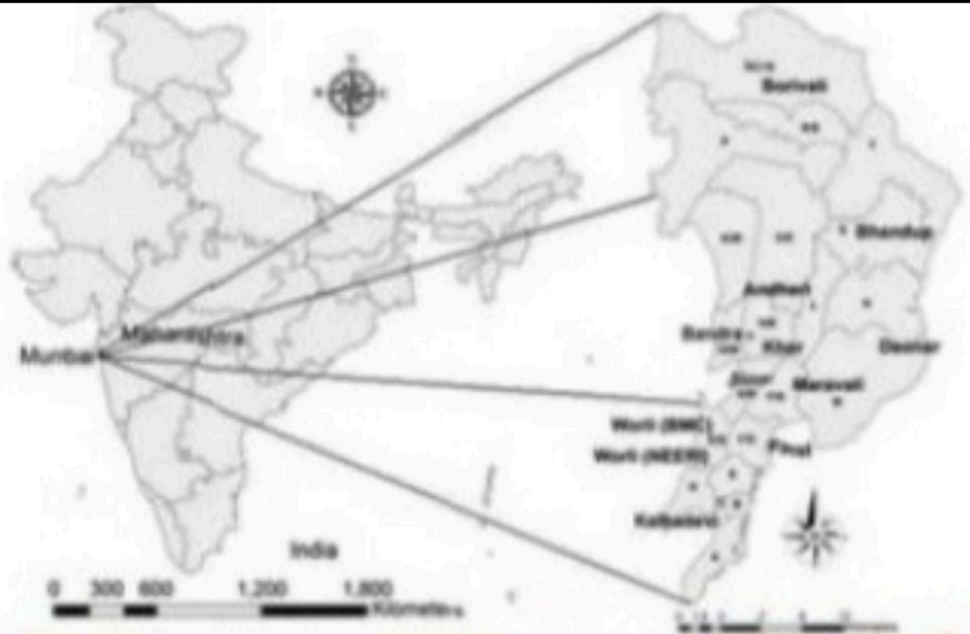
Air quality analysis and prediction in tamilnadu

Introduction :

As the largest growing industrial nation, India is producing record amount of pollutants specifically Co2, pm2.5 etc and other harmful aerial contaminants. Air quality of a particular state or a country is a measure on the effect of pollutants on the respected regions, as per the Indian air quality standard pollutants are indexed in terms of their scale, these air quality indexes indicates the levels of major pollutants on the atmosphere. There are various atmospheric gases which causes pollution on our environment.



Study Area

Mumbai is the capital of the state of Maharashtra in India

One of the objectives of this work was to estimate temporal and spatial variation of air pollution over the region

ways to Reduce Air pollution in Cities:

- Ways to Reduce Air Pollution in cities are using public transportion, Control emissions from industries, promote green spaces and Electric vehicle
- Air pollution is a major problem affecting cities worldwide.Theconcentration of harmful pollutants in cities poses serious health risks for living beings & contributes to environmental degradation

Air Quality Selected Locations

For Date - 4 October 2023

City / State	PM2.5	PM10	AQI
Korba	59	62	98
Delhi	67	76	124

Example for air quality analysis

3. Data Cleaning

3.1. Handling Outliers

From Table 4, the maximum and minimum possible values for the weather dataset can be inferred as shown below:

Band	Dew (°C)	Temperature (°C)	Wind Direction (°)	Wind Speed (Km/Hour)	Humidity (%)
Minimum	-98.2	-93.2	1	0	0
Maximum	36.8	61.8	360	324	100

Table 4: Limits for weather features

Since these values are obtained from the reference for the weather dataset, any value outside these limits can be considered as outliers and confidently discarded.
The London Air Quality Network (LAQN) specifies the following values for the different bands of pollutants [4]:

Band	Index	O_3 (µg/m^3)	NO_2 (µg/m^3)	PM2.5 (µg/m^3)
Low	1-3	0-100	0-200	0-35
Moderate	4-6	101-160	201-400	36-53
High	7-9	161-240	401-600	54-70
Very High	10	241 or more	601 or more	71 or more

Substance	Time weighted average	Averaging time
Cadmium	5 ng/1m3	1 yr
Carbon monoxide	100 mg/m3	15 mins
	60 mg/m3	30 mins
	30 mg/m3	1 hr
	10 mg/m3	8 hrs
Nitrogen dioxide	200 μg/m3	1 hr
	40 μg/m	1 yr
Ozone	150-200 μg/m3	1 hr
	100-120 μg/m3	8 hrs
SO2	500 μg/m3	10 mins
	350 μg/m3	1 hr

Air pollution Monitoring in Tamilnadu:

The nation air quality monitoring programme sponsored by the central pollution control board since 1990 has generate atabase over last 14 years in 10 major Indian cities

The trend analysis showed that suspended particulate matter (SPM) exceeds the cpcd standard in all the cities of the time throughout the year

The concentration ratio of p<10 fraction (human respirable particles)to the total spam varies between 30% to 60%,with coastal cities showing highest percentages

Pollutant	Mechanism	Potential health effects
Particules (small particles less than 10 microns, and particularly less than 2.5 microns aerodynamic diameter)	• Acute: bronchial irritation, inflammation and increased reactivity • Reduced mucociliary clearance • Reduced macrophage response and (?) reduced local immunity • (?) Fibrotic reaction	• Wheezing, exacerbation of asthma • Respiratory infections • Chronic bronchitis and chronic obstructive pulmonary disease • Exacerbation of chronic obstructive pulmonary disease
Carbon monoxide	• Binding with haemoglobin to produce carboxy haemoglobin, which reduces oxygen delivery to key organs and the developing fetus.	• Low birth weight (fetal carboxy-haemoglobin 2–10% or higher) • Increase in perinatal deaths
Polycyclic aromatic hydrocarbons, e.g. benzo[*a*]pyrene	• Carcinogenic	• Lung cancer • Cancer of mouth, nasopharynx and larynx
Nitrogen dioxide	• Acute exposure increases bronchial reactivity • Longer term exposure increases susceptibility to bacterial and viral lung infections	• Wheezing and exacerbation of asthma • Respiratory infections • Reduced lung function in children
Sulphur dioxide	• Acute exposure increases bronchial reactivity • Longer term: difficult to dissociate from effects of particles	• Wheezing and exacerbation of asthma • Exacerbation of chronic obstructive pulmonary disease, cardiovascular disease
Biomass smoke condensates including polycyclic aromatics and metal ions	• Absorption of toxins into lens, leading to oxidative changes	• Cataract

Program:

```
 # data processing, CSV file I/O (e.g. pd.read_csv)

yplot as plt
ns


e.figsize'] = (10, 7)



ings('ignore')

are available in the "../input/" directory.
ing this (by clicking run or pressing Shift+Enter) will list the files in the input directory


./input"))

rite to the current directory are saved as output.
```

```
['lat-lon-indianstates', 'india-air-quality-data', 'indian-states-lat-lon']
```

```
['lat-lon-indianstates', 'india-air-quality-data', 'indian-states-lat-lon']
```

In [2]:
```
data=pd.read_csv('../input/india-air-quality-data/data.csv',encoding="ISO-8859-1")
data.fillna(0, inplace=True)
data.head()
```

Out [2]:

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm	location_monitoring_statio
0	150	February - M021990	Andhra Pradesh	Hyderabad	0	Residential, Rural and other Areas	4.8	17.4	0.0	0.0	0
1	151	February - M021990	Andhra Pradesh	Hyderabad	0	Industrial Area	3.1	7.0	0.0	0.0	0
2	152	February - M021990	Andhra Pradesh	Hyderabad	0	Residential, Rural and other Areas	6.2	28.5	0.0	0.0	0
3	150	March - M031990	Andhra Pradesh	Hyderabad	0	Residential, Rural and other Areas	6.3	14.7	0.0	0.0	0
4	151	March - M031990	Andhra Pradesh	Hyderabad	0	Industrial Area	4.7	7.5	0.0	0.0	0

In [3]:
```
#Function to calculate so2 individual pollutant index(si)
def calculate_si(so2):
    si=0
    if (so2<=40):
     si= so2*(50/40)
    if (so2>40 and so2<=80):
     si= 50+(so2-40)*(50/40)
    if (so2>80 and so2<=380):
     si= 100+(so2-80)*(100/300)
    if (so2>380 and so2<=800):
     si= 200+(so2-380)*(100/800)
    if (so2>800 and so2<=1600):
     si= 300+(so2-800)*(100/800)
    if (so2>1600):
     si= 400+(so2-1600)*(100/800)
    return si
```

4 4.7 5.875

In [4]:
```
#Function to calculate no2 individual pollutant index(ni)
def calculate_ni(no2):
    ni=0
    if(no2<=40):
     ni= no2*50/40
    elif(no2>40 and no2<=80):
     ni= 50+(no2-14)*(50/40)
    elif(no2>80 and no2<=180):
     ni= 100+(no2-80)*(100/100)
    elif(no2>180 and no2<=280):
     ni= 200+(no2-180)*(100/100)
    elif(no2>280 and no2<=400):
     ni= 300+(no2-280)*(100/120)
    else:
     ni= 400+(no2-400)*(100/120)
    return ni
data['ni']=data['no2'].apply(calculate_ni)
df= data[['no2','ni']]
df.head()
```

In [7]:

```
#function to calculate the air quality index (AQI) of every data value
#its is calculated as per indian govt standards
def calculate_aqi(si,ni,spi,rpi):
    aqi=0
    if(si>ni and si>spi and si>rpi):
     aqi=si
    if(spi>si and spi>ni and spi>rpi):
     aqi=spi
    if(ni>si and ni>spi and ni>rpi):
     aqi=ni
    if(rpi>si and rpi>ni and rpi>spi):
     aqi=rpi
    return aqi
data['AQI']=data.apply(lambda x:calculate_aqi(x['si'],x['ni'],x['spi'],x['rpi']),axis=1)
df= data[['sampling_date','state','si','ni','rpi','spi','AQI']]
df.head()
```

Out [7]:

	sampling_date	state	si	ni	rpi	spi	AQI
0	February - M021990	Andhra Pradesh	6.000	21.750	0.0	0.0	21.750
1	February - M021990	Andhra Pradesh	3.875	8.750	0.0	0.0	8.750
2	February - M021990	Andhra Pradesh	7.750	35.625	0.0	0.0	35.625
3	March - M031990	Andhra Pradesh	7.875	18.375	0.0	0.0	18.375
4	March - M031990	Andhra Pradesh	5.875	9.375	0.0	0.0	9.375

Out [6]:

	spm	spi
435737	0.0	0.0
435738	0.0	0.0
435739	0.0	0.0
435740	0.0	0.0
435741	0.0	0.0

In [7]:

```
#function to calculate the air quality index (AQI) of every data value
#its is calculated as per indian govt standards
def calculate_aqi(si,ni,spi,rpi):
    aqi=0
    if(si>ni and si>spi and si>rpi):
     aqi=si
    if(spi>si and spi>ni and spi>rpi):
     aqi=spi
    if(ni>si and ni>spi and ni>rpi):
     aqi=ni
    if(rpi>si and rpi>ni and rpi>spi):
     aqi=rpi
    return aqi
data['AQI']=data.apply(lambda x:calculate_aqi(x['si'],x['ni'],x['spi'],x['rpi']),axis=1)
df= data[['sampling_date','state','si','ni','rpi','spi','AQI']]
df.head()
```

Out [7]:

	sampling_date	state	si	ni	rpi	spi	AQI
0	February - M021990	Andhra Pradesh	6.000	21.750	0.0	0.0	21.750
1	February - M021990	Andhra Pradesh	3.875	8.750	0.0	0.0	8.750
2	February - M021990	Andhra Pradesh	7.750	35.625	0.0	0.0	35.625
3	March - M031990	Andhra Pradesh	7.875	18.375	0.0	0.0	18.375
4	March - M031990	Andhra Pradesh	5.875	9.375	0.0	0.0	9.375

Out [8]:
```
array(['Andhra Pradesh', 'Arunachal Pradesh', 'Assam', 'Bihar',
       'Chandigarh', 'Chhattisgarh', 'Dadra & Nagar Haveli',
       'Daman & Diu', 'Delhi', 'Goa', 'Gujarat', 'Haryana',
       'Himachal Pradesh', 'Jammu & Kashmir', 'Jharkhand', 'Karnataka',
       'Kerala', 'Madhya Pradesh', 'Maharashtra', 'Manipur', 'Meghalaya',
       'Mizoram', 'Nagaland', 'Odisha', 'Puducherry', 'Punjab',
       'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Uttar Pradesh',
       'Uttarakhand', 'Uttaranchal', 'West Bengal',
       'andaman-and-nicobar-islands', 'Lakshadweep', 'Tripura'],
      dtype=object)
```

In [9]:
```
state=pd.read_csv("../input/indian-states-lat-lon/lat.csv")
state.head()
df.head()
```

Out [9]:

	sampling_date	state	si	ni	rpi	spi	AQI
0	February - M021990	Andhra Pradesh	6.000	21.750	0.0	0.0	21.750
1	February - M021990	Andhra Pradesh	3.875	8.750	0.0	0.0	8.750
2	February - M021990	Andhra Pradesh	7.750	35.625	0.0	0.0	35.625
3	March - M031990	Andhra Pradesh	7.875	18.375	0.0	0.0	18.375
4	March - M031990	Andhra Pradesh	5.875	9.375	0.0	0.0	9.375

In [10]:
```
dff=pd.merge(state.set_index("state"),df.set_index("state"), right_index=True, left_index=Tru
dff.head()
```

Out [10]:

	state	lat	lon	sampling_date	si	ni	rpi	spi	AQI
0	Andhra Pradesh	14.750429	78.570026	February - M021990	6.000	21.750	0.0	0.0	21.750
1	Andhra Pradesh	14.750429	78.570026	February - M021990	3.875	8.750	0.0	0.0	8.750
2	Andhra Pradesh	14.750429	78.570026	February - M021990	7.750	35.625	0.0	0.0	35.625
3	Andhra Pradesh	14.750429	78.570026	March - M031990	7.875	18.375	0.0	0.0	18.375
4	Andhra Pradesh	14.750429	78.570026	March - M031990	5.875	9.375	0.0	0.0	9.375

In [11]:
```
from mpl_toolkits.basemap import Basemap
```

Out [10]:

	state	lat	lon	sampling_date	si	ni	rpi	spi	AQI
0	Andhra Pradesh	14.750429	78.570026	February - M021990	6.000	21.750	0.0	0.0	21.750
1	Andhra Pradesh	14.750429	78.570026	February - M021990	3.875	8.750	0.0	0.0	8.750
2	Andhra Pradesh	14.750429	78.570026	February - M021990	7.750	35.625	0.0	0.0	35.625
3	Andhra Pradesh	14.750429	78.570026	March - M031990	7.875	18.375	0.0	0.0	18.375
4	Andhra Pradesh	14.750429	78.570026	March - M031990	5.875	9.375	0.0	0.0	9.375

In [11]:

```
from mpl_toolkits.basemap import Basemap
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
%config InlineBackend.figure_format = 'retina'
```

In [12]:

```

m = Basemap(projection='mill',llcrnrlat=5,urcrnrlat=40, llcrnrlon=60,urcrnrlon=110,lat_ts=20,
```

In [13]:

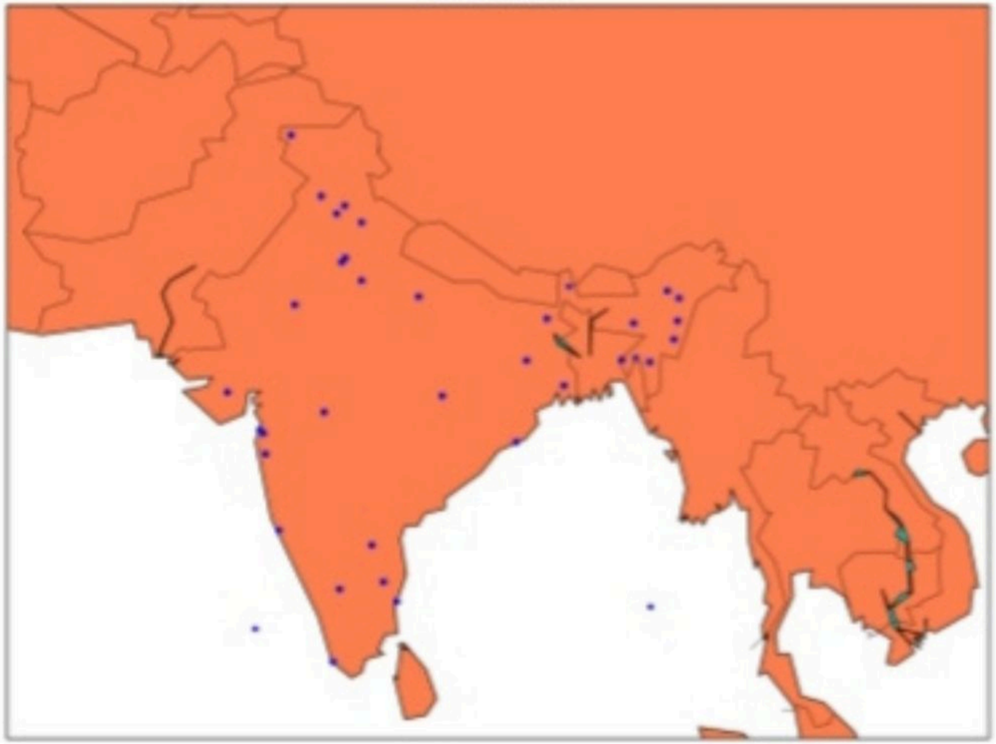
```

longitudes = dff["lon"].tolist()
latitudes = dff["lat"].tolist()
#m = Basemap(width=12000000,height=9000000,projection='lcc',
            #resolution=None,lat_1=80.,lat_2=55,lat_0=80,lon_0=-107.)
x,y = m(longitudes,latitudes)
```

In [14]:

```
fig = plt.figure(figsize=(12,10))
plt.title("All affected areas")
m.plot(x, y, "o", markersize = 3, color = 'blue')
m.drawcoastlines()
m.fillcontinents(color='coral',lake_color='aqua')
m.drawmapboundary()
m.drawcountries()
plt.show()
```

All affected areas



In [15]:

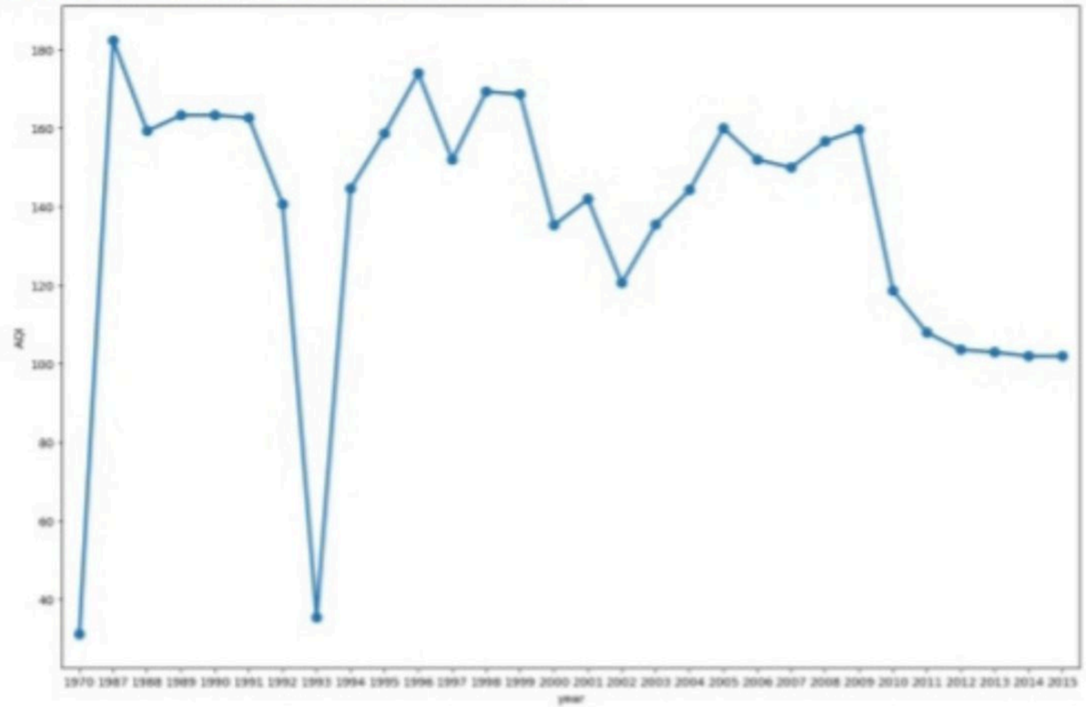
```
#Visualization of AQI across india

data['date'] = pd.to_datetime(data['date'],format='%Y-%m-%d') # date parse
data['year'] = data['date'].dt.year # year
data['year'] = data['year'].fillna(0.0).astype(int)
data = data[(data['year']>0)]

df = data[['AQI','year','state']].groupby(["year"]).median().reset_index().sort_values(by='ye
f,ax=plt.subplots(figsize=(15,10))
sns.pointplot(x='year', y='AQI', data=df)
```

Out [15]: <matplotlib.axes._subplots.AxesSubplot at 0x7f72c28f4cc0>

Out [15]: <matplotlib.axes._subplots.AxesSubplot at 0x7f72c28f4cc0>



In [16]:
```
#setting up date parameter
import warnings
import itertools
import dateutil
```

Output:

```
import statsmodels.api as sm
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import seaborn as sns
%matplotlib inline
df=data[['AQI','date']]
df["date"] = pd.to_datetime(df['date'])
df.tail(20)
```

Out [16]:

	AQI	date
435722	118.333333	2015-11-05
435723	118.666667	2015-11-07
435724	140.666667	2015-11-10
435725	133.666667	2015-11-11
435726	105.000000	2015-11-16
435727	112.666667	2015-11-20
435728	121.333333	2015-11-26
435729	120.000000	2015-11-29
435730	120.666667	2015-12-03
435731	125.000000	2015-12-06
435732	121.666667	2015-12-09
435733	127.000000	2015-12-12
435734	122.666667	2015-12-15
435735	117.000000	2015-12-18
435736	120.000000	2015-12-21
435737	121.000000	2015-12-24
435738	130.333333	2015-12-29
435739	0.000000	1970-01-01
435740	0.000000	1970-01-01

In [17]:

```
#Calculating the yearly mean for the data
df=df.set_index('date').resample('M')["AQI"].mean()
df.head()
```

Out [17]:

```
date
1970-01-31    49.654762
1970-02-28          NaN
1970-03-31          NaN
1970-04-30          NaN
1970-05-31          NaN
Freq: M, Name: AQI, dtype: float64
```

In [18]:

```
#preprocessing the data values
data=df.reset_index(level=0, inplace=False)
data = data[np.isfinite(data['AQI'])]
data=data[data.date != '1970-01-31']
data = data.reset_index(drop=True)
data.head()
```

Out [18]:

	date	AQI
0	1987-01-31	242.438652
1	1987-02-28	235.787929
2	1987-03-31	294.558772
3	1987-04-30	202.012681
4	1987-05-31	307.991667

In [19]:

```
#visualizing the processed data of AQI

df=data.set_index('date')
```

Summary of SPM levels during the year 2006 in Vijayawada

Location	Type of area	Average in µg/m3	Std dev	n	Air quality	% vio
Auto nagar	I	242	30	89	M	0
Benz circle	R	178	20	101	H	17

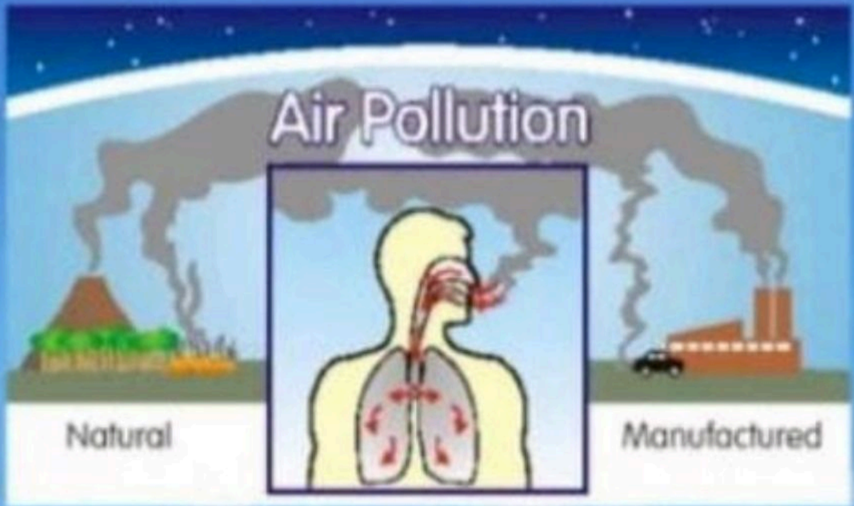
Health aspects

- The health effects of air pollution are both immediate & delayed.
- The immediate are borne by the respiratory system, the resulting state is acute bronchitis.
- If air pollution is intense, it may result even in death by suffocation
- The delayed effects most commonly linked with air pollution are chronic bronchitis, lung cancer, bronchial asthma, emphysema and respiratory allergies.

Effects of air pollution

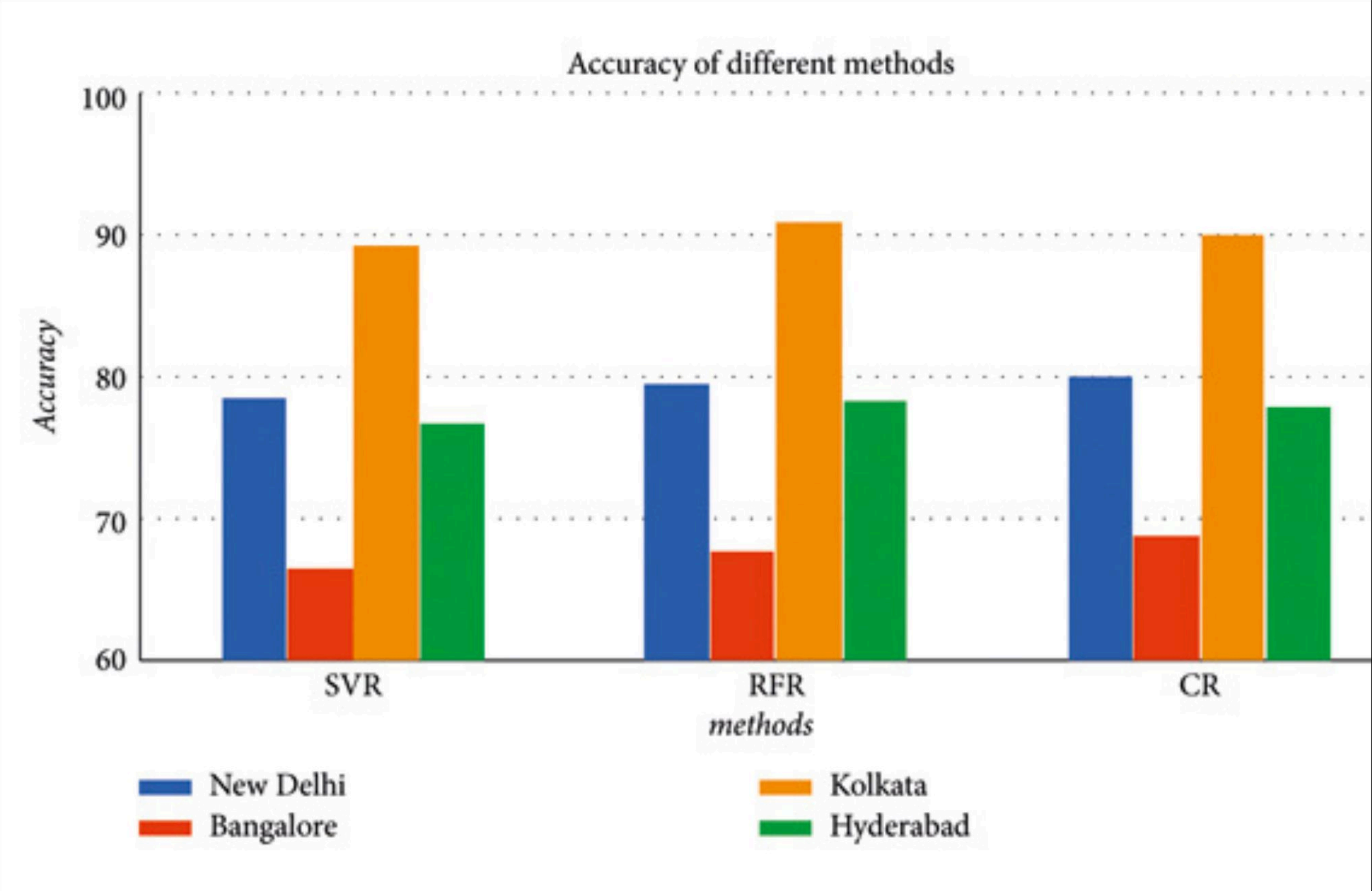
Health aspects

Social & economic aspects

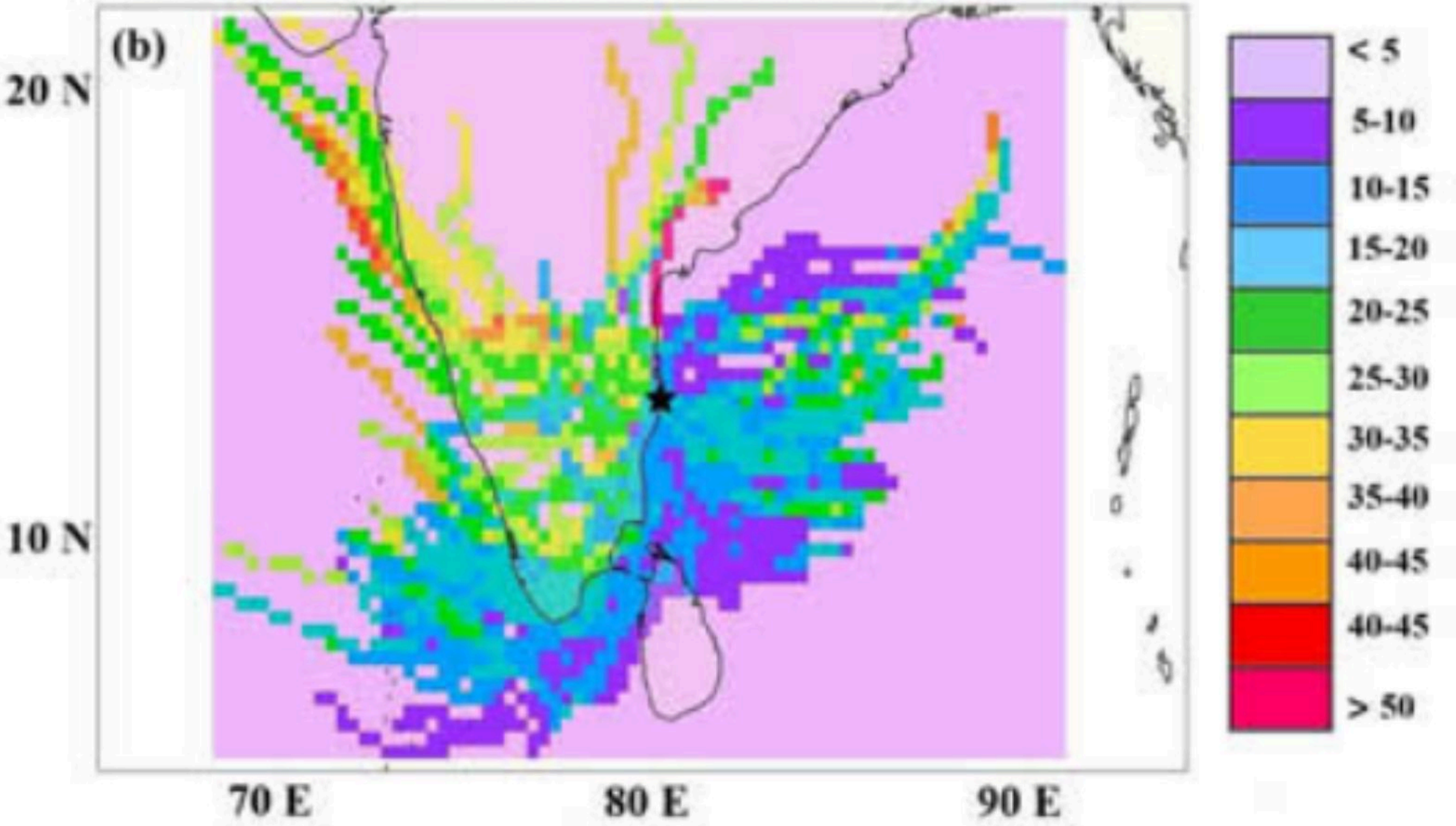


Benefits:

- Lessen the possibility of disease from stroke
- Heart disease
- Lung cancer as well as chronic
- Acute respiratory illnesses such as asthma


Accuracy of different methods
100
90
80
70
60
Accuracy
SVR
RFR
CR
methods
New Delhi
Bangalore
Kolkata
Hyderabad

CWT analysis of PM2.5 over Chennai using two days back trajectories by HYSPLIT during COVID-19 lockdown period in India

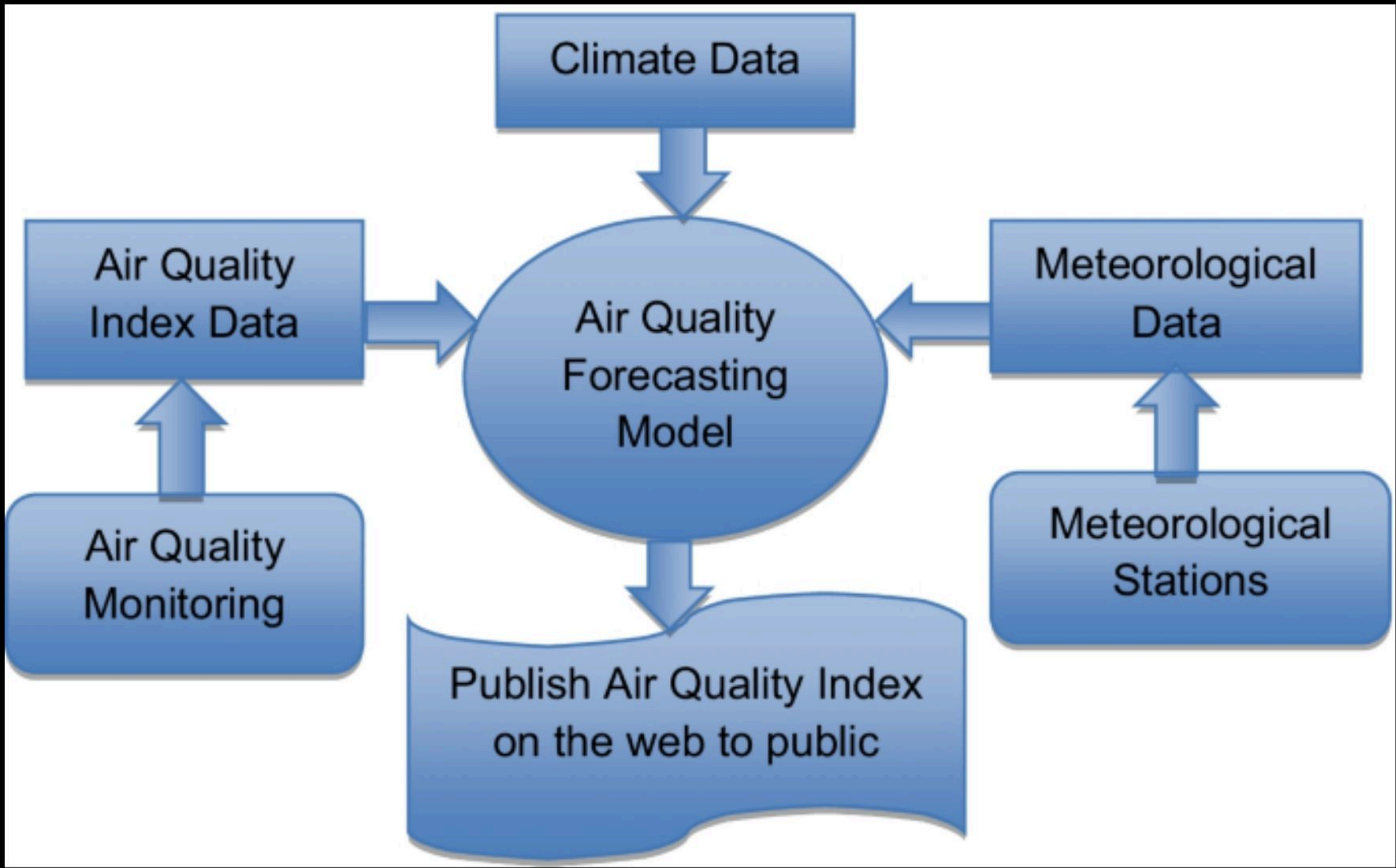


Monitoring of air pollution:

- Monitoring is done to keep a track on quality of air with a view to collect information & improve it
- The best indicators are - SO_2,smoke & suspended particles
- These are monitored on a daily basic & the results are collected by a central agency

Air pollution monitoring in India

- The national air quality monitoring programme sponsored by the central pollution control board since 1990 has generated database over last 14 years in 10 major Indian cities.
- The trend analysis showed that suspended particulate matter (SPM) exceeds the cpcb standards in all the cities of the time throughout the year.
- The concentration ratio of p<10 fraction (human respirable particles) to the total spm varies between 30% to 60%, with coastal cities showing highest percentages.


Climate Data
Air Quality Index Data
Air Quality Forecasting Model
Meteorological Data
Air Quality Monitoring
Meteorological Stations
Publish Air Quality Index on the web to public

Conclusion:

While the effects of air pollution on materials, vegetation & animals can be measured, health effects on human s can only be estimated from epidemiological evidence

All of the evidence we have suggests that air pollutants threaten human health and well-being to an extent that control of these pollutants is necessary