11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Cambridge, United Kingdom

AIR QUALITY INDEX ESTIMATIONS FROM MEASUREMENTS AND STATISTICAL MODELLING TECHNIQUES OVER DELHI

Manju Mohan[1], Anurag Kandya[1] and Manish Yadav[2]

[1]Centre for Atmospheric Sciences,
Indian Institute of Technology Delhi, India

Email:mmohan6@hotmail.com

[2] 531, Dr. Mukherjee Nagar, New Delhi, India

AIR QUALITY MODELLING & FORECASTING

- In air pollution problems, the air quality models are used to predict concentrations of one or more species in space and time as related to the dependent variables.
- Air Quality Models provide the ability to asses the current and future air quality in order to enable well versed policy decisions.
- Thus, air quality models play an important role in providing information for better and more efficient air quality management planning.

Air Quality Models

Physical Models

Simulation of nature on smaller scales in laboratory

Statistical Models

Calculate ambient air concentration from statistical analysis of past air monitoring data even in the absence of emission data. They are good for short term forecasting

Deterministic Models

Calculate ambient air concentration using a solution of various equations representing the relevant physical processes

BACKGROUND

- Deterministic models are subject to model evaluation on case to case basis for often point sources, and relatively less frequently for area and line sources.

- Ultimate objective of the urban modelling is to consider all types of emission sources with the combination of the above models.

- Considering the uncertainties involved in the input data and parameterisations of the underlying physics, the acceptable performance measures are far from the ideal or expected values.

- Based on the comparable performance of statistical models to the deterministic models, it shall be appropriate to consider including statistical models in the model inter-comparison exercise.

OBJECTIVES OF THE STUDY

- Air pollution forecasting from different statistical techniques for the four criteria pollutants namely NO_2, SO_2, Suspended Particulate Mater (SPM) and PM_{10} over Delhi

- Air Quality Index assessment of the above four pollutants over Delhi based on the predicted values and observations

- To study trends of AQI in this megacity considering lot of regulatory/air pollution control measures during past decade

METHODS OF PREDICTION

- Single Exponential Smoothing Method (SES).
- Adaptive Response Rate SES Method (ARRSES).
- Holt's Linear Method (HLM)
- AutoRegressive eXogenous Method (ARX)
- Auto Regressive Integrated Moving Average Method (ARIMA)

Techniques are based on exponential smoothing and regression analysis

SINGLE EXPONENTIAL SMOOTHING (SES)

The equation for the forecast is given by

$$F_{t+1} = F_t + \alpha (Y_t - F_t)$$

Where,

F_{t+1} = forecast for the time t+1

F_t = forecast for the time t

Y_t = observation at time t

α = constant between 0 and 1

The above equation involves a basic principle of negative feedback. The past forecast error is used to correct the next forecast in a direction opposite to that of the error. There will be an adjustment until the error is corrected.

ADAPTIVE RESPONSE RATE SES (ARRSES)

- The basic equation for the forecasting with the method of ARRSES is

$$F_{t+1} = \alpha_t Y_t + (1- \alpha_t) F_t$$

where

- $\alpha_{t+1} = | A_t/M_t |$
- $A_t = \beta E_t + (1-\beta) A_{t-1}$
- $M_t = \beta |E_t| + (1-\beta) M_{t-1}$
- $E_t = Y_t - F_t$
- β = constant between 0 and 1

ARRSES method is an SES method where α value is systematically & automatically, changed from period to period to allow for changes in the pattern of the data.

HOLT'S LINEAR METHOD (HLM)

- The forecast for Holt's linear exponential smoothing is found using two smoothing constants, α and β (with values between 0 and 1), and three equations,

$$\mathbf{F_{t+m} = L_t + b_t m}$$
$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}$$

This method allows forecasting of data with trends.

L_t : level of the series at time t
b_t : slope of the time series at time t.
α,β : constants between 0 and 1

ARX (AutoRegressive eXogenous) MODEL

The model structure is based on least squares analysis where AR refers to the autoregressive part A(q) y(t) and X refers to the extra input B(q) u(t) called the exogenous variable

$$A(q)\, y(t) = B(q)\, u(t-n_k) + e(t)$$

Where

$y(t)$	is the prediction term
$u(t-n_k)$	is the observation term
$e(t)$	is the error term
$A(q)$	$=1+ a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$
$B(q)$	$=b_1+b_2 q^{-1} + \ldots + b_{n_b} q^{-n_b+1}$
q^{-1}	is the delay operator

Models suggested using the ARX method

1. SO_2

Model used is ARX (2 2 1)

$$A(q)\,y(t) = B(q)\,u(t) + e(t)$$

$A(q) = 1 - 0.3792\,q^{-1} - 0.2059\,q^{-2}$

$B(q) = 0.3792\,q^{-1}$

2. $\underline{NO_2}$

Model used is ARX (2 2 1)

$$A(q)\,y(t) = B(q)\,u(t) + e(t)$$

$$A(q) = 1 - 0.3399\,q^{-1} - 0.1541\,q^{-2}$$

$$B(q) = 0.3399\,q^{-1} + 0.1541\,q^{-2}$$

3. SPM

Model used is ARX (1 1 1):

$$A(q)\, y(t) = B(q)\, u(t) + e(t)$$

$$A(q) = 1 - 0.4718\, q^{-1}$$

$$B(q) = 0.4718\, q^{-1}$$

4. PM_{10}

Model used is ARX (1 1 1)

$$A(q)\, y(t) = B(q)\, u(t) + e(t)$$

$$A(q) = 1 - 0.4714\, q^{-1}$$

$$B(q) = 0.4714\, q^{-1}$$

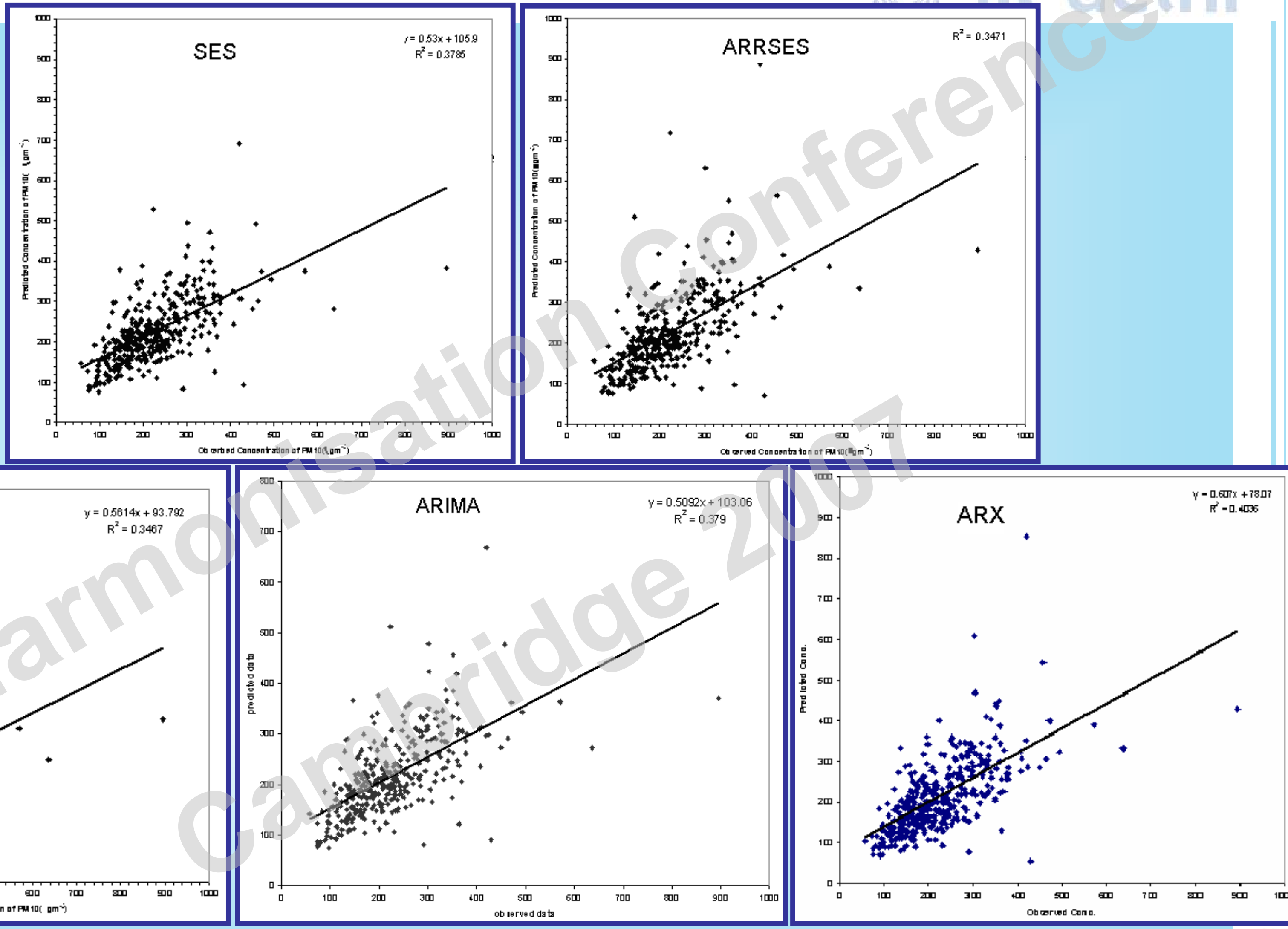
Auto Regressive Integrated Moving Average (ARIMA) Method

- An **autoregressive integrated moving average (ARIMA)** model is a generalisation of an autoregressive moving average or (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series.

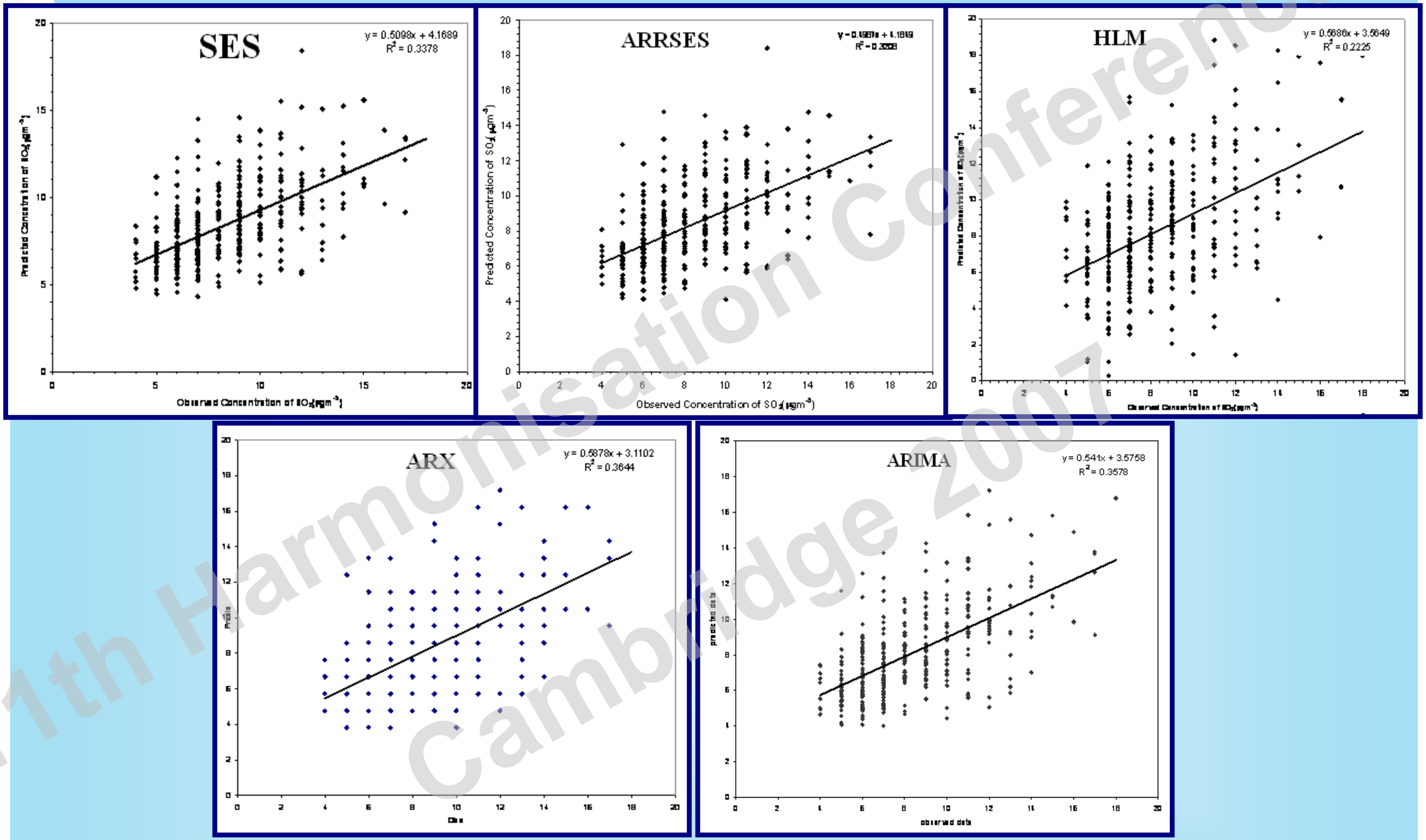
- The model is generally referred to as an ARIMA(p,d,q) model.

ARIMA (p,d,q):

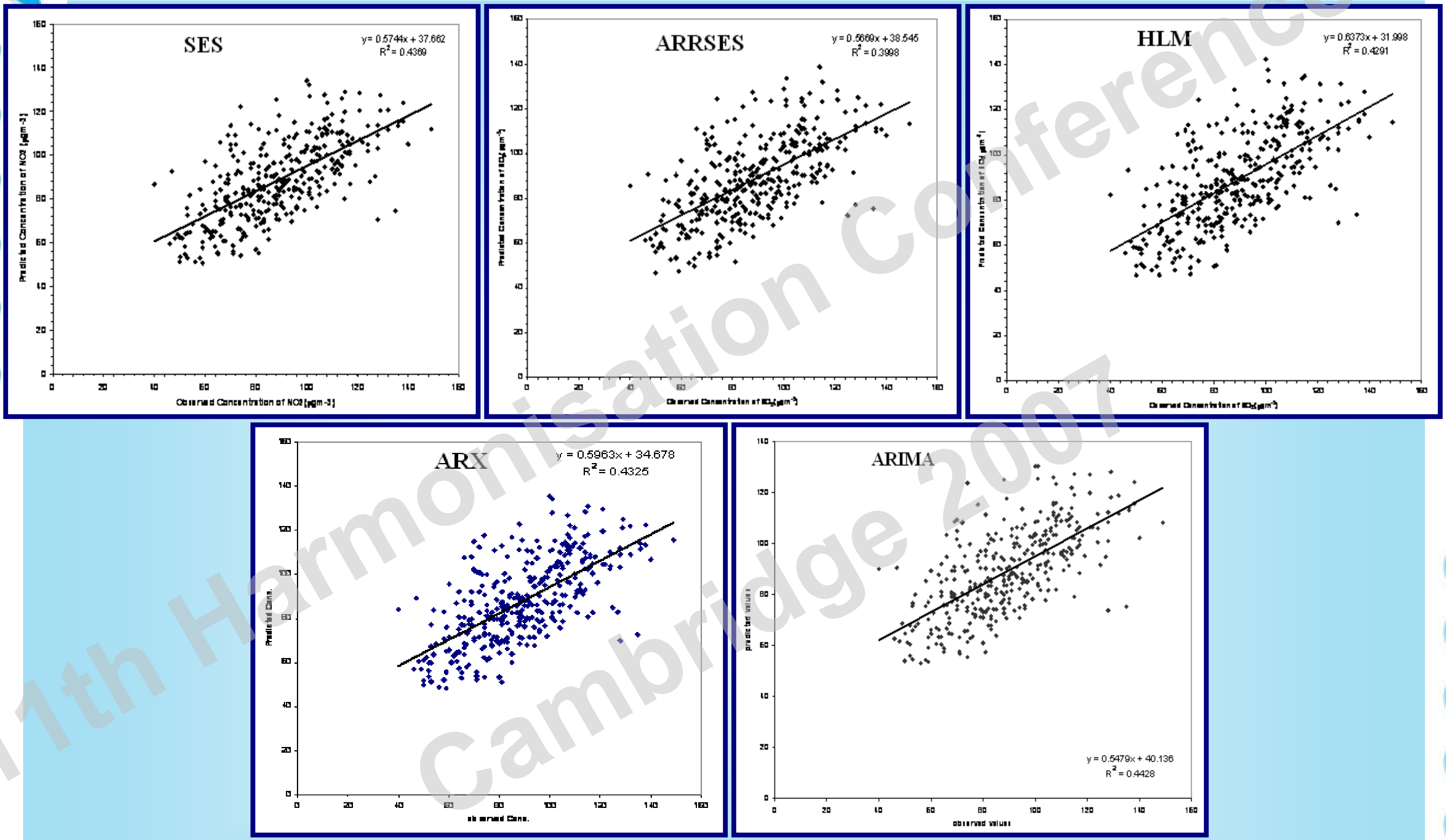
p = order of auto-regressive process
d = degree of differencing involved
q = order of moving averages process.



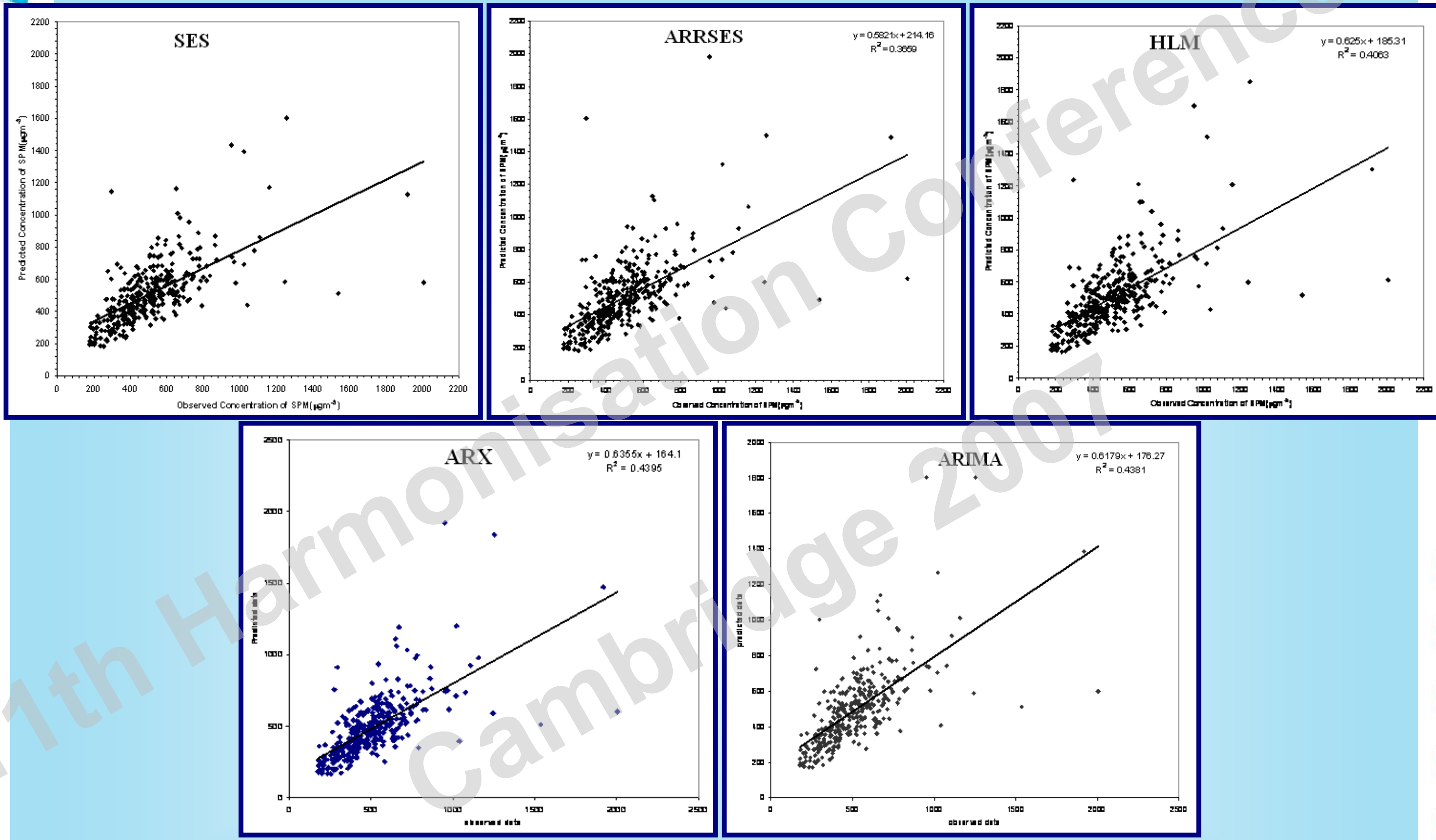
Scatter Plot of observed vs. predicted data of PM_{10} using different methods (one day prediction)



Scatter Plot of observed vs. predicted data of SO_2 using different methods (one day prediction)



Scatter Plot of observed vs. predicted data of NO_2 using different methods (one day prediction)



Scatter Plot of observed vs. predicted data of SPM using different methods (one day prediction)

Mohan *et al*

PERFORMANCE MEASURE

For all the statistical techniques used for predicting the concentration SO_2.

Performance Measures	SES	ARRSES	HLM	ARX	ARIMA	ARX	ARIMA	ARX	ARIMA
	1 DAY					4 DAY		7 DAY	
r	0.581	0.566	0.472	0.604	0.598	0.563	0.564	0.569	0.568
FB	-0.01	0	0.01	0.04	0.01	0.08	0.0	0.08	0.0
MG	0.98	0.99	1.05	1.05	1.0	1.07	0.98	1.07	0.98
VG	1.08	1.08	1.25	1.09	1.06	1.08	1.06	1.08	1.06
NMSE	0.08	0.08	0.13	0.08	0.06	0.09	0.06	0.09	0.06
SD	2.39	2.43	3.16	2.42	2.58	2.36	2.56	2.34	2.54
RMSEU	1.94	1.97	2.91	2.12	1.58	1.76	1.15	1.73	1.16
RMSES	1.38	1.41	1.21	1.21	1.90	1.7	2.16	1.71	2.14
RMSE	2.38	2.42	3.15	2.441	2.471	2.447	2.447	2.433	2.434
d	0.75	0.74	0.68	0.77	0.53	0.71	0.39	0.71	0.04
FAC 2	0.977	0.977	0.909	0.974	1.0	0.977	1.0	0.977	1.0

PERFORMANCE MEASURE

For all the statistical techniques used for predicting the concentration NO_2.

Performance Measures	SES	ARRSES	HLM	ARX	ARIMA	ARX	ARIMA	ARX	ARIMA
	1 DAY					4 DAY		7 DAY	
R^2	0.437	0.399	0.429	0.433	0.443	0.387	0.44	0.405	0.436
r	0.661	0.632	0.655	0.658	0.665	0.622	0.663	0.636	0.66
FB	0	0	0	0.01	0.0	0.02	0.0	0.02	0.0
MG	0.99	0.99	1	1.01	0.99	1.01	0.99	1.01	0.99
VG	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04
NMSE	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.04
SD	16.18	17.01	17.03	16.49	17.73	16.67	18.06	16.29	18.26
RMSEU	13.53	14.41	15.25	14.17	12.83	12.92	14.11	12.41	14.28
RMSES	8.83	8.99	7.53	8.45	12.10	10.65	11.73	10.71	11.22
RMSE	16.15	16..98	17.01	16.5	17.64	16.74	18.35	16.39	18.16
d	0.81	0.79	0.81	0.81	0.7	0.77	0.72	0.78	0.71
FAC 2	0.997	0.997	0.997	0.997	1.0	0.997	1.0	0.997	1.0

PERFORMANCE MEASURE

For all the statistical techniques used for predicting the concentration SPM.

Performanc Measures	SES	ARRSES	HLM	ARX	ARIMA	ARX	ARIMA	ARX	ARIMA
	1 DAY					4 DAY		7 DAY	
R^2	0.405	0.366	0.406	0.439	0.438	0.380	0.387	0.388	0.387
r	0.636	0.605	0.637	0.663	0.661	0.617	0.622	0.623	0.622
FB	0	0	0.01	0.04	0.04	0.08	0.07	0.08	0.07
MG	0.98	0.99	1.01	1.04	1.04	1.07	1.05	1.07	1.05
VG	1.08	1.1	1.09	1.08	1.04	1.09	1.09	1.09	1.09
NMSE	0.11	0.13	0.12	0.12	0.13	0.13	0.13	0.13	0.14
SD	180.1	195.48	189.03	180.39	180.57	179.95	183.76	177.98	181.83
RMSEU	148.8	171.22	169.06	160.57	156.85	136.44	138.68	133.25	135.58
RMSES	100.9	93.77	84.1	84.27	90.87	123.36	124.02	124.63	124.93
RMSE	179.78	195.21	188.82	181.34	181.27	183.94	186.05	182.45	184.37
d	0.78	0.77	0.79	0.8	0.8	0.75	0.75	0.76	0.75
FAC 2	0.99	0.997	0.997	0.997	0.973	0.997	0.97	0.997	0.977

PERFORMANCE MEASURE

For all the statistical techniques used for predicting the concentration RSPM.

Performance Measures	SES	ARRSES	HLM	ARX	ARIMA	ARX	ARIMA	ARX	ARIMA
	1 DAY					4 DAY		7 DAY	
R^2	0.379	0.347	0.347	0.404	0.379	0.362	0.365	0.369	0.367
r	0.615	0.589	0.588	0.635	0.616	0.601	0.604	0.607	0.606
FB	0	-0.02	0.02	0.05	0.03	0.09	0.03	0.09	0.03
MG	0.98	0.98	1.02	1.05	1.01	1.07	1.01	1.07	1.01
VG	1.11	1.12	1.13	1.12	1.1	1.11	1.1	1.11	1.1
NMSE	0.11	0.14	0.13	0.12	0.1	0.13	0.11	0.13	0.11
SD	78.76	88.52	84.82	79.97	66.65	77.91	66.48	77.25	66.80
RMSEU	64.37	80.29	73.59	70.45	51.85	57.93	49.86	56.98	49.63
RMSES	45.18	37.2	42.18	38.91	42.11	55.11	45.07	55.48	45.08
RMSE	78.55	88.49	84.82	80.48	66.79	79..96	67.21	79.52	67.05
d	0.77	0.76	0.76	0.79	0.74	0.74	0.72	0.75	0.73
FAC 2	0.963	0.957	0.951	0.969	0.962	0.966	0.962	0.966	0.962

A COMPARISION OF THE CORRELATION COEFFICIENTS FOR ITO SITE

YEAR	METHODS	POLLUTANTS			
		SO_2	NO_2	SPM	RSPM
2004	SES	0.581	0.661	0.636	0.615
1998 - 2002	SES	0.566	0.532	0.588	0.436
2004	ARRSES	0.566	0.632	0.605	0.589
1998 - 2002	ARRSES	0.557	0.481	0.503	0.370

Data set considered using $\alpha = 0.6$ as obtained from the analysis of the year 2003

INFERENCES

- For one day prediction ARX method is the most suitable considering the overall behavior of performance measures for all the pollutants.
- For 4 day and 7 day predictions, ARIMA technique is comparatively better than ARX. However, performance from all techniques were close to one another
- All necessary coefficients are based on 2004 data. Comparison of performance of various methods based on one year data (i.e., 2003) with that of five years data shows one year predictions are more accurate.
- Time span applicable to statistical models is crucial for their performance and shall be standarised site-wise.

INTRODUCTION TO AQI

- Air Quality Index (AQI), is a system for transforming air pollution levels into a single number, and aims at providing information about air quality in simple terms to general public.
- The AQI focuses on health effects we may experience within a few hours or days after breathing polluted air.
- The AQI varies from 0 to 500. The higher the AQI value, the greater the level of air pollution and the greater the health concern.

AIR QUALITY INDEX ESTIMATIONS

- In this study, estimation and prediction of Air Quality Index from different methods has been carried out.
- Air Quality Indices have been estimated using daily pollution data (1996-2004) of Delhi and also through different statistical techniques.
- The measured and statistically calculated Air Quality in terms of Vulnerabilty Scores have been compared.

1. AQI USING MAX. OPERATOR FUNCTION

$$A_i = \frac{C_i}{S_i} * 100$$

Where, A_i = Sub index of pollutant i

C_i = Concentration of pollutant i

S_i = Air quality standard for pollutant i

- Maximum Value of A_i is reported as the AQI
- This method has been used by USEPA and also by CPCB in a modified for AQI estimation.

AQI Values	Levels of Health Concern	Colors
When the AQI is in this range:	...air quality conditions are:	...as symbolized by this color:
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Description of Air Quality Index Values

VULNERABILITY INDEX

Vulnerable score (VST) is obtained from the following expression:

$$VS_T = \sum_{i=1}^{n} X_i T_i$$

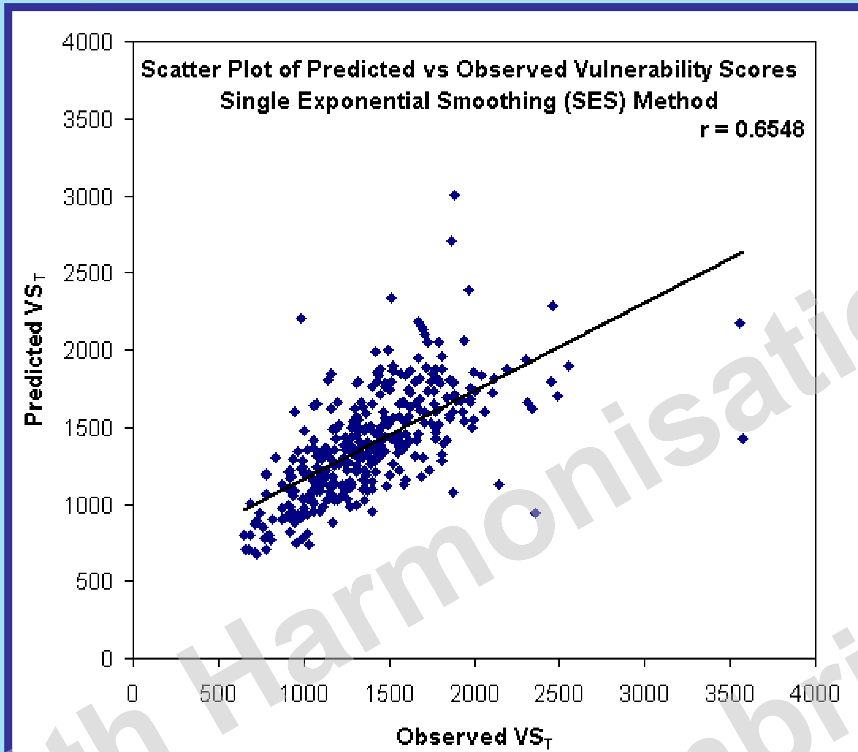
where
Xi = concentration of *i*th air pollutant,
Ti = toxicity weighing factors for *i*th air pollutant,
n = the number of air pollutants.

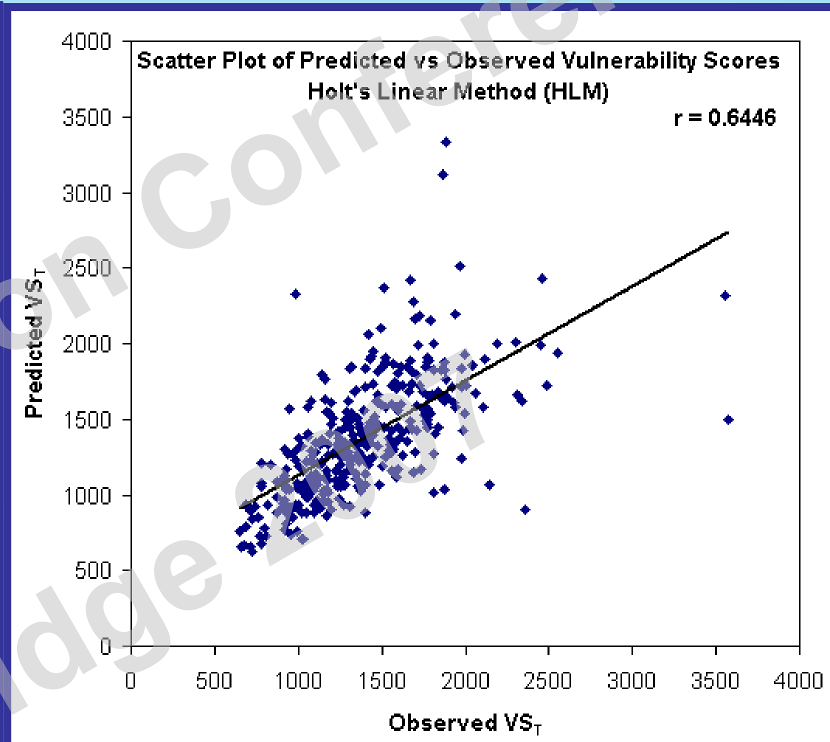
Pollutant	Relative Weight
NO_x	4.5
PM_{10}	2.3
SO_2	1.4

the toxicity weighing factors as given by World Bank

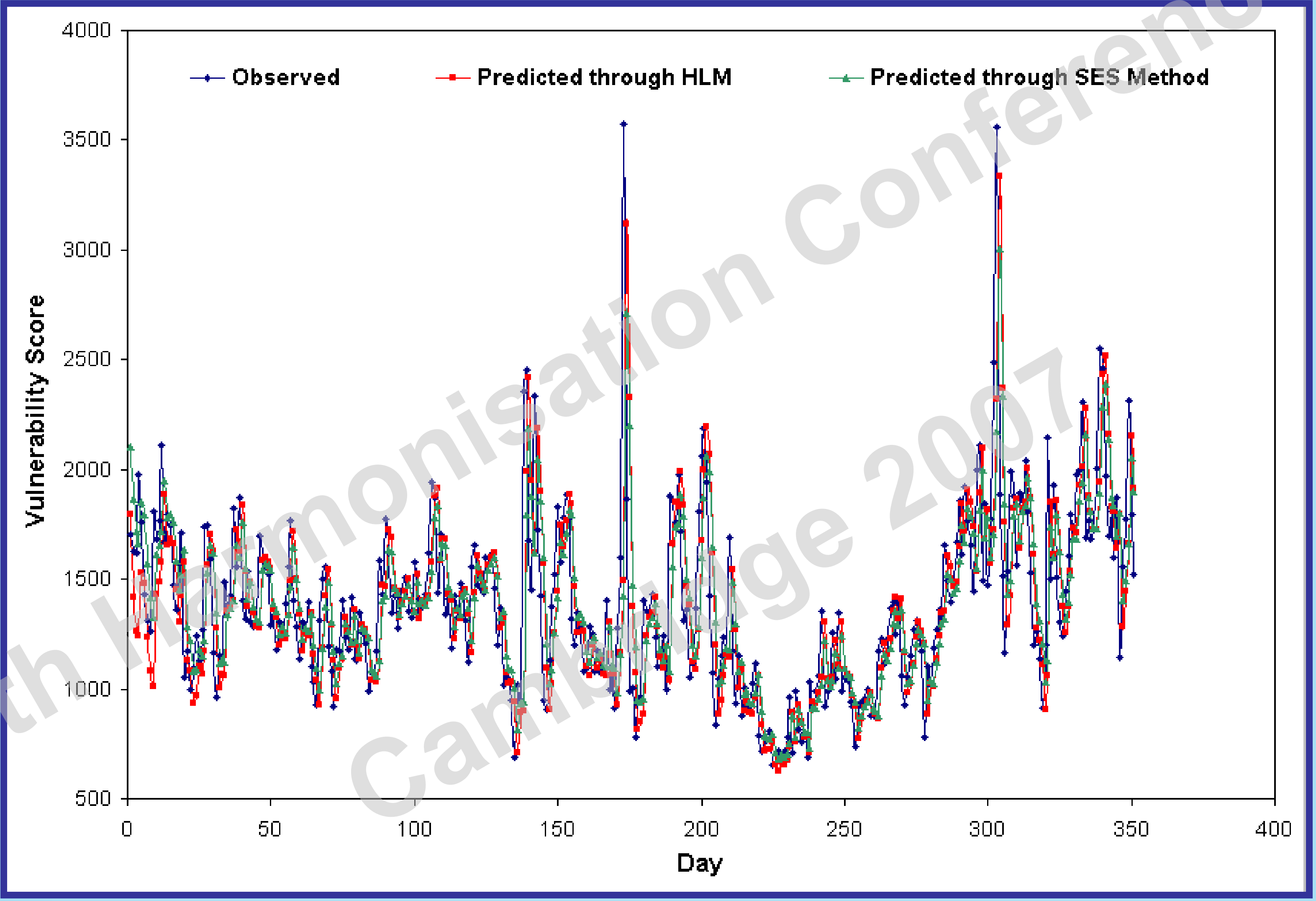
Total Vulnerability Score VS_T	Vulnerability Index VI
> 4420	High
4420-3315	Medium High
3315-2210	High
2210-1661	Medium High
1661-1113	Medium
1113-517	Low
< 517	Very Low

Description of Vulnerability Index Values





Scatter Plot of Observed vs. Predicted Data of Vulnerability Scores using Two Different Methods

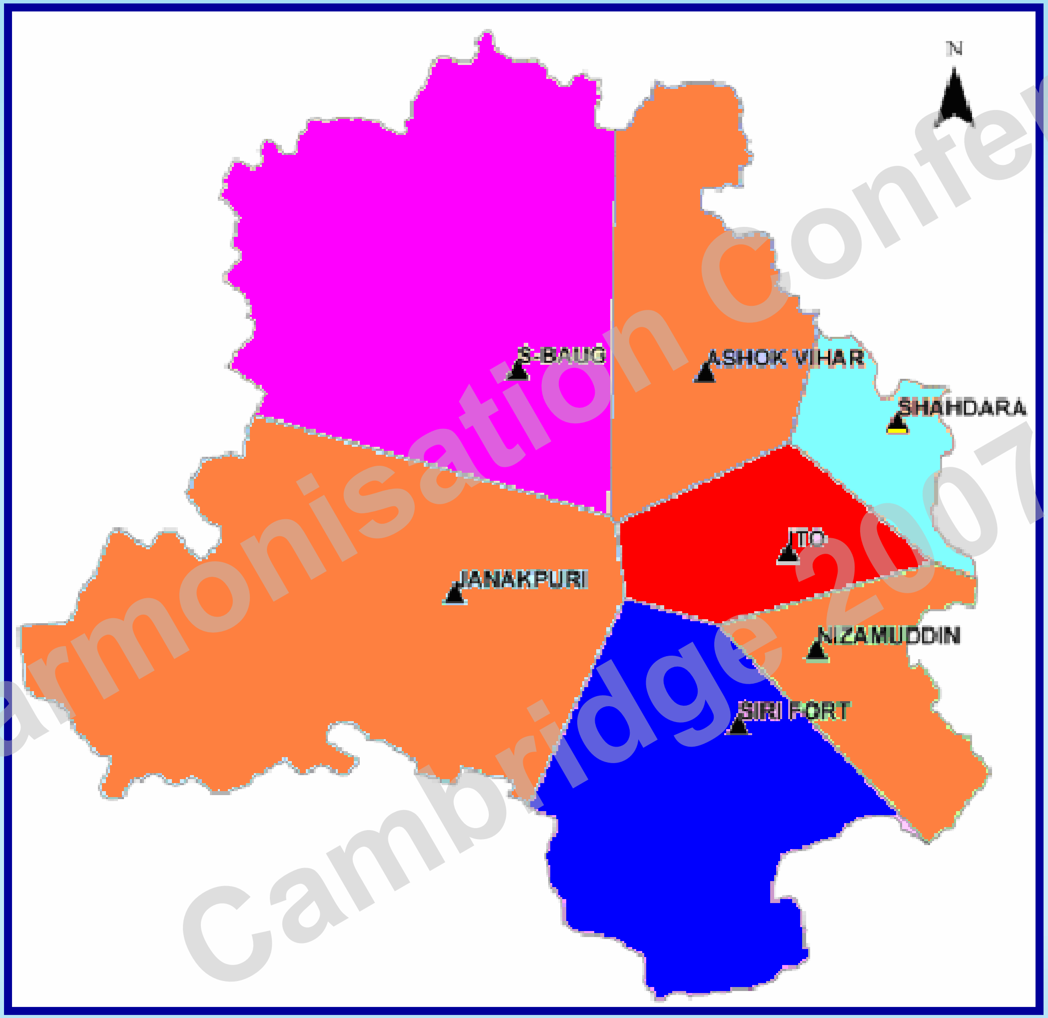


Time Series of Observed and Predicted Vulnerability Scores

Mohan *et al*

AIR QUALITY MONITORING IN DELHI

- The locations have been categorized on land use basis i.e. Residential, Industrial and Traffic Intersection.
- The stations coming under residential area are Ashok Vihar, Siri Fort, Janakpuri and Nizamuddin, under Industrial area are Shahzada Bagh and Shahdara and ITO is a traffic site junction.
- Four pollutants viz., SO_2, NO_2, SPM and RSPM have been identified for regular monitoring at these locations



Location of Ambient Air Quality Monitoring Stations in Delhi

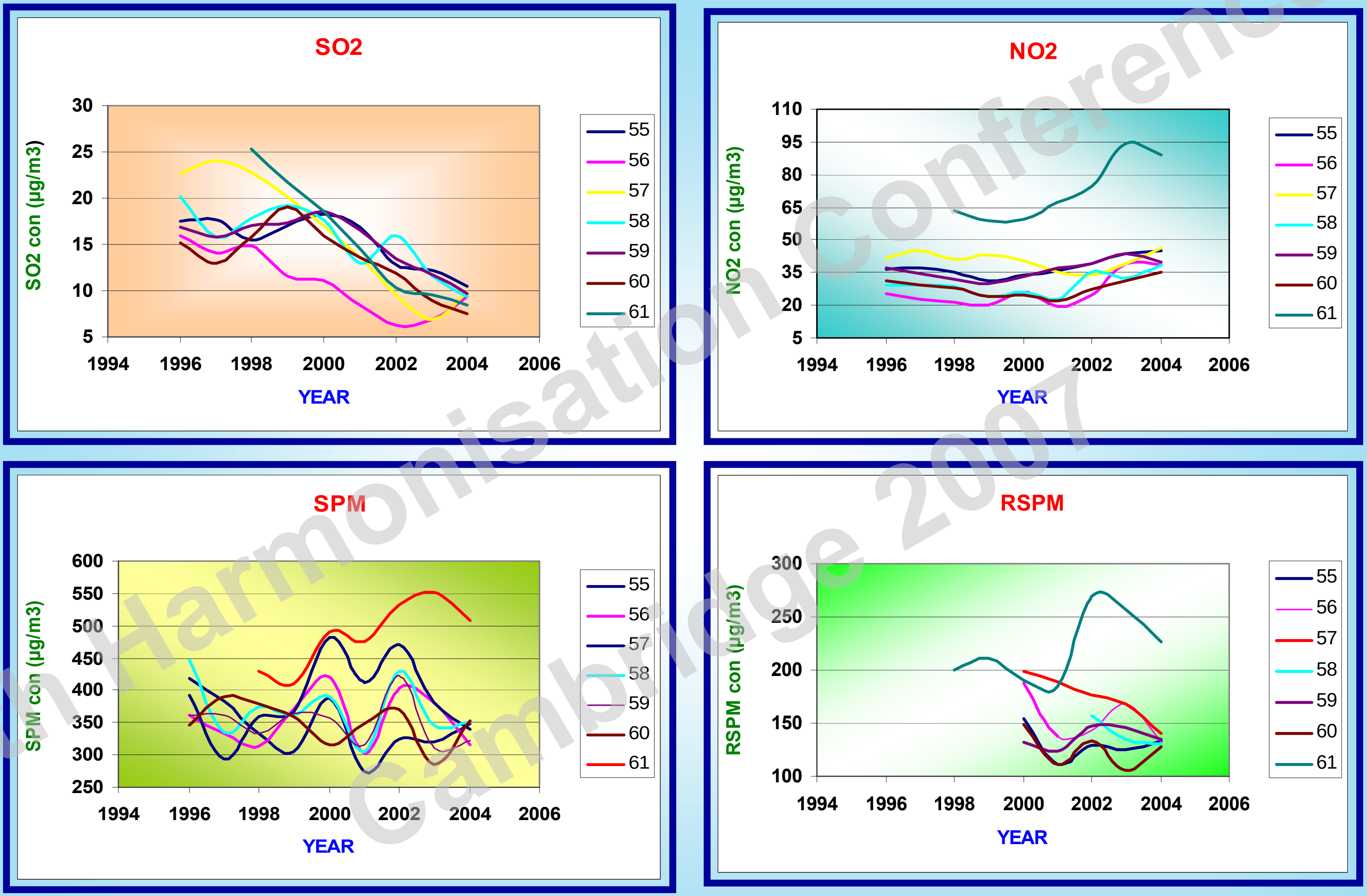
SEASONAL AQI ANALYSIS

- For the seasonal analysis, following was the break up taken into consideration

Sr. No	SEASON	MONTHS
1.	Winter	December, January and February
2.	Summer	March, April, May and June
3.	Monsoon	July, August and September
4.	Post Monsoon	October and November

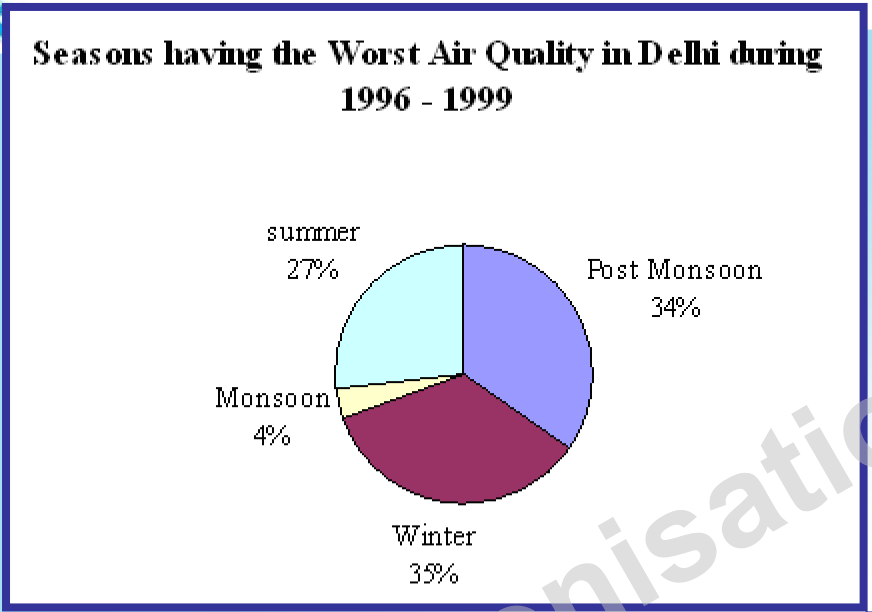
RESULTS & DISCUSSION

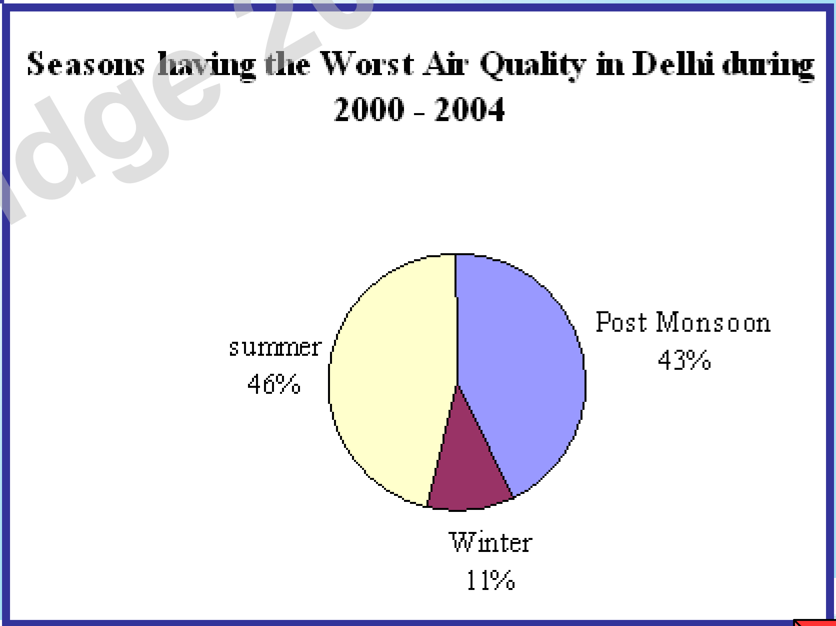
- In 1996, it was generally the winters that used to have the worst type of air quality, but in 2004, summers followed by post-monsoon show the worst type of air quality.
- Change of season for worst AQI from Winter to Summer may be due to increased photochemical reactions and change in the nature of emissions owing to different control measures
- A consistent gradual increase in NO_2 levels from 2001 onwards reflects the effect of changed emission patterns in the city and more photochemical reactions.
- Overall, majority of sampling stations showed improvement in AQI from 1996 to 2004. This is the period when regulatory measures were implemented in a major way.



Yearly Variation of SO_2 NO_2, SPM and RSPM across Delhi

Mohan *et al*


Seasons having the Worst Air Quality in Delhi during 1996 - 1999
summer 27%
Post Monsoon 34%
Monsoon 4%
Winter 35%


Seasons having the Worst Air Quality in Delhi during 2000 - 2004
summer 46%
Post Monsoon 43%
Winter 11%

Mohan *et al*

CONCLUSIONS

- A shift in worst AQI season from winter to summer is noted
- An increased NO_2 concentration at all sites increased from 2000 onwards.
- ARX method is most suitable for predicting One Day concentrations while ARIMA technique scores well for 4 and 7 days prediction. However, all techniques showed reasonable and close performances
- Vulnerability Scores estimated with predicted pollutant concentrations compare well with those estimated with observed values.
- Considering the performance evaluation of statistical model techniques it is recommended that the the best performing statistical model be included along-side the deterministic models for the model inter-comparison exercise

Thanks