially and temporally represent all the landuse and emission activities. We observe that 172 of the 194 deviations occurred in cities with less than this minimum. Remaining 22 deviations occurred in cities that have 4-15 monitors. **Cities with 15+ monitors never deviated from Benford's law, considering the 0.25 rule of thumb**.

In other words, if there is a violation, it is often from a city with a small monitoring network (under 4 stations). In general, the southern cities are more likely to deviate from Benford's law (more green dots in the figure) because of lower ambient concentrations and lesser spread of the first digits.

No. of stations in the city	No. of data points	No. of deviations from Benford's law
1-3	534	172
4-15	67	22
15-35	9	0

Data from small monitoring networks is known to exhibit inconsistencies as they are not capable of representing the landuse and emission patterns of the urban

[13] Reanalysed database for all India at 0.1° resolution and as state-district averages is available @ https://urbanemissions.info and the global files at 0.1 ° and 0.01 ° resolution are available @ https://sites.wustl.edu/acag/datasets/surface-pm2-5

airshed. To catch the heterogeneity of the activities in the airshed, the network size must be at least 4-5 stations covering at least a residential, commercial, industrial, traffic, and background location[14].

In 2023, there were only 22 cities operating at least 4 monitors at some point during the year. The deviation from Benford's law in these 22 cities (as ED), is presented in the following table. We also present the maximum number of monitors operated in that year in the parentheses. We report NA when there are less than 300 days of data reported in a year (for example in Pune in 2022)

Table: Euclidian Deviation (ED) from Benford's Law and number of monitoring stations operational in the city between 2018 and 2023. The cells with ED > 0.25 are highlighted. Only cities with minimum 4 stations operational in 2023 are studied.

	2018	2019	2020	2021	2022	2023
Agra	0.1 (1)	0.09 (1)	0.07 (1)	0.12 (5)	0.19 (6)	0.25 (6)
Ahmedabad	0.24 (1)	0.4 (1)	0.29 (1)	0.24 (9)	0.27 (9)	0.36 (9)
Bengaluru	0.3 (10)	0.26 (10)	0.33 (10)	0.34 (10)	0.31 (10)	0.34 (13)
Chennai	0.25 (3)	0.25 (4)	0.4 (8)	0.4 (8)	0.36 (9)	0.33 (8)
Delhi	0.18 (35+)	0.14 (35+)	0.09 (35+)	0.11 (35+)	0.22 (35+)	0.15 (37)
Faridabad	0.17 (1)	0.12 (1)	0.14 (4)	0.12 (4)	0.21 (4)	0.18 (4)
Ghaziabad	0.17 (1)	0.17 (4)	0.08 (4)	0.16 (4)	0.21 (4)	0.18 (4)
Gurugram	0.24 (2)	0.11 (2)	0.16 (4)	0.12 (4)	0.25 (4)	0.24 (4)
Guwahati	NA	NA	0.19 (2)	0.11 (2)	0.14 (4)	0.09 (4)
Gwalior	NA	NA	0.12 (2)	0.13 (2)	0.13 (3)	0.16 (4)
Hyderabad	0.28 (6)	0.19 (6)	0.18 (6)	0.21 (6)	0.25 (14)	0.34 (14)
Jaipur	0.44 (3)	0.3 (3)	0.24 (3)	0.28 (3)	0.3 (3)	0.31 (6)
Jodhpur	0.38 (1)	0.36 (1)	0.27 (1)	0.33 (1)	0.28 (1)	0.38 (5)
Kanpur	0.13 (1)	0.08 (1)	0.12 (2)	0.1 (4)	0.2 (4)	0.25 (4)
Kolkata	NA	0.11 (7)	0.16 (7)	0.08 (7)	0.1 (7)	0.05 (7)
Lucknow	0.21 (4)	0.11 (4)	0.11 (4)	0.11 (6)	0.19 (6)	0.21 (6)
Moradabad	0.15 (1)	NA	NA	0.17 (3)	0.3 (6)	0.32 (6)
Navi Mumbai	0.33 (1)	0.3 (3)	0.11 (3)	0.1 (3)	0.14 (3)	0.11 (7)
Noida	0.17 (2)	0.14 (4)	0.09 (4)	0.11 (4)	0.18 (4)	0.18 (4)
Patna	0.15 (1)	0.11 (4)	0.04 (6)	0.16 (6)	0.15 (6)	0.15 (6)
Pune	0.27 (1)	0.28 (1)	0.3 (6)	0.28 (8)	NA	0.4 (8)
Raipur	NA	NA	NA	NA	NA	0.21 (4)
Varanasi	0.21 (1)	0.27 (1)	0.11 (1)	0.15 (4)	0.24 (4)	0.27 (4)

Deviation from Benford's law only means that additional information is necessary to explain the change in the data patterns in city's air quality. In Indian cities, we suspect "small sample sizes" and "heterogeneity among the locations" as the main reason to explain the deviations.

[14] "Data from small monitoring networks is not reliable: Case of Indian cities" (2024) SIM-air working paper series #48-2024 @ https://urbanemissions.info. This paper presents mathematical guidance for 4-5 minimum stations using the AQI data from the cities and the margin-of-error concept.

5. City Case Studies

In this investigation, we limited city level analysis to:

- Delhi (15-35 category) for benchmarking purposes
- Three major cities in the 4-15 monitors category: Hyderabad, Chennai, Bengaluru
- Varanasi (4-15 category since 2022) as it reported a 72% reduction in 2024, the highest among all the cities under the National Clean air programme (NCAP)[15]
- Udaipur (1-station category)

For each city, we present the deviation from Benford law as ED value and present the year-on-year deviations using the two-sample K-S test and the PDF non-overlapping areas methods.

PDF assessment is limited to years 2018 to 2023 and year-on-year assessments are presented as two-sets – (a) all years against 2018 to show overall change in the trend distribution and (b) between consecutive years to show any immediate change from emission management schemes.

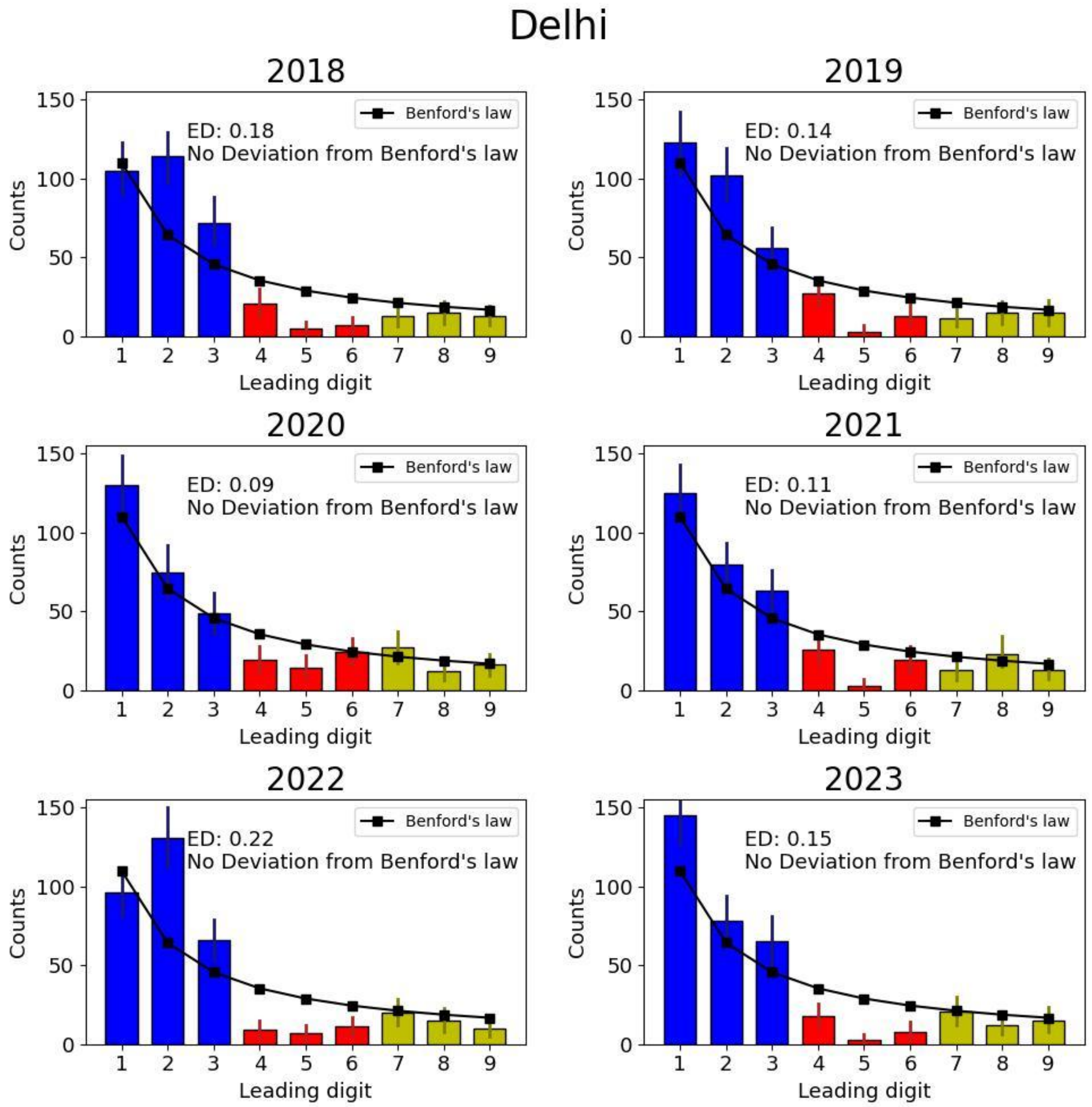
Year-on-year, we expect the trends and distributions to deviate. As described in the methods section, data from each city and year undergoes normalised two sample K-S test to make it less sensitive to the central tendencies and shape. The deviation then observed is more when the nature of the distribution itself changed or when there is a sudden change in the air quality values. We also present the non-overlapping area percentage for each of this comparison in the city specific graphs.

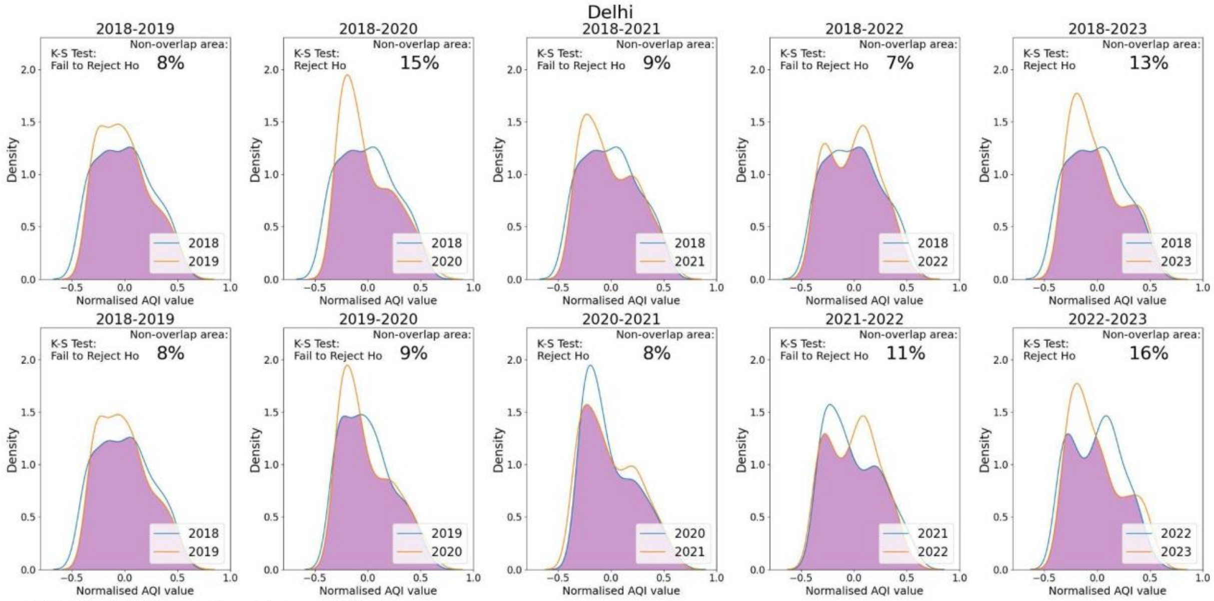
The investigative steps are outlined only as a method to spot these dramatic shifts in the deviations which if it gets tagged as an anomaly, requires additional information to explain the deviation.

All the data, calendar plots, codes to make the plots are available on our github repository and accessible from here @ https://urbanemissions.info and @ https://urbanemissions.info/india-ncap-aqi-indian-cities-2015-2023

India's approved methodology for calculating AQI from pollutant concentrations and five other methodologies can be explored as a MS Excel based calculator, along with instructional videos @ https://urbanemissions.info/tools

[15] 5-year NCAP analysis @ https://carboncopy.info/air-in-delhi-patna-most-polluted-navi-mumbai-worsening-5-yr-ncap-analysis





Null Hypothesis (Ho): Two samples are from same distribution

Delhi operated 35+ monitoring stations between 2018-2023, peaking at 40 stations on some days[16]. It is the highest number of monitors in any city in India and the only city that never deviated from Benford's law. For this reason, we would treat Delhi as a benchmark city and compare other cities with it.

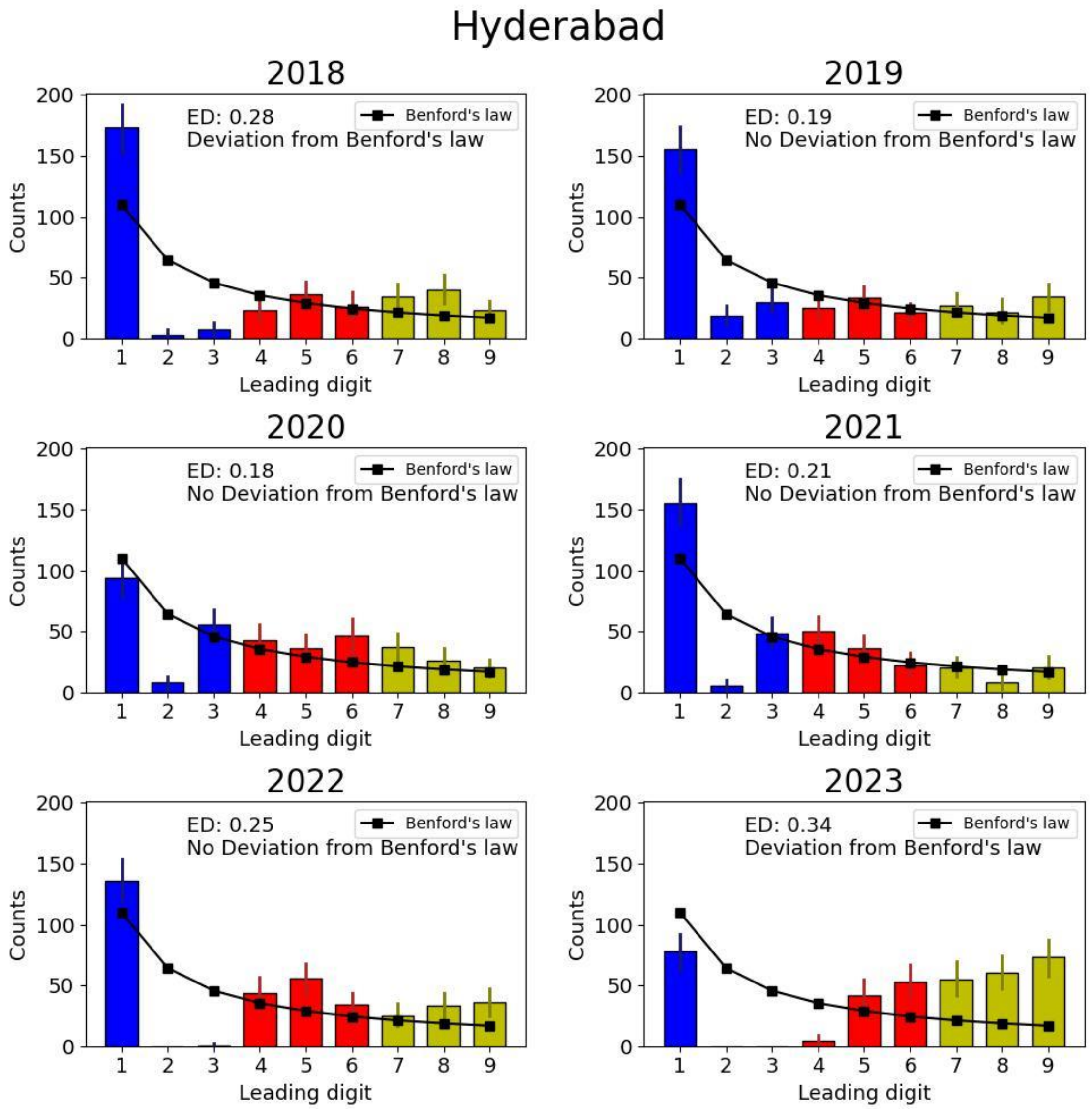
The non-overlap area is under the acceptable limit (20%) for all the year-on-year comparisons. The two sample K-S test detected two deviations:

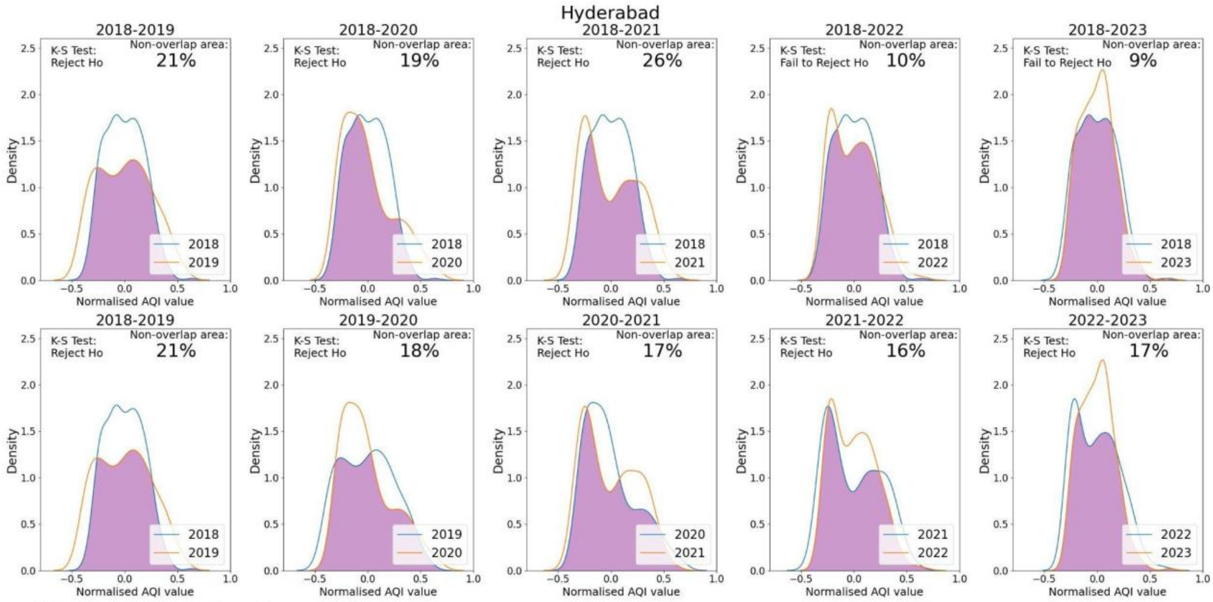
- 2020-21 deviations can be attributed to the pollution drops from COVID-19 lockdown restrictions. In 2021 there is a revival in economic activity and more pollution when compared to 2020 and hence the disappearance of the leftward skew that occurred in 2020.
- 2022-2023 deviation is an interesting one because we can see that the nature of the distribution itself has changed: from a double peak distribution in 2022 to a left-skewed distribution in 2023. This can be looked up into more detail. There is a study from Princeton University, which attributed a significant share of the changes observed in 2023 to shifting meteorological patterns[17].

Similar conclusions also apply to other cities in the national capital region (NCR) of Delhi – Gurugram, Noida, Faridabad, and Ghaziabad. All these cities have multiple monitoring stations (minimum 4), with higher consistency in the data reporting from stations at representative locations across their airsheds.

[16] More detailed long-term analysis on "What is polluting Delhi's air: Review from 1990 to 2022" is available @ https://www.mdpi.com/2071-1050/15/5/4209

[17] "Recent PM2.5 air quality improvements in India benefited from meteorological variation" (2024) @ https://www.nature.com/articles/s41893-024-01366-y





Null Hypothesis (Ho): Two samples are from same distribution

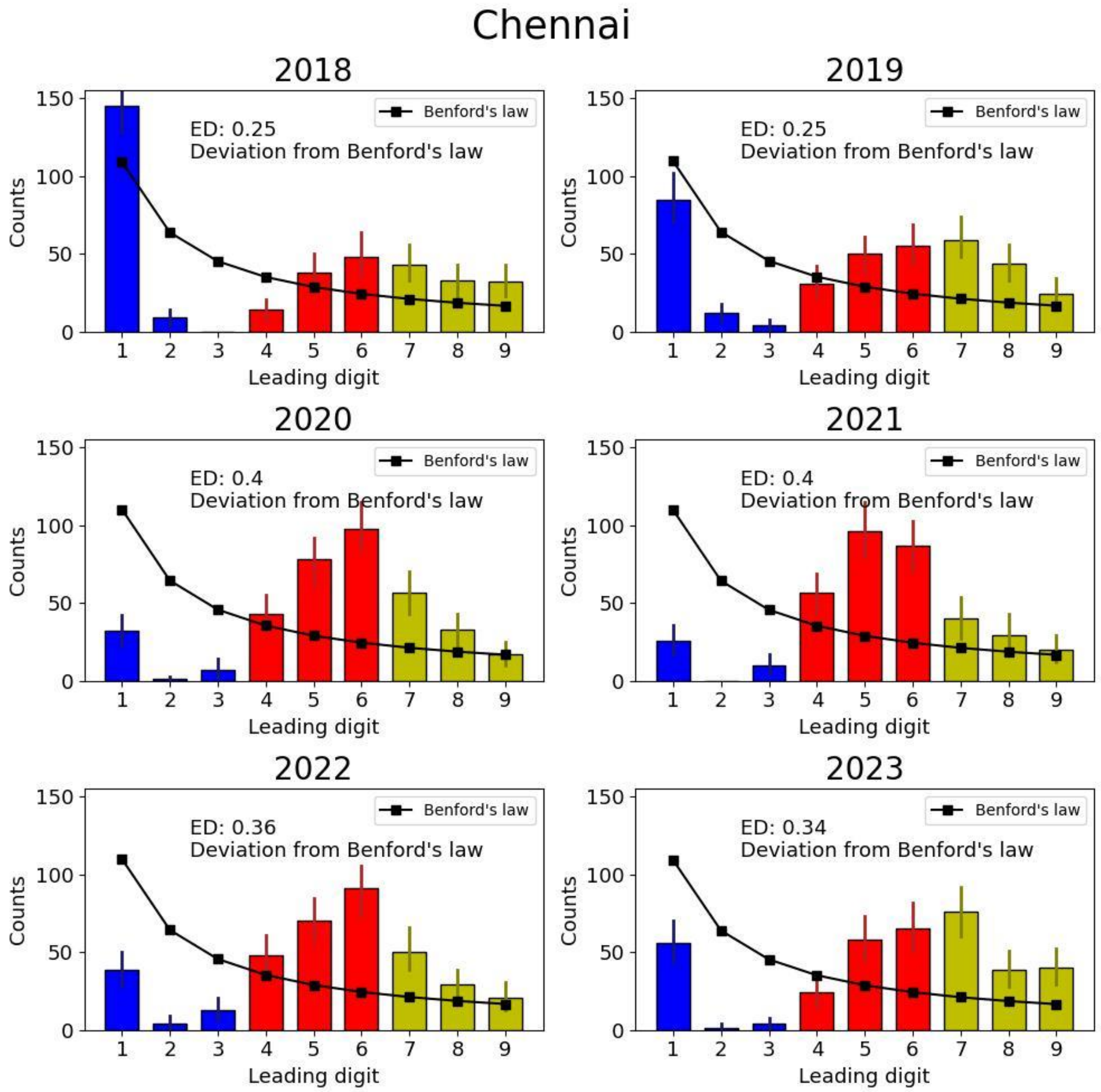
Hyderabad saw deviation from Benford's law twice: 2018 and 2023. However, in 2018 the domination of smaller leading digits continued while in 2023 trailing leading digits dominated (with smaller numbers mostly missing). The deviation in 2023 asks for deeper investigation into the increasing AQI values over the years. A general conclusion is that the overall background concentrations in the city have increased over the years.

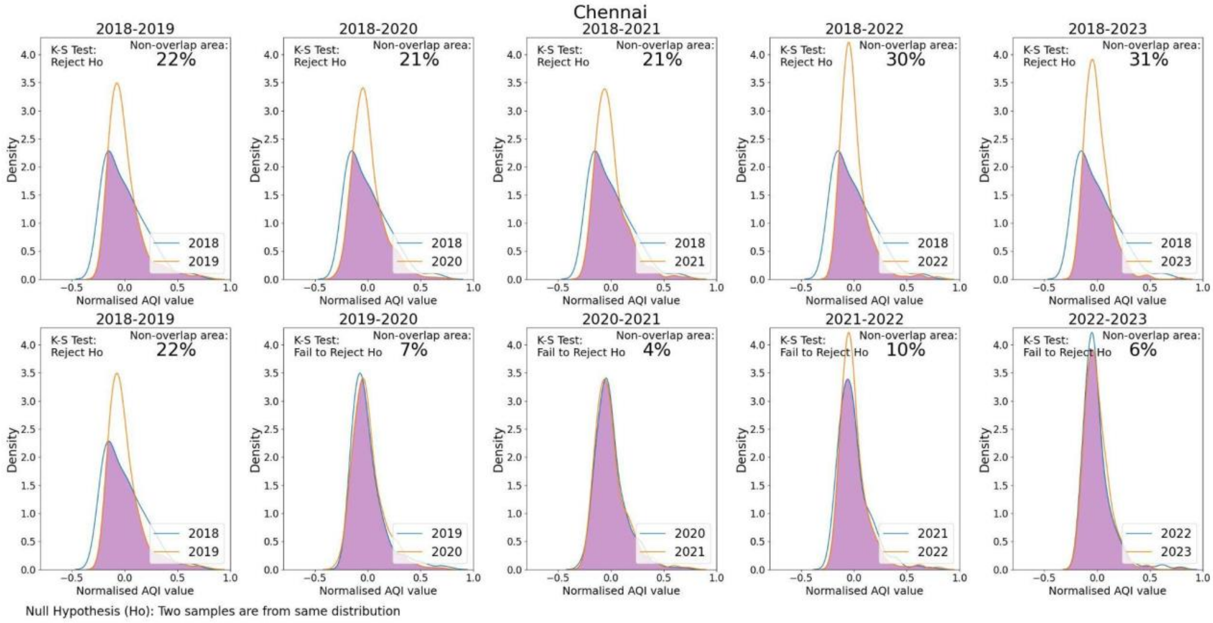
The non-overlap area is under the acceptable limit (20%) for all the year-on-year comparisons. The deviation from Benford's law in 2018 is detected by the K-S test as well.

Interestingly changes in the year-on-year distributions were observed - there is a significant left skew in 2019 which should be explored further. In 2022, Hyderabad had a double peak distribution, which changed in 2023. Every year, Hyderabad's air quality distribution is different when compared to the previous year. This unpredictability needs justification.

Lack of consistency in the availability of monitoring data can also lead to these deviations. Hyderabad increased its monitor network from 3 to 14 stations between 2018 and 2023. However, the operation of monitors was sporadic with different numbers of monitors operating on different number of days.

	No. of reporting days in					
No. of Stations	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	0	0	0	2	0	0
2	0	0	2	2	0	0
3	8	5	17	14	7	0
4	32	39	35	48	12	1
5	116	145	124	154	56	1
6	209	176	188	145	72	1
7	0	0	0	0	1	2
8	0	0	0	0	1	17
9	0	0	0	0	4	24
10	0	0	0	0	17	48
11	0	0	0	0	31	40
12	0	0	0	0	48	75
13	0	0	0	0	68	105
14	0	0	0	0	48	51


Chennai
2018
ED: 0.25
Deviation from Benford's law
2019
ED: 0.25
Deviation from Benford's law
2020
ED: 0.4
Deviation from Benford's law
2021
ED: 0.4
Deviation from Benford's law
2022
ED: 0.36
Deviation from Benford's law
2023
ED: 0.34
Deviation from Benford's law
Benford's law
Counts
Leading digit
0
50
100
150
1
2
3
4
5
6
7
8
9


Chennai
2018-2019
K-S Test:
Reject Ho
Non-overlap area:
22%
2018-2020
K-S Test:
Reject Ho
Non-overlap area:
21%
2018-2021
K-S Test:
Reject Ho
Non-overlap area:
21%
2018-2022
K-S Test:
Reject Ho
Non-overlap area:
30%
2018-2023
K-S Test:
Reject Ho
Non-overlap area:
31%
2018-2019
K-S Test:
Reject Ho
Non-overlap area:
22%
2019-2020
K-S Test:
Fail to Reject Ho
Non-overlap area:
7%
2020-2021
K-S Test:
Fail to Reject Ho
Non-overlap area:
4%
2021-2022
K-S Test:
Fail to Reject Ho
Non-overlap area:
10%
2022-2023
K-S Test:
Fail to Reject Ho
Non-overlap area:
6%
Density
Normalised AQI value
2018
2019
2020
2021
2022
2023
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
−0.5
0.0
0.5
1.0
Null Hypothesis (Ho): Two samples are from same distribution

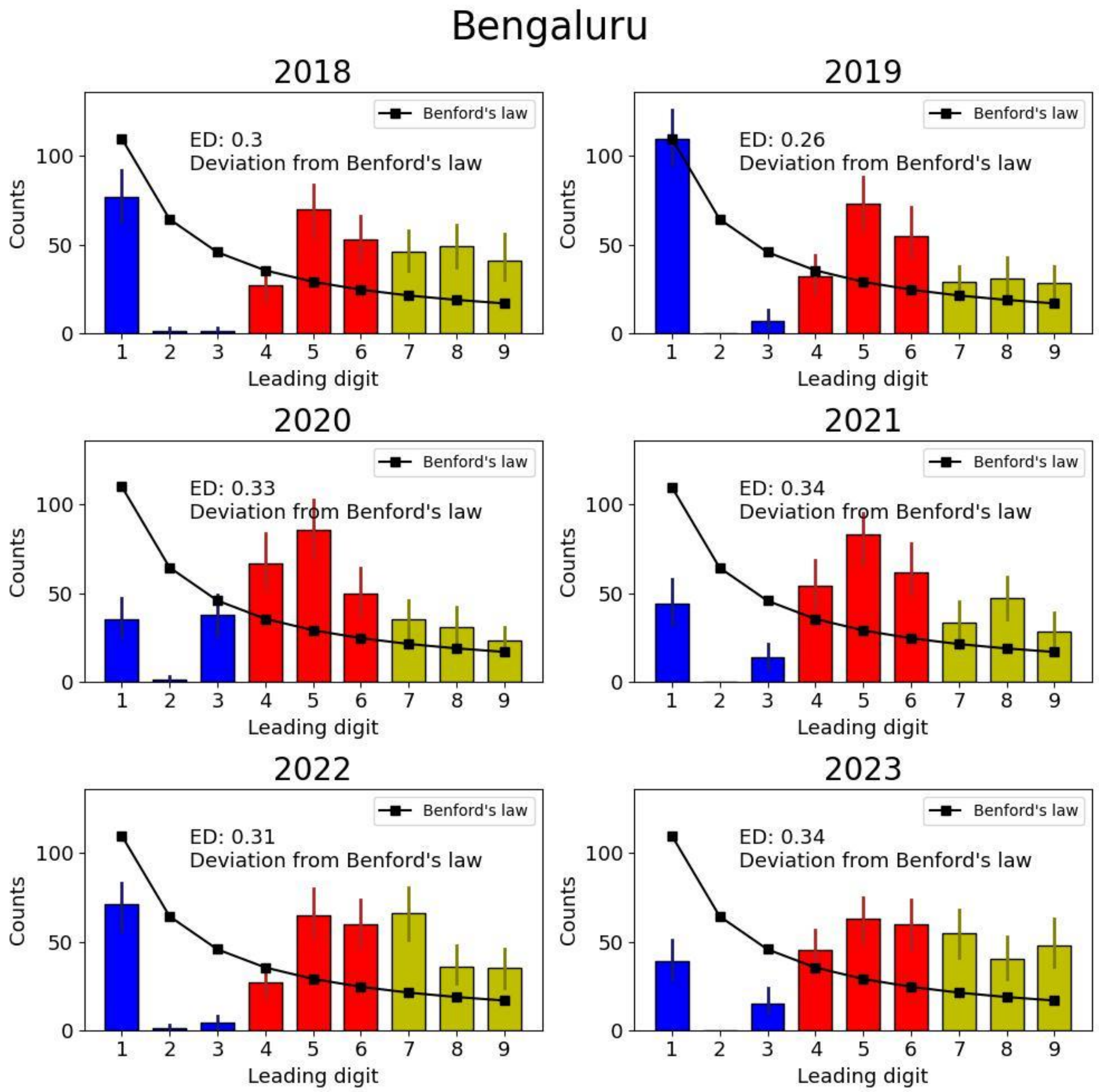
Chennai presents an interesting case, which never followed Benford's law. Middle leading digits were more dominant in all the years. While this might flag potential misreporting, the year-on-year comparison of PDFs tells a different story. After one major change in the distribution between 2018-19, the air quality distributions in Chennai have remained consistent in the later years.

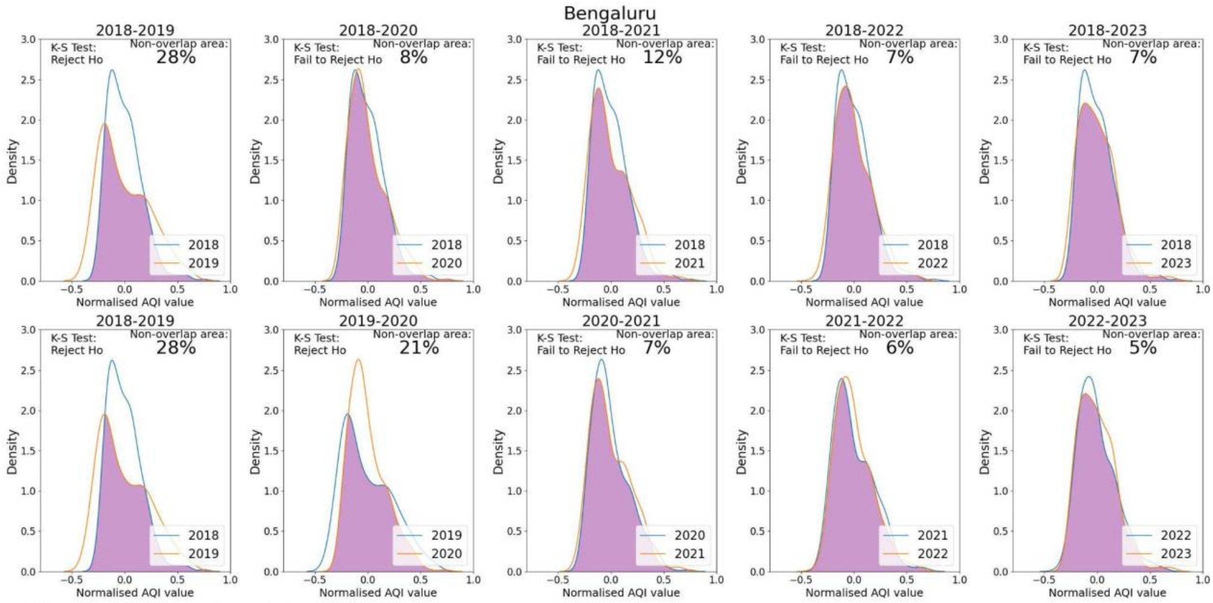
In 2018, Chennai had only 1-3 monitors, which can result in statistical bias in their reporting. If we consider this as a reason for the major shift in the distribution, then we can observe the consistency in distributions later, is a result of network expansion and consistent reporting from the stations.

	No. of reporting days in					
No. of Stations	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	13	25	0	0	0	2
2	177	89	5	0	0	2
3	172	201	87	3	0	4
4	0	49	228	12	6	6
5	0	0	12	33	24	15
6	0	0	18	91	31	81
7	0	0	15	145	128	126
8	0	0	1	81	149	127
9	0	0	0	0	27	0

Chennai thus presents a case study where despite deviation from Benford's law, the year-on-year comparison of PDFs present a better understanding of potential deviations with no data red-flags.

Missing lower order numbers also raise a background question – what is the impact of natural sources like sea-salt emissions? Are these emissions and general background contributions (like shipping) being high enough to sustain the AQI numbers above 20s-30s values (a shift to the right)?





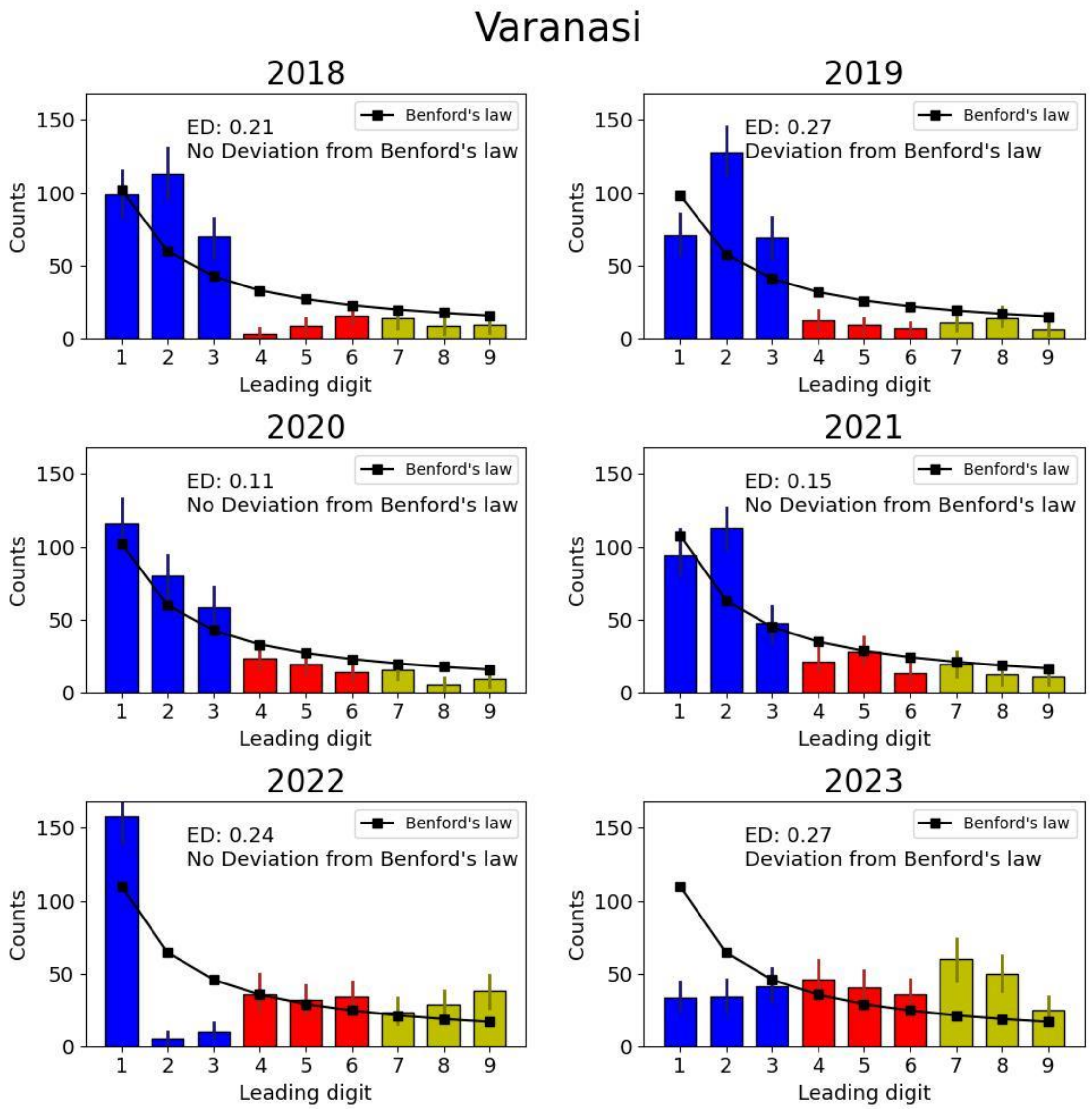
Null Hypothesis (Ho): Two samples are from same distribution

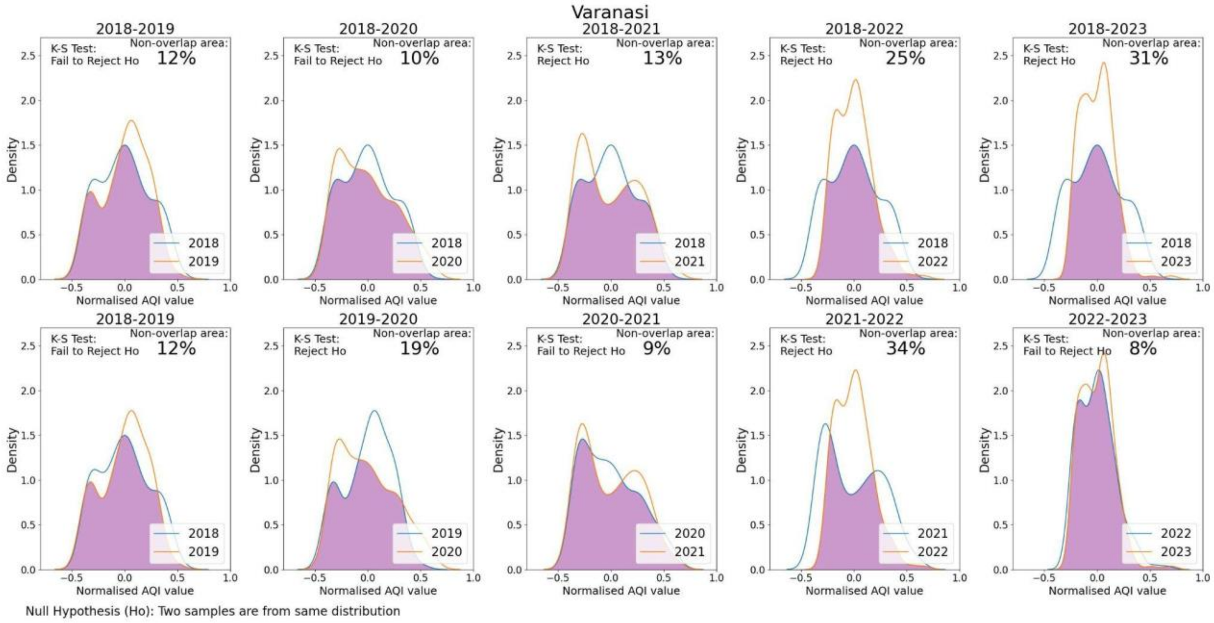
Bengaluru's story is like that of Chennai. It deviated from Benford's law every year due to dominance of middle leading digits. But the air quality distributions have been consistent over the past three years.

In 2019-20, deviation in the distribution can be explained as an influence of the COVID-19 lockdown restrictions. Since Bengaluru's air pollution is dominated by vehicle exhaust and road-dust, the changes in 2020 during the lockdown period were immediate and significant with most of the traffic off-roads.

In 2018, Bengaluru's monitor network has not been consistent with different numbers of monitors reporting data for different number of days. It can be observed that as the network expanded, the air quality distributions have become predictable.

	No. of reporting days in					
No. of stations	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	7	0	0	0	0	0
2	15	0	0	0	1	0
3	27	1	0	0	3	0
4	57	4	0	3	7	0
5	80	3	2	0	31	1
6	20	32	15	25	55	6
7	42	47	45	64	79	16
8	48	122	113	122	86	35
9	48	110	126	107	78	51
10	21	46	65	44	25	69
11	0	0	0	0	0	86
12	0	0	0	0	0	67
13	0	0	0	0	0	34


Varanasi
2018
ED: 0.21
No Deviation from Benford's law
Benford's law
2019
ED: 0.27
Deviation from Benford's law
Benford's law
2020
ED: 0.11
No Deviation from Benford's law
Benford's law
2021
ED: 0.15
No Deviation from Benford's law
Benford's law
2022
ED: 0.24
No Deviation from Benford's law
Benford's law
2023
ED: 0.27
Deviation from Benford's law
Benford's law
Counts
0
50
100
150
Leading digit
1
2
3
4
5
6
7
8
9

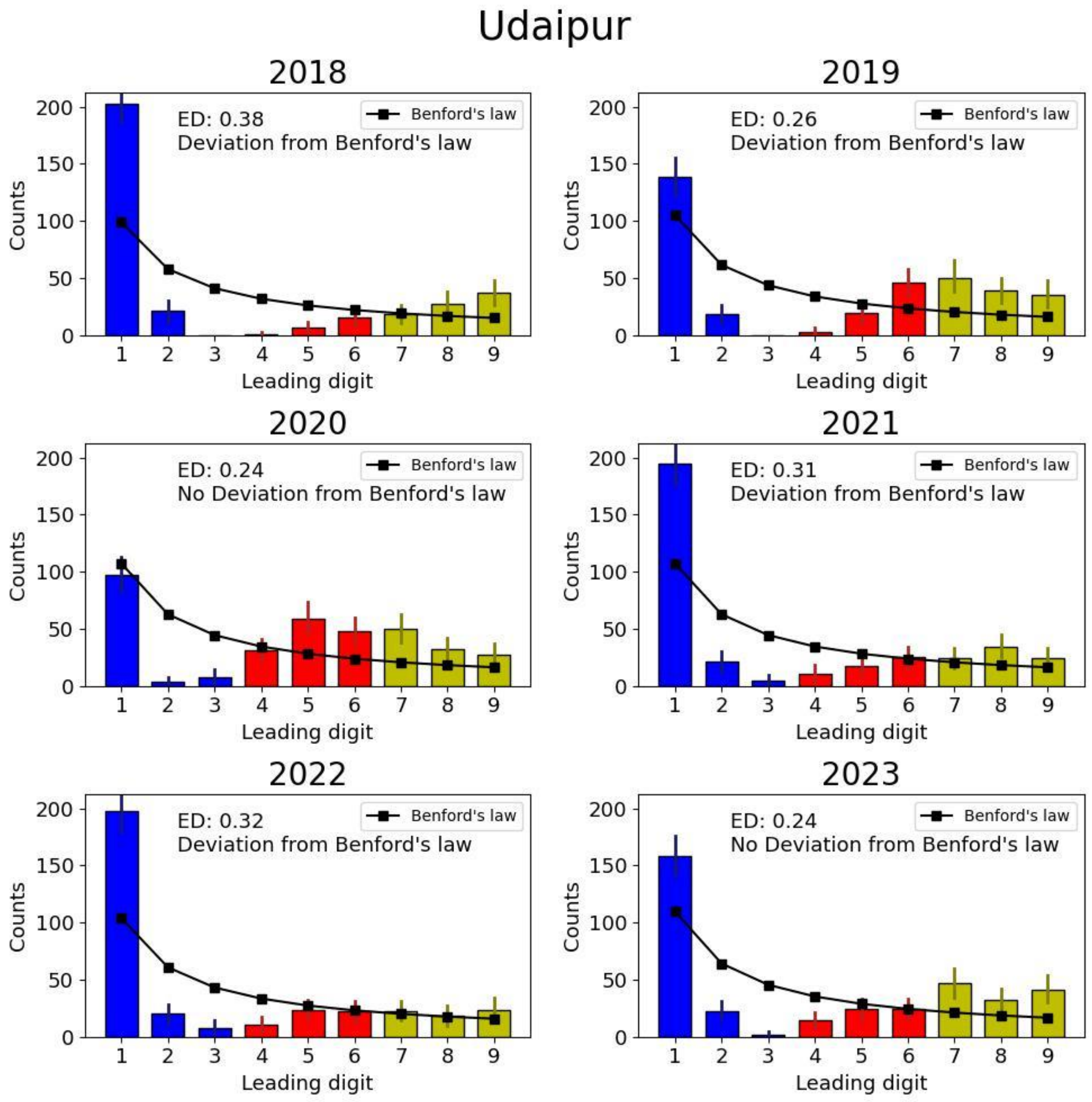

Varanasi
2018-2019
K-S Test: Fail to Reject Ho
Non-overlap area: 12%
2018
2019
2018-2020
K-S Test: Fail to Reject Ho
Non-overlap area: 10%
2018
2020
2018-2021
K-S Test: Reject Ho
Non-overlap area: 13%
2018
2021
2018-2022
K-S Test: Reject Ho
Non-overlap area: 25%
2018
2022
2018-2023
K-S Test: Reject Ho
Non-overlap area: 31%
2018
2023
2018-2019
K-S Test: Fail to Reject Ho
Non-overlap area: 12%
2018
2019
2019-2020
K-S Test: Reject Ho
Non-overlap area: 19%
2019
2020
2020-2021
K-S Test: Fail to Reject Ho
Non-overlap area: 9%
2020
2021
2021-2022
K-S Test: Reject Ho
Non-overlap area: 34%
2021
2022
2022-2023
K-S Test: Fail to Reject Ho
Non-overlap area: 8%
2022
2023
Density
0.0
0.5
1.0
1.5
2.0
2.5
Normalised AQI value
−0.5
0.0
0.5
1.0
Null Hypothesis (Ho): Two samples are from same distribution

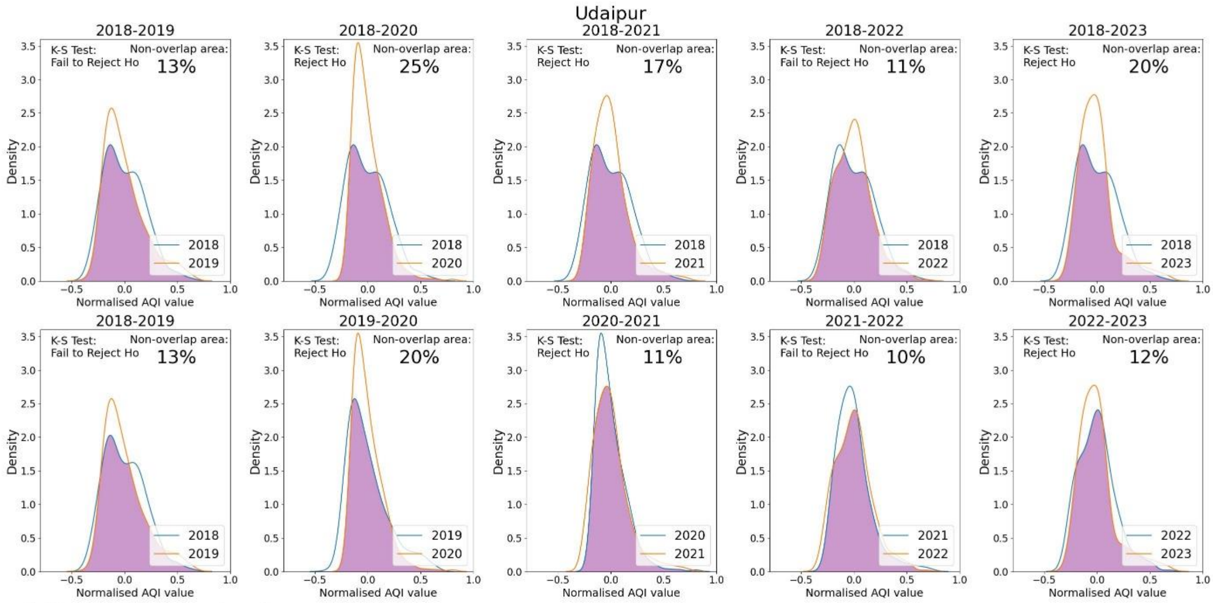
Varanasi air quality between 2018 and 2022, there has been a dominance of smaller leading digits as expected by Benford's law. In 2023, there is a sudden shift in the way the leading digits are distributed. This is unlike the deviations from Benford's law that we observed in Chennai and Bengaluru, where middle leading digits have a consistent dominance. This sudden shift in the way the leading digits are distributed calls for a closer inspection.

The K-S tests also indicate a sudden shift that started in the year 2022 (with non-overlap area of 34%) and continued in 2023. A left skew appeared in 2022 and 2023 that changed the air quality distribution significantly in Varanasi. In these years, Varanasi also reported more values with 4 monitors.

	No. of reporting days in					
No. of Stations	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	339	327	339	166	0	0
2	0	0	0	10	10	1
3	0	0	0	33	64	38
4	0	0	0	149	291	326

One line of reasoning behind this major shift in 2022's distribution is that, like in Chennai and Bengaluru, the new data from the growing monitoring network shifted the distribution. But if one additional monitor shifted the distribution to the left by such an extent, a closer look is necessary on the working of the monitor.





Null Hypothesis (Ho): Two samples are from same distribution

Udaipur started ambient monitoring in 2017 using one station. As of September 2024, the city continues to operate only one monitoring station. Udaipur is included as a case study only to illustrate the low confidence intervals in the data, with Benford deviations calculated for 4 out 6 years and K-S test reject comparisons among consecutive years for 3 out 5 sets.

Data from cities with only one monitoring station must be used with caution.

6. Conclusions

Cities are on high alert to show improvements in the city to access financial resources for emissions management (as designed under the national clean air programme – NCAP).

In this working paper, we demonstrated a 2-step process to identify these changes and flag them if the changes are too dramatic between years (based on statistical methods) and if yes, request an explanation for the same.

Few concerns regarding how the AQI is reported by CPCB:

- Firstly, AQI bulletins report a city average. But the number of monitors active on a day has high variance, especially as observed in Hyderabad and Bengaluru. The AQI averages thus calculated may not be comparable and useful for longitudinal analysis.
- Secondly, the annual distributions of these unreliable AQI averages would not be useful for public policy purposes. If the very nature of distribution is rapidly shifting from one year to the other, there cannot be a planned policy response. There is indeed a chance that this rapid shift is truly because of natural conditions.
- Finally, cities with denser and consistent networks of monitors deviated less from Benford's law. Even if they deviated from Benford's law, their annual distributions are consistent and thus predictable. Such AQI data is more useful for policy work. In other words, for good policy support, all efforts to increase the city monitoring networks to operate 10-15+ stations must be encouraged at all costs.

7. Annexure: Recommended Minimum No. of Monitoring Locations for Indian Airsheds Under NCAP

Based on the guidelines issued by the Central Pollution Control Board for ambient monitoring in 2003, the following minimums were calculated.

Full publication on the methods are published here
Plugging the ambient air monitoring gaps in India's national clean air programme (NCAP) airsheds (Atmospheric Environment, 2023)
https://www.sciencedirect.com/science/article/pii/S1352231023001383

And the associated databases are published here
https://urbanemissions.info

Table: Characteristics of airsheds designated for NCAP non-attainment cities. B = cities included in the airshed from the NCAP list; C = cities included in the airshed, but not on the NCAP list; D = airshed size in grids of equal size (0.01°); E = total airshed population (in million); F = fraction of grids designated as urban using built-up area; G = fraction of population in the urban grids; H, I, J, K = number of continuous monitoring stations recommended for tracking PM, SO_2, NO_2, and Others respectively.

	State/UT	Airshed	B	C	D	E	F	G	H	I	J	K
1	Andhra Pradesh	Anantapur			30 x 30	0.6	8%	60%	10	6	8	2
2	Andhra Pradesh	Chitoor			30 x 30	0.5	8%	50%	9	5	7	2
3	Andhra Pradesh	Eluru		Hanuman Junction	30 x 30	0.7	8%	50%	10	6	8	2
4	Andhra Pradesh	Kadapa			30 x 30	0.5	6%	62%	9	6	8	2
5	Andhra Pradesh	Kurnool			30 x 30	0.7	10%	65%	10	6	9	3
6	Andhra Pradesh	Nellore			30 x 30	0.8	15%	66%	12	7	9	3
7	Andhra Pradesh	Ongole			30 x 30	0.5	9%	54%	9	5	7	2
8	Andhra Pradesh	Rajahmundry			30 x 30	1.4	25%	55%	17	9	10	4
9	Andhra Pradesh	Srikakulam			30 x 30	0.7	8%	41%	10	6	8	2
10	Andhra Pradesh	Vijayawada	Guntur	Tenali	50 x 50	3.1	23%	65%	22	11	10	6
11	Andhra Pradesh	Vishakhapatnam		Anakapalle	50 x 50	2.9	18%	68%	20	11	10	6
12	Andhra Pradesh	Vizianagaram			30 x 30	0.9	9%	47%	12	8	10	3
13	Assam	Guwahati	Byrnahati	Dispur	40 x 30	1.7	36%	73%	18	9	10	4
14	Assam	Nagaon			30 x 30	1.2	47%	20%	36	8	10	3
15	Assam	Nalbari			30 x 30	0.9	31%	56%	11	8	10	3
16	Assam	Sibsagar			30 x 30	0.5	19%	32%	12	5	7	2
17	Assam	Silchar			30 x 30	1.1	14%	18%	19	8	10	3
18	Bihar	Gaya			30 x 30	1.6	18%	30%	19	9	10	4
19	Bihar	Muzaffarpur			30 x 30	2.7	42%	30%	35	11	10	6
20	Bihar	Patna			60 x 40	7.0	38%	46%	43	17	10	10
21	Chandigarh	Chandigarh	Dera Bassi, Parwanoo	Panchkula, Kalka	50 x 40	2.9	40%	76%	23	11	10	6
22	Chhattisgarh	Korba			40 x 40	0.9	11%	58%	12	7	10	3
23	Chhattisgarh	Raipur	Bhillai	Durg	60 x 30	3.2	29%	76%	22	11	10	6

24	Delhi	Delhi	Faridabad, Ghaziabad, Noida	Greater Noida, Gurugram, Palwal, Manesar, Sonipat	100 x 100	32.8	43%	79%	101	20	10	23
25	Gujarat	Ahmedabad		Gandhi Nagar	50 x 50	7.9	40%	79%	38	18	10	10
26	Gujarat	Rajkot			30 x 30	1.5	24%	80%	16	9	10	4
27	Gujarat	Surat		Hazira	50 x 50	5.8	23%	61%	30	15	10	9
28	Gujarat	Vadodara			30 x 30	2.6	34%	82%	21	10	10	5
29	Himachal Pradesh	Kala Amb			30 x 30	0.4	7%	29%	9	5	7	2
30	Himachal Pradesh	Nalagarh	Baddi		30 x 30	0.3	20%	62%	9	5	7	2
31	Himachal Pradesh	Paonta Sahib			20 x 20	0.2	12%	53%	7	4	5	2
32	Himachal Pradesh	Sunder Nagar			20 x 20	0.2	22%	63%	8	4	6	2
33	Jammu & Kashmir	Jammu			30 x 30	1.3	47%	65%	19	8	10	3
34	Jammu & Kashmir	Srinagar			30 x 30	2.1	56%	77%	23	10	10	5
35	Jharkhand	Dhanbad			60 x 40	3.8	23%	39%	28	12	10	7
36	Jharkhand	Jamshedpur		Bokaro, Jaropokhar	40 x 40	2.2	12%	61%	16	10	10	5
37	Jharkhand	Ranchi			40 x 40	1.9	20%	58%	17	9	10	4
38	Karnataka	Bangalore			60 x 60	11.7	50%	81%	50	20	10	12
39	Karnataka	Devanagere			30 x 30	0.9	12%	65%	12	7	10	3
40	Karnataka	Gulburga			30 x 30	0.8	10%	71%	11	7	9	3
41	Karnataka	Hubli-Dharwad			30 x 30	1.3	18%	77%	14	8	10	3
42	Madhya Pradesh	Bhopal			40 x 40	2.6	23%	86%	19	10	10	5
43	Madhya Pradesh	Gwalior			30 x 30	1.4	17%	71%	15	9	10	4
44	Madhya Pradesh	Indore	Dewas, Ujjain	Mhow, Pitampura	80 x 80	5.5	11%	51%	26	15	10	9
45	Madhya Pradesh	Jabalpur			40 x 40	1.9	15%	75%	16	9	10	4
46	Madhya Pradesh	Sagar			30 x 30	0.5	8%	61%	9	6	8	2
47	Maharashtra	Akola			30 x 30	0.8	10%	64%	11	7	9	3
48	Maharashtra	Amravati			30 x 30	0.9	10%	74%	12	8	10	3
49	Maharashtra	Aurangabad			40 x 40	1.9	16%	73%	16	9	10	4
50	Maharashtra	Chandrapur			30 x 30	0.7	12%	73%	11	7	9	3
51	Maharashtra	Jalgaon			30 x 30	0.8	10%	66%	11	7	9	3
52	Maharashtra	Jalna			30 x 30	0.6	7%	51%	9	6	8	2
53	Maharashtra	Kolhapur	Sangli		60 x 40	3.9	23%	47%	26	12	10	7
54	Maharashtra	Latur			30 x 30	0.8	10%	60%	11	7	9	3
55	Maharashtra	Mumbai	Badlapur, Navi Mumbai, Thane, Ulhasnagar, Vasai Virar	Kalyan, Karjat	80 x 80	25.1	21%	78%	67	20	10	19
56	Maharashtra	Nagpur			40 x 40	3.6	28%	88%	23	12	10	7
57	Maharashtra	Nashik			40 x 40	2.6	29%	75%	20	10	10	5
58	Maharashtra	Pune		Pimpri-Chinchwad, Hinjewadi	40 x 40	6.8	60%	86%	40	17	10	10
59	Maharashtra	Solapur			30 x 30	1.1	16%	79%	13	8	10	3
60	Nagaland	Dimapur			30 x 30	0.5	22%	80%	10	5	7	2
61	Nagaland	Kohima			30 x 30	0.2	5%	54%	7	4	6	2
62	Orissa	Angul	Talcher		40 x 40	0.7	11%	39%	12	7	9	3
63	Orissa	Balasore			30 x 30	0.8	8%	36%	12	7	9	3
64	Orissa	Bhubaneswar	Cuttack, Kalinga Nagar		40 x 40	3.2	21%	60%	22	11	10	6
65	Orissa	Rourkela			30 x 30	1.2	16%	56%	15	8	10	3
66	Punjab	Amritsar		Tarn Taran	40 x 40	2.2	38%	69%	21	10	10	5
67	Punjab	Jalandhar		Phagwara	40 x 40	1.9	44%	65%	22	9	10	4

68	Punjab	Khanna	Gobindgarh		30 x 30	0.7	37%	69%	14	7	9	3
69	Punjab	Ludhiana		Philaur	40 x 40	2.7	45%	78%	23	11	10	6
70	Punjab	Naya Nangal		Una	30 x 30	0.5	29%	65%	11	5	7	2
71	Punjab	Pathankot/Dera Baba	Damtal		30 x 30	0.7	30%	70%	13	7	9	3
72	Punjab	Patiala			60 x 40	1.8	22%	48%	19	9	10	4
73	Rajasthan	Alwar			30 x 30	0.9	18%	67%	13	7	10	3
74	Rajasthan	Jaipur			40 x 40	4.5	54%	90%	31	13	10	8
75	Rajasthan	Jodhpur			40 x 40	1.9	26%	83%	17	9	10	4
76	Rajasthan	Kota			30 x 30	1.1	25%	83%	14	8	10	3
77	Rajasthan	Udaipur			30 x 30	1.4	27%	71%	16	9	10	4
78	Tamil Nadu	Chennai			50 x 50	10.9	44%	83%	46	20	10	12
79	Tamil Nadu	Madurai		Singrauli	30 x 30	2.1	27%	86%	18	10	10	5
80	Tamil Nadu	Thoothukudi			40 x 40	0.9	11%	66%	12	7	10	3
81	Tamil Nadu	Trichy			30 x 30	1.8	31%	78%	18	9	10	4
82	Telangana	Hyderabad	Patancheru, Sangareddy		60 x 60	9.0	36%	85%	39	20	10	11
83	Telangana	Nalgonda			30 x 30	0.4	6%	44%	8	5	7	2
84	Uttar Pradesh	Agra			40 x 40	3.7	22%	66%	23	12	10	7
85	Uttar Pradesh	Allahabad			40 x 40	3.7	31%	49%	28	12	10	7
86	Uttar Pradesh	Anpara			40 x 40	0.8	15%	65%	12	7	9	3
87	Uttar Pradesh	Bareily			30 x 30	2.4	25%	63%	20	10	10	5
88	Uttar Pradesh	Firozabad			30 x 30	1.5	11%	43%	15	9	10	4
89	Uttar Pradesh	Gajraula			30 x 30	0.8	16%	43%	13	7	9	3
90	Uttar Pradesh	Gorakhpur			30 x 30	2.3	44%	60%	24	10	10	5
91	Uttar Pradesh	Jhansi			30 x 30	0.9	17%	72%	13	8	10	3
92	Uttar Pradesh	Kanpur		Unnao	40 x 40	4.0	23%	70%	24	13	10	8
93	Uttar Pradesh	Khurja		Bulandshahr	30 x 30	1.2	14%	32%	16	8	10	3
94	Uttar Pradesh	Lucknow		Barabanki	60 x 60	6.4	22%	54%	32	16	10	10
95	Uttar Pradesh	Meerut			30 x 30	2.5	42%	73%	23	10	10	5
96	Uttar Pradesh	Moradabad			30 x 30	2.0	29%	51%	21	10	10	5
97	Uttar Pradesh	Raebareli			30 x 30	1.1	7%	27%	14	8	10	3
98	Uttar Pradesh	Varanasi			40 x 40	4.6	52%	57%	37	13	10	8
99	Uttarakhand	Dehradun			30 x 30	1.1	31%	82%	15	8	10	3
100	Uttarakhand	Kashipur			30 x 30	1.0	22%	46%	16	8	10	3
101	Uttarakhand	Rishikesh		Haridwar	30 x 30	0.8	20%	75%	12	7	9	3
102	West Bengal	Asansol	Durgapur	Ranigunj	60 x 40	3.6	26%	43%	27	12	10	7
103	West Bengal	Haldia			40 x 40	2.2	11%	7%	34	10	10	5
104	West Bengal	Kolkata	Barrackpore, Howrah		60 x 60	20.4	50%	61%	82	20	10	17

www.urbanemissions.info