AIR QUALITY MONITORING

PHASE 2: INNOVATION

PROJECT: AIR QUALITY MONITORING

INTRODUCTION:

Air is a basic requirement for the survival and development of all lives on Earth. It affects healthand influences the development of the economy. Today, due to the development of industrialization,the increase in the number of private cars, and the burning of fossil fuels, air quality is decreasing, withincreasingly serious air pollution. There are many pollutants in the atmosphere, such as SO_2, NO_2,CO_2, NO, CO, NO_x, PM2.5, and PM10. Internationally, a large number of scholars have conductedresearch on air pollution and air quality forecasts, concentrating on the forecasting of contaminants.Air pollution affects the life of a society, and even endangers the survival of mankind. During theIndustrial Revolution, there was a dramatic increase in coal use by factories and households, andthe smog caused significant morbidity and mortality, particularly when combined with stagnantatmospheric conditions. During the Great London Smog of 1952, heavy pollution for 5 days caused atleast 4000 deaths [1,2]. This episode highlighted the relationship between air pollution and humanhealth, yet air pollution continues to be a growing problem in cities and households around the world.

Air pollution is made up of a mixture of gases and particles in harmful amounts that are releasedinto the atmosphere due to either natural or human activities [3]. The sources of pollutants can bedivided into two categories:

1) Natural sources
2) Anthropogenic sources

(1) Natural sources:

Natural pollution sources are natural phenomena that discharge harmful substances or haveharmful effects on the environment. Natural phenomena, such as volcanic eruptions and forest fires,will result in air pollutants, including SO2, CO2, NO2, CO, and sulfate.

(2) Anthropogenic (man-made) sources:

Man-made sources such as the burning of fuels, discharges from industrial production processes,and transportation emissions are the main sources of air pollution. There are many kinds of pollutantsemitted by man-made pollution sources, including hydrogen, oxygen,nitrogen, sulfur, metal compounds,and particulate matter.With the increasing world population and the developing world economy, the demand for energyin the world has increased dramatically. The large-scale use of fossil energy globally has also ledto a series of environmental problems that have received much attention due to their detrimentaleffects on human health and the environment [3–5]. Air pollution is a fundamental problem in manyparts of the world, with two important concerns: the impact on human health, such as cardiovasculardiseases, and the impact on the environment, such as acid rain, climate change, and global warming.

The construction of this paper

Air Pollution Assessment:

In recent years, air pollution accidents have occurred frequently, which have damaged the economy and human life. To assess the extent of the damage, air pollution control must be evaluatedin order to have a quantitative understanding of pollution.Int. J. Environ. Res. Public

Health **2018**, 15, 780 6 of 44The assessment of air pollution is identify and measure the degree and scope of damage causedby environmental pollution cover the economic, legal, technical and other means reasonably [35–37].

Two of the more mature assessment methods will be described. The market value methodis a type of cost benefit analysis method. It uses the change of product yield and profit causedby the environmental quality change to measure the economic loss related to the environmentalquality change.Environmental pollution and damage caused by air pollution can be prevented, restored, orreplaced by the existing environmental functions. Therefore, the cost of preventing, restoring, orreplacing the original functional protection facilities can be used to estimate the loss caused bypollution or damage to the environment. This method is called the engineering cost method.The main equation and the meaning of the variables in those methods are given in Table 1, andthe flowchart of the assessment methods is given in Figure 2.

Two of the more mature assessment methods will be described. The market value method is a typeof cost benefit analysis method. It uses the change of product yield and profit caused by theenvironmental quality change to measure the economic loss related to the environmental qualitychange.

Environmental pollution and damage caused by air pollution can be prevented, restored, orreplaced by the existing environmental functions. Therefore, the cost of preventing, restoring, orreplacing the original functional protection facilities can be used to estimate the loss caused bypollution or damage to the environment. This method is called the engineering cost method.

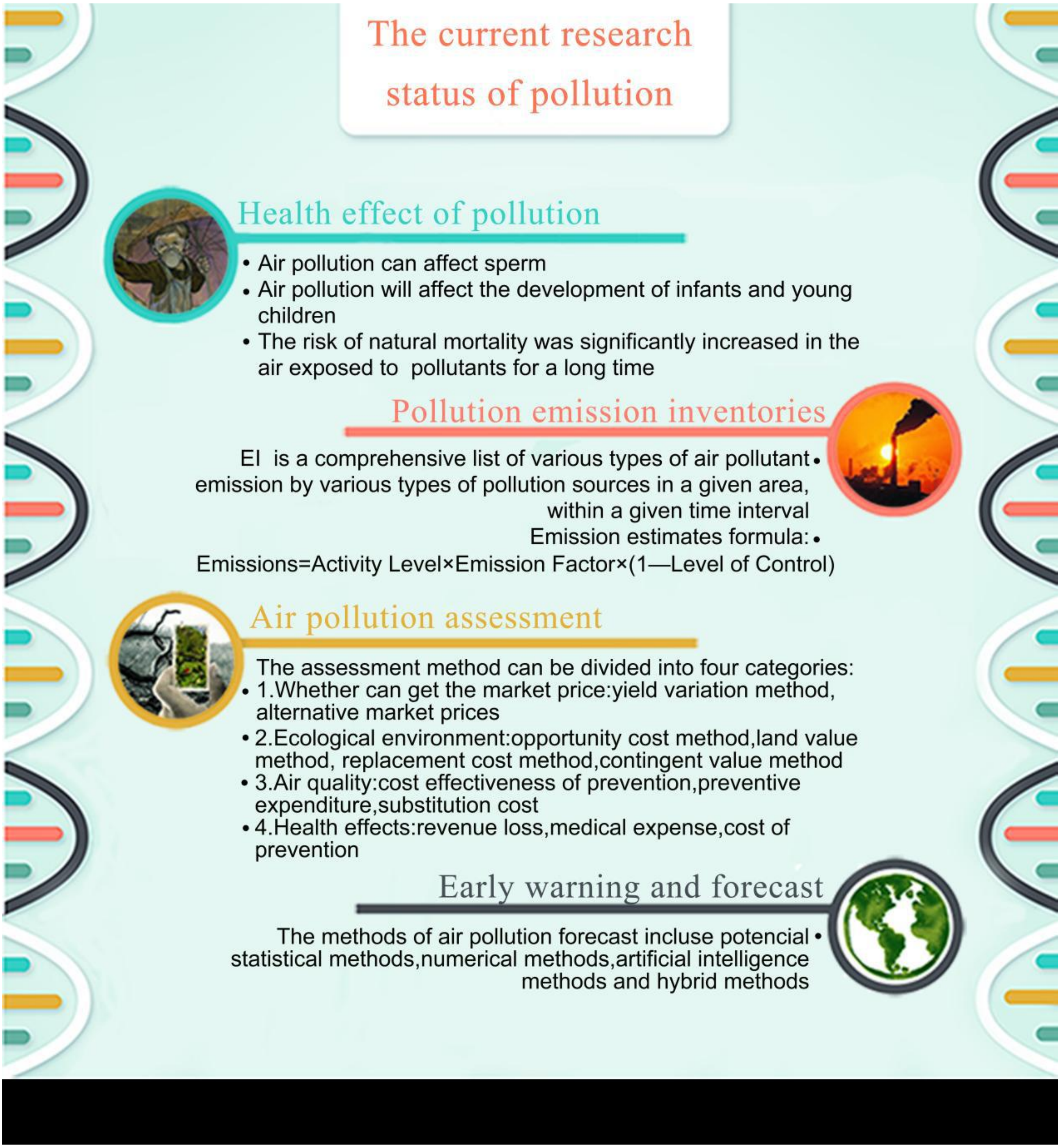
Air Pollution Early Warning and Forecast:

The most important function of air pollution early warning systems is to report the air quality torelevant departments when the air quality reaches the early warning standard. A complete pollution.

Warning system includes the pollutant, resource, and scope of influence [44]. Air quality forecasting is an effectiveway of protecting public health by providing an earlywarning against harmful air pollutants [9]. Urban air pollution events can be forecasted by meteorological elements to provide an early warning. Therefore, in the face of more and more urban air pollution incidents, in addition to risk prevention management and emergency measures, air pollution forecastsshould also include the emergency warnings as an important part of the whole emergency system.

The early warning system for air pollution is triggered before the heavy pollution of urbanair, according to the forecast of meteorological elements. Corresponding emergency measures areinitiated as early as possible to reduce the discharge of pollutants and mitigate the consequences.

Many countries have early warning systems for pollution. For example, the Air Quality Index (AQI)value is an index for the classification of the early warning level in China, and the early warning levelis determined according to the upper limit of the pollution forecast. Therefore, the forecasting of airpollution as the basis for pollution warning systems and pollution control should be highly valued by all countries.



THE CURRENT RESURCH STATUS OF POLLUTION

TYPES OF PREDICTIVE MODELING TO FORECAST AIR QUALITY TRENDS:

Linear regression modeling (LM),
Support vector regression (SVR),
Artificial neural network (ANN),
Decision tree (DT),
Random forest (RF),
Extreme gradient boosting tree (XGB)

Linear regression modeling (LM):

Air quality is predicted using the R squared value. R square determines the proportion of variance in the dependent variable of the system that can be explained by the independent variable. It is a statistical measure in a regression model. It is also called a coefficient of determination.

Support vector regression (SVR):

The is paper provides Support Vector Machine (SVM) model to forecast the quality of air with AQI for the upcoming 15 hours. The proposed model predicts the amount of pollutants such as PM 2.5, PM10, NO, NO2, NH3 and shows the better performance than linear prediction models when compared in terms of RMSE.

Artificial neural network (ANN):

Based on the analysis conducted, model with neural network structure 7-20-4 produces the best performance in the prediction of air quality compared to the first model based on the values of R and the prediction accuracy.

Decision tree (DT):

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

Random forest (RF):

In an urban sensing system, an algorithm (RAQ) based on a random forest concept is proposed to predict the urban area air quality through the use of historical air quality data, meteorology data, historical traffic and road status as well as POI distribution information

Extreme gradient boosting tree (XGB):

What is XGBoost? XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

PROBLEM:

IP [1] :

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
```

```
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

IP [2] :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
/opt/conda/lib/python3.10/site-packages/scipy/__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.23.5
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

IP [3] :

```
airqualitydatset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29531 entries, 0 to 29530
Data columns (total 16 columns):
```

OUTPUT:

```
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   City        29531 non-null  object
 1   Date        29531 non-null  object
 2   PM2.5       24933 non-null  float64
 3   PM10        18391 non-null  float64
 4   NO          25949 non-null  float64
 5   NO2         25946 non-null  float64
 6   NOx         25346 non-null  float64
 7   NH3         19203 non-null  float64
 8   CO          27472 non-null  float64
 9   SO2         25677 non-null  float64
 10  O3          25509 non-null  float64
 11  Benzene     23908 non-null  float64
 12  Toluene     21490 non-null  float64
 13  Xylene      11422 non-null  float64
 14  AQI         24850 non-null  float64
 15  AQI_Bucket  24850 non-null  object
dtypes: float64(13), object(3)
memory usage: 3.6+ MB
```

IP [4] :

```
airqualitydatset["City"].unique()
```

OUTPUT:

```
array(['Ahmedabad', 'Aizawl', 'Amaravati', 'Amritsar', 'Bengaluru',
       'Bhopal', 'Brajrajnagar', 'Chandigarh', 'Chennai', 'Coimbatore',
       'Delhi', 'Ernakulam', 'Gurugram', 'Guwahati', 'Hyderabad',
       'Jaipur', 'Jorapokhar', 'Kochi', 'Kolkata', 'Lucknow', 'Mumbai',
       'Patna', 'Shillong', 'Talcher', 'Thiruvananthapuram',
       'Visakhapatnam'], dtype=object)
```

IP [5] :

```
airqualitydatset = airqualitydatset.dropna()
airqualitydatset  # Dropped all null values in the Dataset as we could not average.
```

OUTPUT:

	city	date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	BENENE	TOLUENE	AQI	AQI_Bucket
0	Ahamedabad	2015.01.01	Nan	NAn	12.5	18.22	17.5	NaN	0.92	27.64	133.36	0.00	0.02	NaN	NaN
1	Ahamedabad	2015.01.02	Nan	Nan	15.7	15.69	16.46	NaN	0.97	24.55	34.06	5.50	5.50	NaN	NaN
2	Ahamedabad	2015.01.03	Nan	Nan	17.40	0.92	29.70	NaN	17.40	29.07	30.70	16.40	16.40	NaN	NaN
3	Ahamedabad	2015.01.04	Nan	Nan	1.70	0.97	17.97	NaN	1.70	18.59	36.08	10.14	10.14	NaN	NaN
4	Ahamedabad	2015.01.05	Nan	Nan	22.10	21.42	37.76	NaN	22.10	39.31	39.31	18.89	18.89	NaN	NaN

IP [6] :

```
airqualitydatset.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6236 entries, 2123 to 29529
```

OUTPUT:

```
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
```

```
 0   City        6236 non-null   object
 1   Date        6236 non-null   object
 2   PM2.5       6236 non-null   float64
 3   PM10        6236 non-null   float64
 4   NO          6236 non-null   float64
 5   NO2         6236 non-null   float64
 6   NOx         6236 non-null   float64
 7   NH3         6236 non-null   float64
 8   CO          6236 non-null   float64
 9   SO2         6236 non-null   float64
 10  O3          6236 non-null   float64
 11  Benzene     6236 non-null   float64
 12  Toluene     6236 non-null   float64
 13  Xylene      6236 non-null   float64
 14  AQI         6236 non-null   float64
 15  AQI_Bucket  6236 non-null   object
dtypes: float64(13), object(3)
memory usage: 828.2+ KB
```

IP [7] :

```
airqualitydatset.mean()
```

```
/tmp/ipykernel_20/855534917.py:1: FutureWarning: The default value of numeric_only in DataFra
me.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_
only=None' is deprecated. Select only valid columns or specify the value of numeric_only to silenc
e this warning.
  airqualitydatset.mean()
```

OUTPUT:

```
PM2.5       61.327365
PM10       123.418321
NO          17.015191
NO2         31.708190
NOx         32.448956
NH3         20.737070
CO           0.984344
SO2         11.514426
O3          36.127691
Benzene      3.700361
Toluene     10.323696
Xylene       2.557439
AQI        140.510103
dtype: float64
```

IP [8] :

```
airqualitydatset.max()
```

OUTPUT:

```
City          Visakhapatnam
Date             2020-07-01
```

```
PM2.5             639.19
PM10              796.88
NO                159.22
NO2               140.17
NOx               224.09
NH3               166.7
CO                16.23
SO2               70.39
O3                162.33
Benzene           64.44
Toluene           103.0
Xylene            125.18
AQI               677.0
AQI_Bucket        Very Poor
dtype: object
```

IP [9] :

```
airqualitydatset.min()
```

OUTPUT:

```
City          Amaravati
Date          2015-01-01
PM2.5         2.0
PM10          7.8
NO            0.25
NO2           0.17
NOx           0.17
NH3           0.12
CO            0.0
SO2           0.71
O3            1.55
Benzene       0.0
Toluene       0.0
Xylene        0.0
AQI           23.0
AQI_Bucket    Good
dtype: object
```

IP [10] :

```
#mostreadingAQI.value_counts()
#mostreadingAQI = mostreadingAQI[['City', 'AQI_Bucket']]
#mostreadingAQIGood = mostreadingAQI.where(mostreadingAQI["City"] == "Amaravati")
#mostreadingAQI.dropna()
#mostreadingAQI = mostreadingAQI.where(mostreadingAQI["AQI_Bucket"] == "Good")

#mostreadingAQI
#mostreadingAQI["AQI_Bucket"].value_counts()
```

IP [11] :

```
mostreadingAQI = airqualitydatset[["City","AQI_Bucket"]]
mostreadingAQI.sort_values(['City'],inplace=True,ascending=True)
mostreadingAQI.groupby(['AQI_Bucket'])
mostreadingAQI
#mostreadingAQI.sort_values(['City'],inplace=True,ascending=True)
mostreadingAQI.value_counts(['City', 'AQI_Bucket'])
```

```
/tmp/ipykernel_20/565188966.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  mostreadingAQI.sort_values(['City'],inplace=True,ascending=True)
```

OUTPUT:

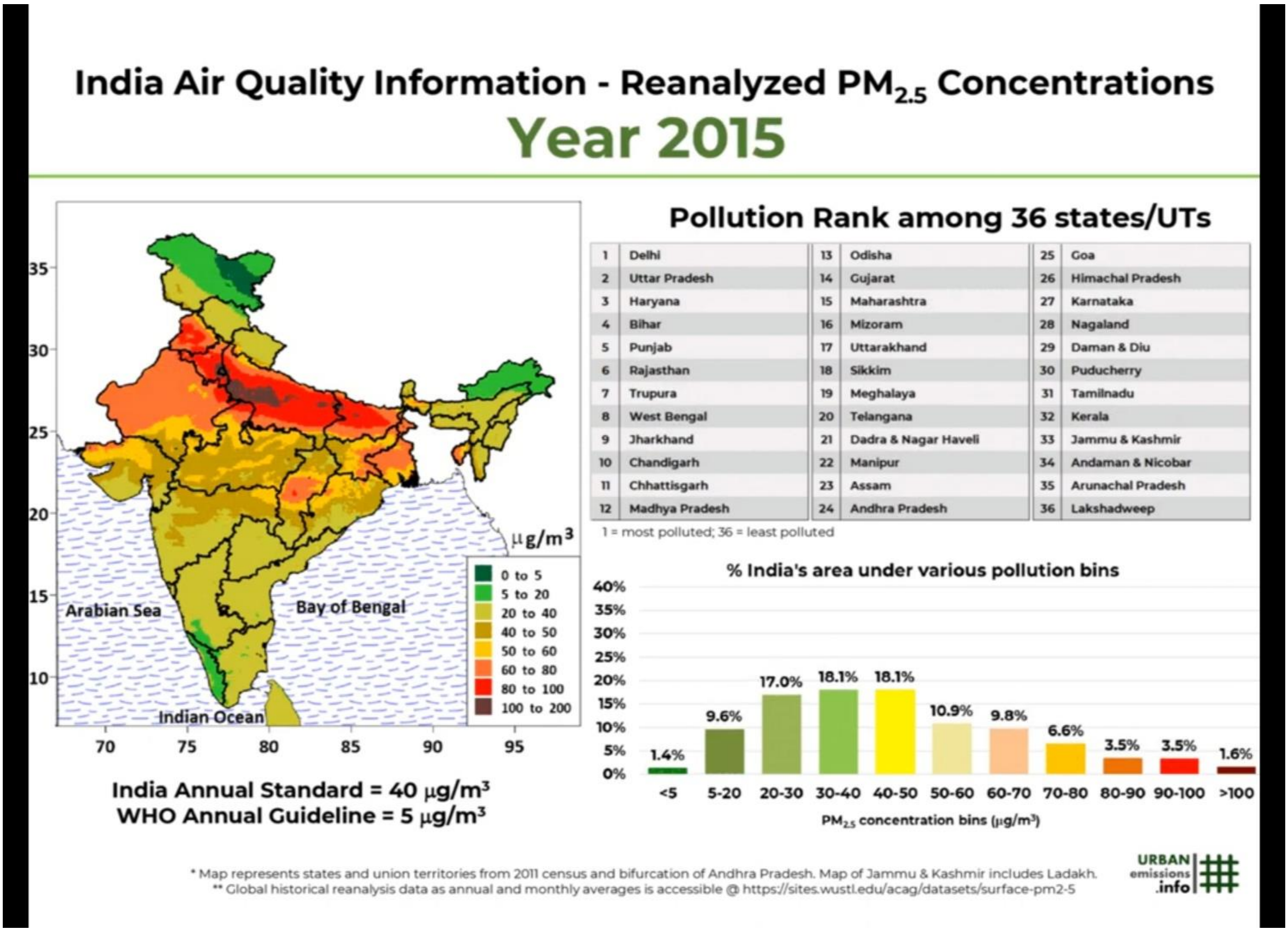
```
City              AQI_ Bucket
Hyderabad  Moderate    810
                Satisfactory   645
Visakhapatnam  Moderat  555
            Satisfactory     438
Delhi          Moderate    360
            Poor           328
            Very Poor       315
Amaravati      Satisfactory    305
Amritsar       Moderate        252
            Satisfactory    252
Amaravati      Moderate        191
Chandigarh     Satisfactory    145
Kolkata        Satisfactory    139
Hyderabad      Good            126
Delhi          Severe          117
            Satisfactory    104
Patna          Moderate        103
Amaravati      Good            101
Gurugram       Moderate         97
Kolkata        Good             95
            Moderate        87
            Poor            71
Visakhapatnam  Poor             70
Chandigarh     Moderate         66
Visakhapatnam  Good             50
Amritsar       Poor             49
Chandigarh     Good             48
Amritsar       Very Poor        47
Patna          Poor             42
Amaravati      Poor             41
Amritsar       Good             34
Patna          Satisfactory     31
Hyderabad      Poor             30
Gurugram       Satisfactory     20
Visakhapatnam  Very Poor        18
Chandigarh     Poor             15
Patna          Very Poor        14
Amaravati      Very Poor         8
Hyderabad      Severe            4
```

```
             Very Poor        3
Chandigarh     Very Poor        3
Kolkata       Very Poor        2
Amritsar      Severe           2
Gurugram       Poor            2
Patna         Severe           1
dtype: int64
```

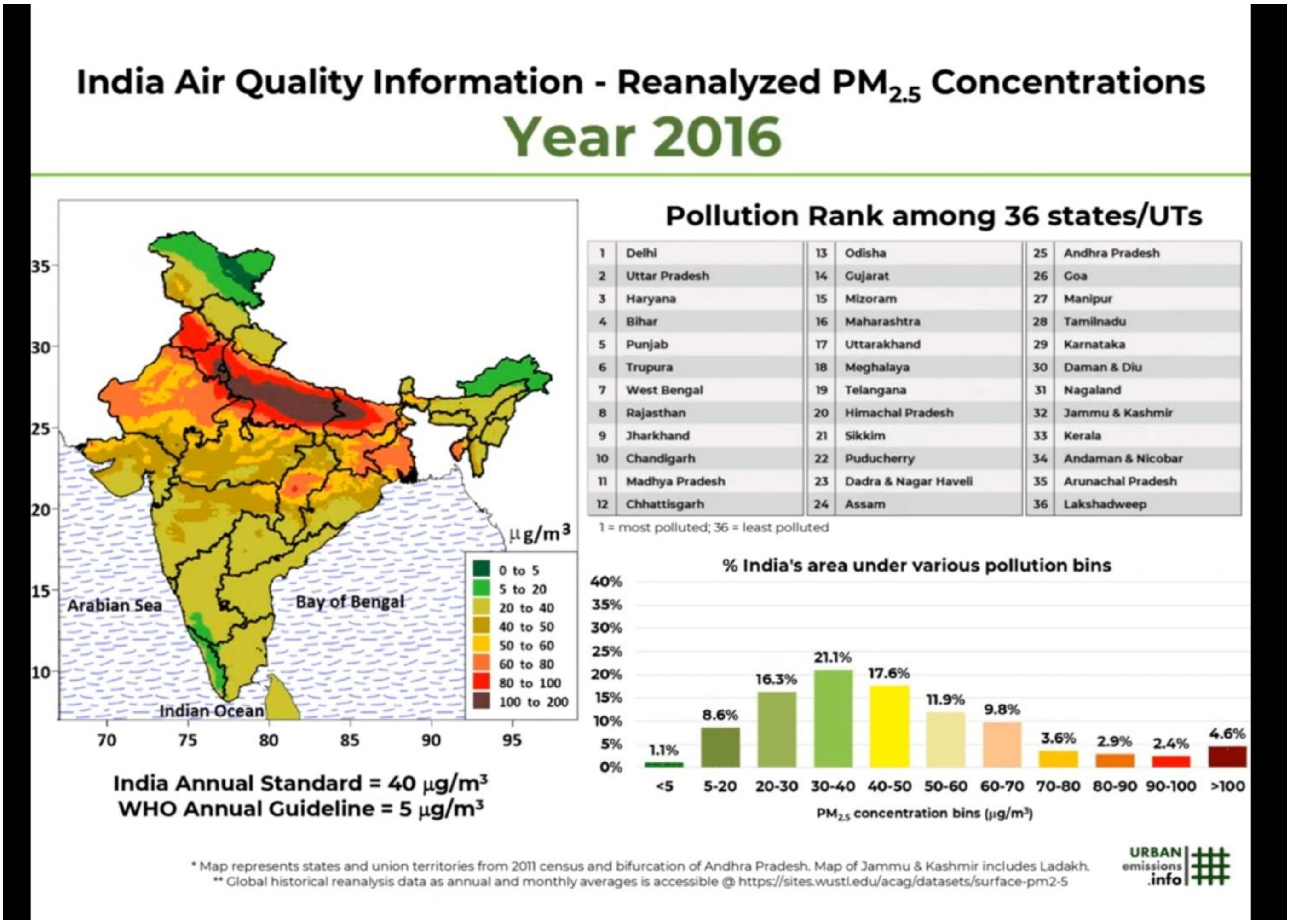
INDIA AIR QUALITY INFORMATION REANALYZED PM2.5 CONCENTRATIONS:

YEAR – 2015 TO 2020

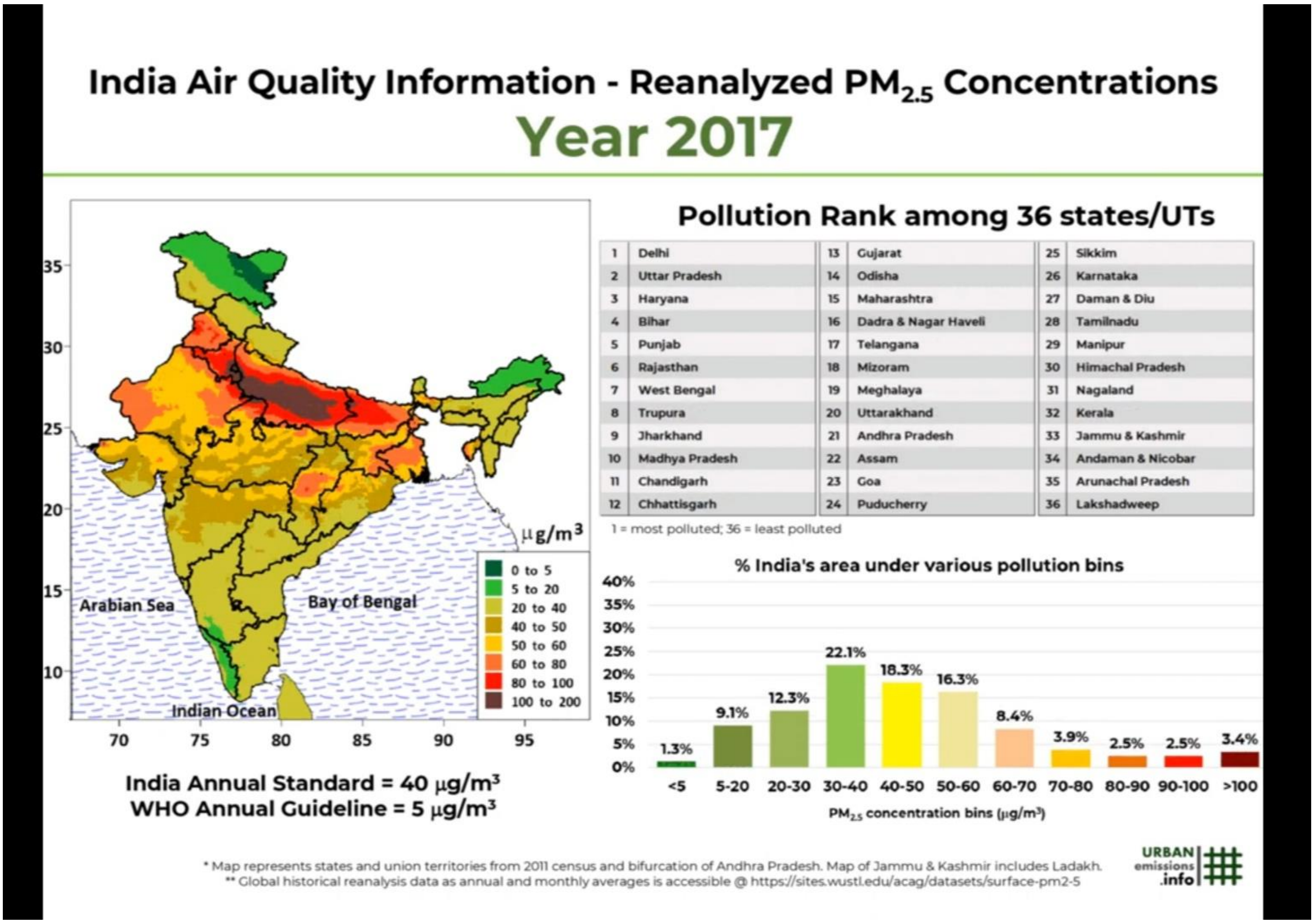
YEAR - 2015



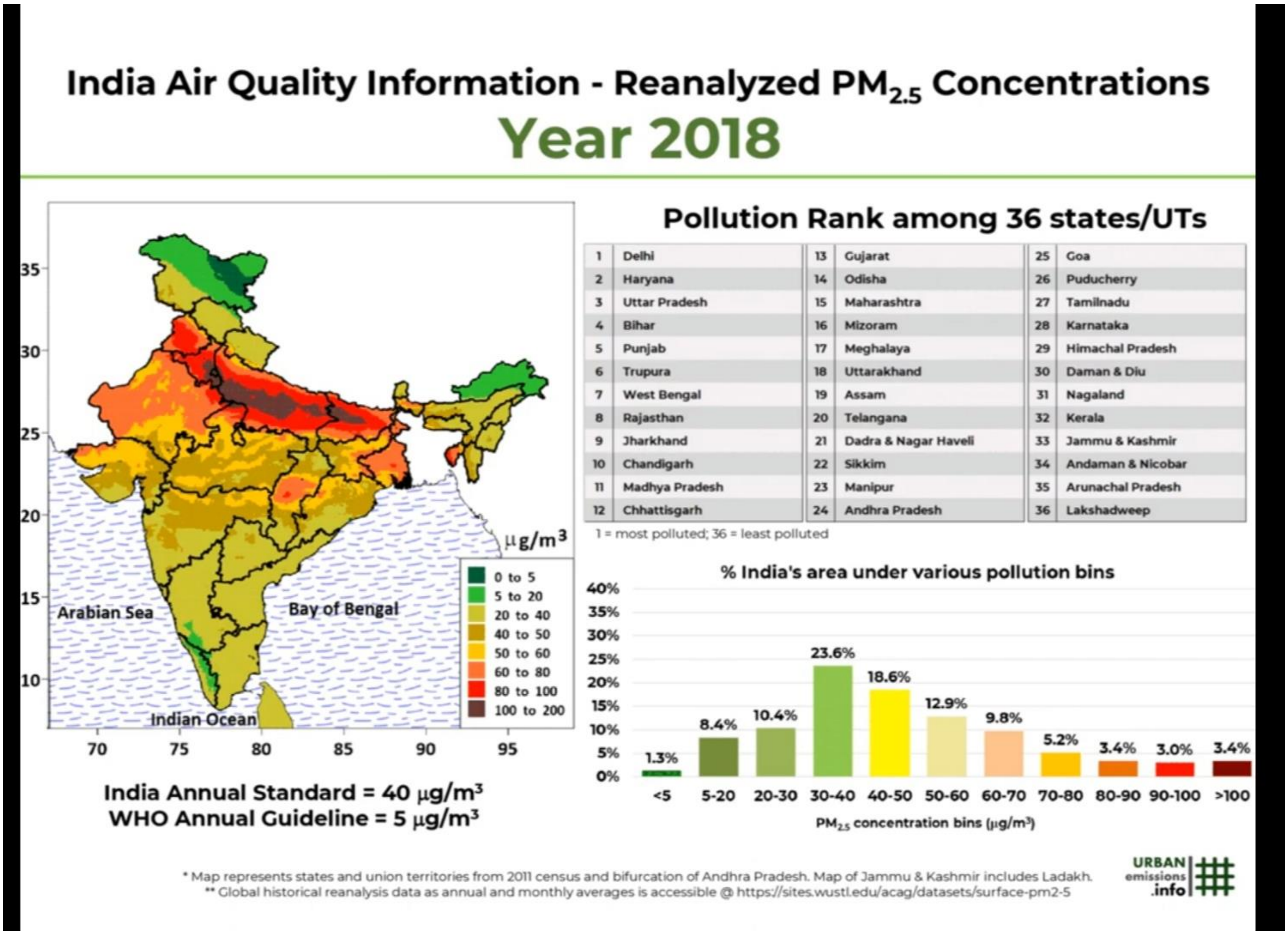
YEAR – 2016



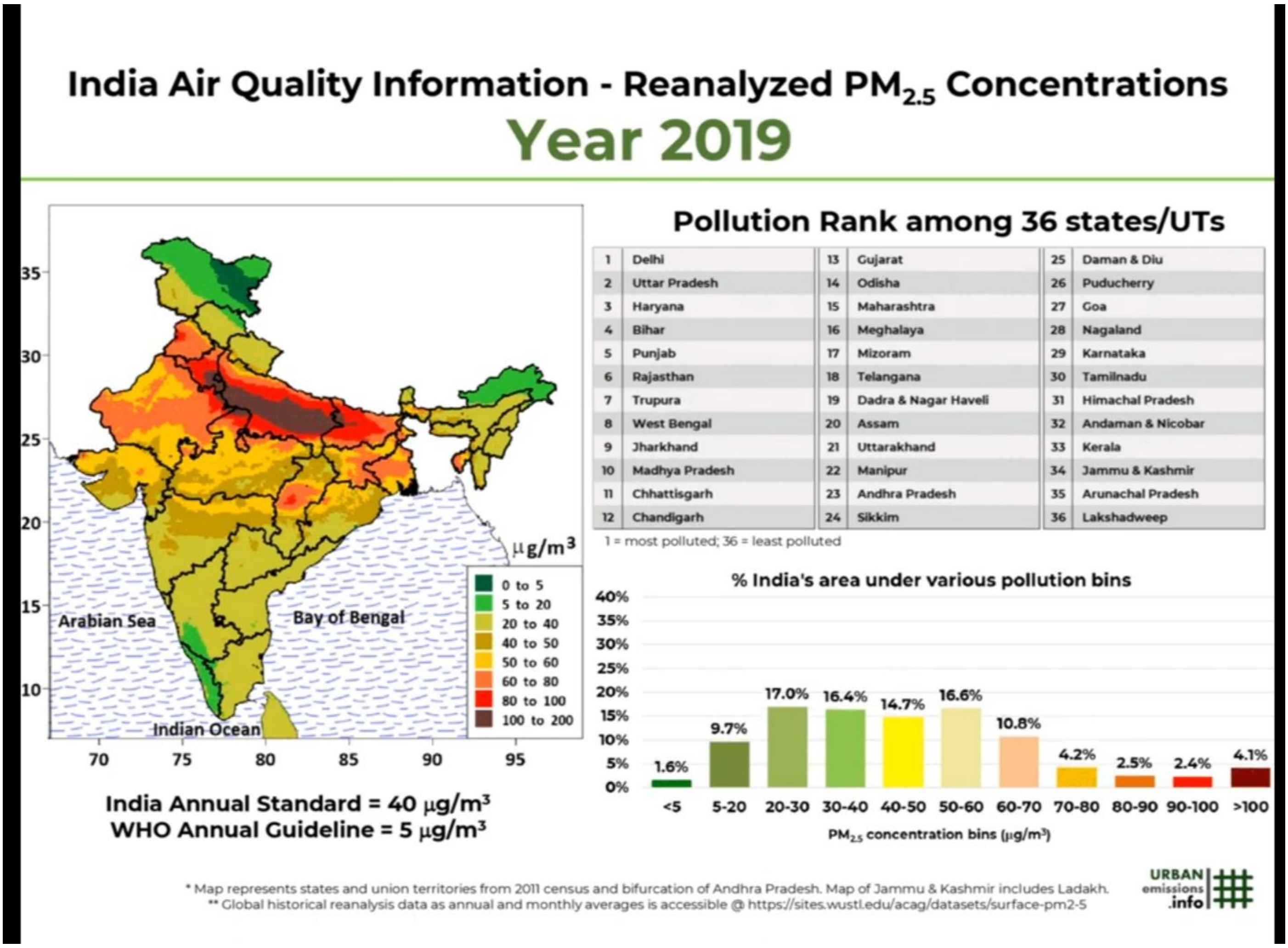
YEAR – 2017


India Air Quality Information - Reanalyzed $PM_{2.5}$ Concentrations
Year 2017
Pollution Rank among 36 states/UTs
1 Delhi
2 Uttar Pradesh
3 Haryana
4 Bihar
5 Punjab
6 Rajasthan
7 West Bengal
8 Trupura
9 Jharkhand
10 Madhya Pradesh
11 Chandigarh
12 Chhattisgarh
13 Gujarat
14 Odisha
15 Maharashtra
16 Dadra & Nagar Haveli
17 Telangana
18 Mizoram
19 Meghalaya
20 Uttarakhand
21 Andhra Pradesh
22 Assam
23 Goa
24 Puducherry
25 Sikkim
26 Karnataka
27 Daman & Diu
28 Tamilnadu
29 Manipur
30 Himachal Pradesh
31 Nagaland
32 Kerala
33 Jammu & Kashmir
34 Andaman & Nicobar
35 Arunachal Pradesh
36 Lakshadweep
1 = most polluted; 36 = least polluted
Arabian Sea
Bay of Bengal
Indian Ocean
$\mu g/m^3$
0 to 5
5 to 20
20 to 40
40 to 50
50 to 60
60 to 80
80 to 100
100 to 200
% India's area under various pollution bins
1.3% 9.1% 12.3% 22.1% 18.3% 16.3% 8.4% 3.9% 2.5% 2.5% 3.4%
<5 5-20 20-30 30-40 40-50 50-60 60-70 70-80 80-90 90-100 >100
$PM_{2.5}$ concentration bins ($\mu g/m^3$)
India Annual Standard = 40 $\mu g/m^3$
WHO Annual Guideline = 5 $\mu g/m^3$
* Map represents states and union territories from 2011 census and bifurcation of Andhra Pradesh. Map of Jammu & Kashmir includes Ladakh.
** Global historical reanalysis data as annual and monthly averages is accessible @ https://sites.wustl.edu/acag/datasets/surface-pm2-5
URBAN emissions .info

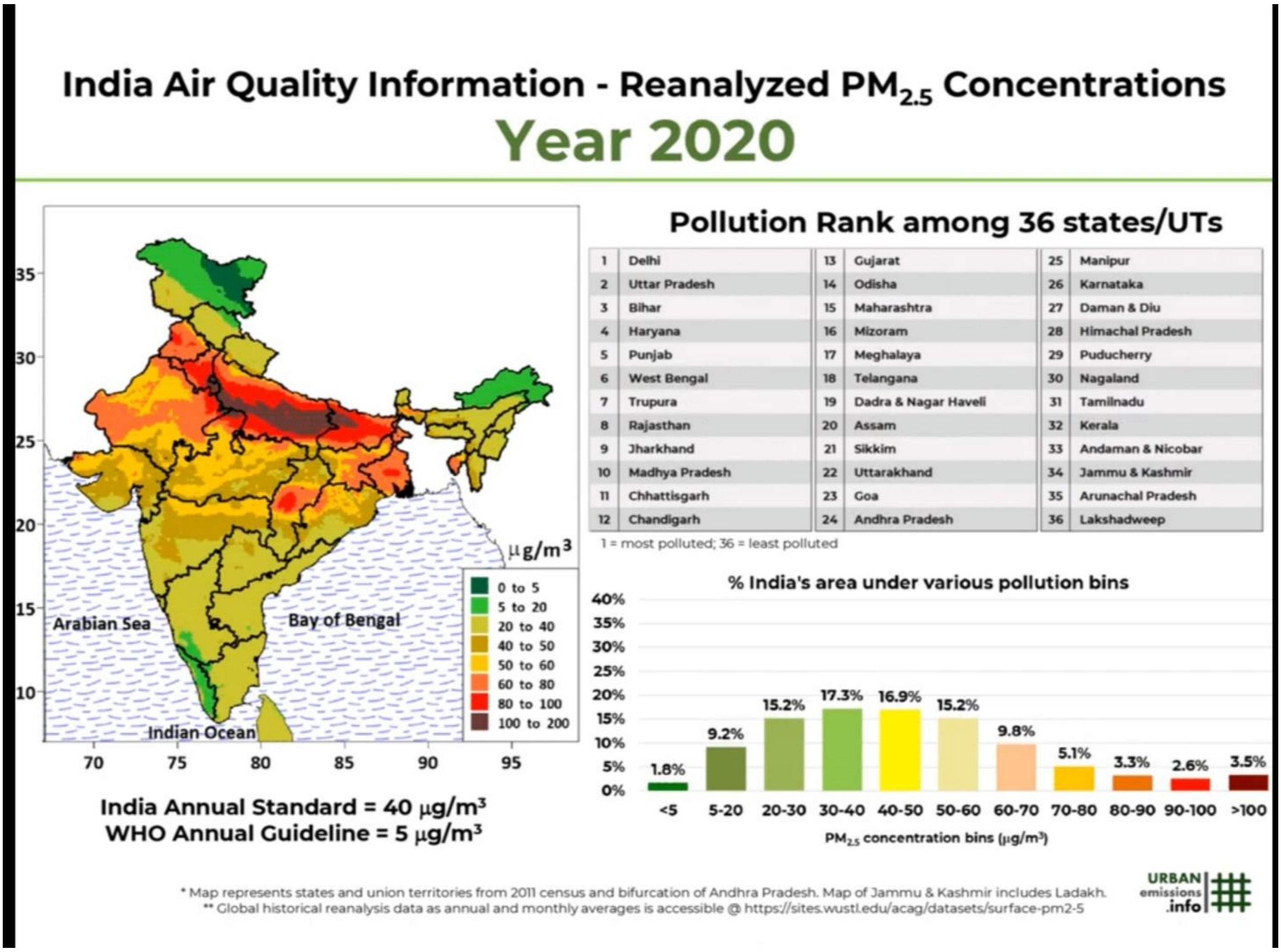
YEAR – 2018


India Air Quality Information - Reanalyzed PM2.5 Concentrations
Year 2018
Pollution Rank among 36 states/UTs
1 Delhi
2 Haryana
3 Uttar Pradesh
4 Bihar
5 Punjab
6 Trupura
7 West Bengal
8 Rajasthan
9 Jharkhand
10 Chandigarh
11 Madhya Pradesh
12 Chhattisgarh
13 Gujarat
14 Odisha
15 Maharashtra
16 Mizoram
17 Meghalaya
18 Uttarakhand
19 Assam
20 Telangana
21 Dadra & Nagar Haveli
22 Sikkim
23 Manipur
24 Andhra Pradesh
25 Goa
26 Puducherry
27 Tamilnadu
28 Karnataka
29 Himachal Pradesh
30 Daman & Diu
31 Nagaland
32 Kerala
33 Jammu & Kashmir
34 Andaman & Nicobar
35 Arunachal Pradesh
36 Lakshadweep
1 = most polluted; 36 = least polluted
Arabian Sea
Bay of Bengal
Indian Ocean
μg/m³
0 to 5
5 to 20
20 to 40
40 to 50
50 to 60
60 to 80
80 to 100
100 to 200
% India's area under various pollution bins
<5: 1.3%
5-20: 8.4%
20-30: 10.4%
30-40: 23.6%
40-50: 18.6%
50-60: 12.9%
60-70: 9.8%
70-80: 5.2%
80-90: 3.4%
90-100: 3.0%
>100: 3.4%
PM2.5 concentration bins (μg/m³)
India Annual Standard = 40 μg/m³
WHO Annual Guideline = 5 μg/m³
* Map represents states and union territories from 2011 census and bifurcation of Andhra Pradesh. Map of Jammu & Kashmir includes Ladakh.
** Global historical reanalysis data as annual and monthly averages is accessible @ https://sites.wustl.edu/acag/datasets/surface-pm2-5
URBAN emissions .info

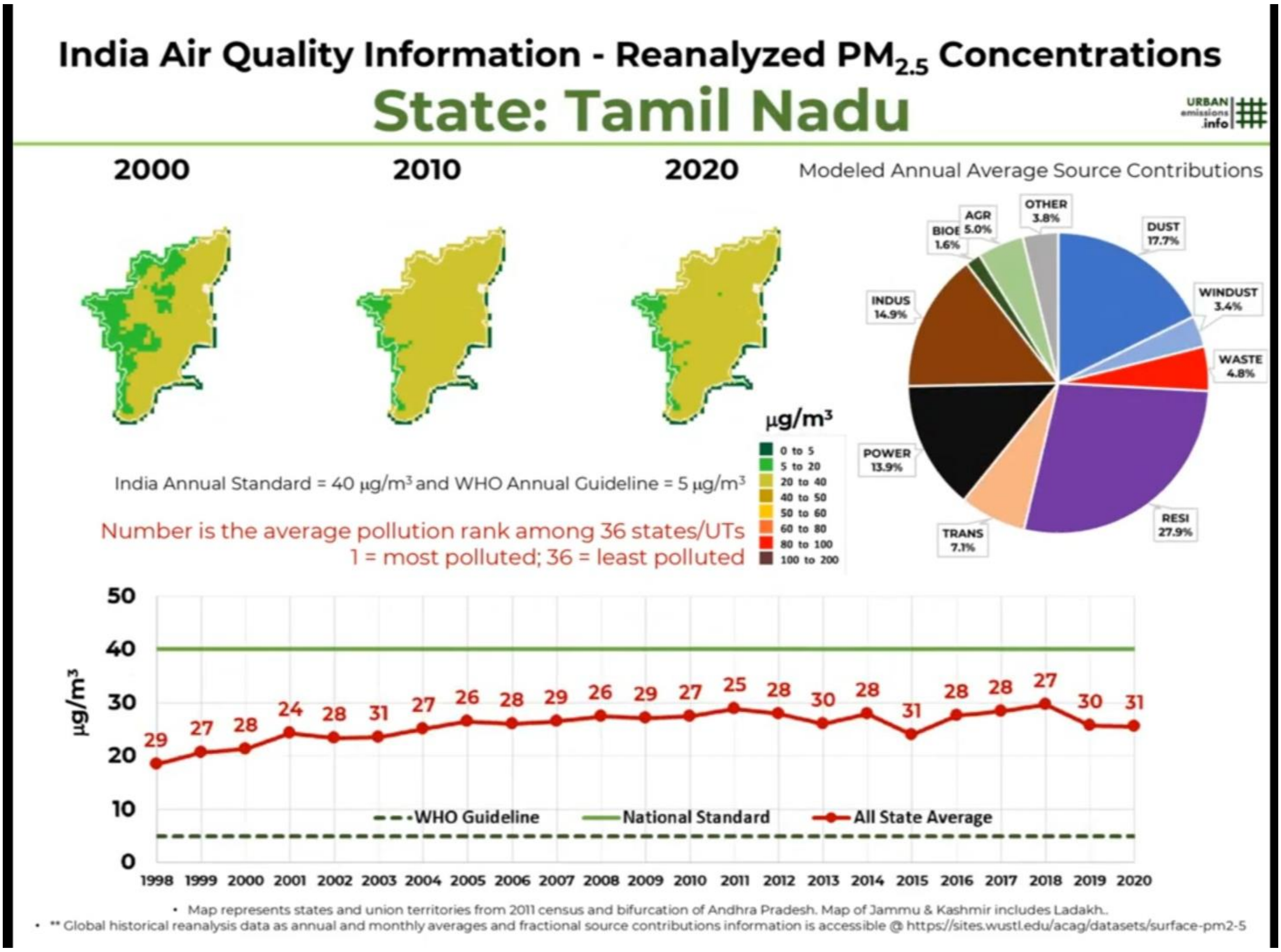
YEAR – 2019


India Air Quality Information - Reanalyzed $PM_{2.5}$ Concentrations
Year 2019
Pollution Rank among 36 states/UTs
1 Delhi
2 Uttar Pradesh
3 Haryana
4 Bihar
5 Punjab
6 Rajasthan
7 Trupura
8 West Bengal
9 Jharkhand
10 Madhya Pradesh
11 Chhattisgarh
12 Chandigarh
13 Gujarat
14 Odisha
15 Maharashtra
16 Meghalaya
17 Mizoram
18 Telangana
19 Dadra & Nagar Haveli
20 Assam
21 Uttarakhand
22 Manipur
23 Andhra Pradesh
24 Sikkim
25 Daman & Diu
26 Puducherry
27 Goa
28 Nagaland
29 Karnataka
30 Tamilnadu
31 Himachal Pradesh
32 Andaman & Nicobar
33 Kerala
34 Jammu & Kashmir
35 Arunachal Pradesh
36 Lakshadweep
1 = most polluted; 36 = least polluted
Arabian Sea
Bay of Bengal
Indian Ocean
μg/m³
0 to 5
5 to 20
20 to 40
40 to 50
50 to 60
60 to 80
80 to 100
100 to 200
35
30
25
20
15
10
70
75
80
85
90
95
India Annual Standard = 40 μg/m³
WHO Annual Guideline = 5 μg/m³
% India's area under various pollution bins
40%
35%
30%
25%
20%
15%
10%
5%
0%
1.6%
9.7%
17.0%
16.4%
14.7%
16.6%
10.8%
4.2%
2.5%
2.4%
4.1%
<5
5-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
>100
$PM_{2.5}$ concentration bins (μg/m³)
* Map represents states and union territories from 2011 census and bifurcation of Andhra Pradesh. Map of Jammu & Kashmir includes Ladakh.
** Global historical reanalysis data as annual and monthly averages is accessible @ https://sites.wustl.edu/acag/datasets/surface-pm2-5
URBAN emissions .info

YEAR – 2020



TAMIL NADU AIR QUALITY INFORMATION – REANALYZED PM2.5 CONCENTRATION:



CUNCLUTION AND FUTURE WORK (PHASE 2):

Project conclution:

In the phase 2 coclution, we wil summarize the key finding and insides from the advanced regression techniques. We will reiterate the impact of these techniques on improving the Air quality monitoring.