

**Particulate air pollution data for Coimbatore, India: real
time monitoring and modeling with data-interoperability
measures**

Thesis submitted to the
BHARATHIAR UNIVERSITY, COIMBATORE
for the award of
DEGREE OF DOCTOR OF PHILOSOPHY
in
ENVIRONMENTAL SCIENCES

by
Nishadh K A
(Reg. No. 2010R1562)



Sálim Ali Centre for Ornithology and Natural History (SACON)
(Aided by the Ministry of Environment, Forests and climate change, Govt. of India)
Coimbatore 641 108, India

September 2016

I dedicate this thesis to my loving parents, my wife Powsiya and my teachers

Certificate

This is certify that the thesis, entitled **“Particulate air pollution data for Coimbatore, India: real time monitoring and modeling with data-interoperability measures”** submitted to the Bharathiar University, in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in **Environmental Sciences** is a record of original research work done by **Mr. Nishadh K A** during the period **April 2010 - September 2016** of his research in the **Department of Environmental Sciences at Sálím Ali Centre for Ornithology and Natural History (SACON), Anaikatty, Coimbatore - 641 108** under my supervision and guidance and the thesis has not formed the basis for the award of any Degree /Diploma /Associateship /Fellowship or other similar title of any candidate of any University.

Signature of the Guide

Countersigned

Director

Declaration

I **Mr. Nishadh. K. A** hereby declare that the thesis, entitled “**Particulate air pollution data for Coimbatore, India: real time monitoring and modeling with data-interoperability measures**”, submitted to the Bharathiar University, in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in **Environmental Sciences** is a record of original and independent research work done by me during **April 2010 - September 2016** under the Supervision and Guidance of **Dr P A Azeez, Department of Environmental Sciences at Sálím Ali Centre for Ornithology and Natural History (SACON), Anaikatty, Coimbatore – 641 108**, and it has not formed the basis for the award of any Degree/ Diploma/ Associateship/ Fellowship or other similar title to any candidate in any University.

Signature of the candidate

Acknowledgements

I express my hearty thanks to Dr. P. A. Azeez, my mentor and PhD guide for his confidence in me as a researcher and giving enviable freedom to explore and pursue. I felt his research guidance rather *Improvisational theatre* play! I thank him for handholding me in to *terra incognita* by being my co-actor in the long, perhaps a slow moving play, and for being inspirational for a lifetime. I will remain thankful to him for all support, valuable guidance, amity and more than all his patience.

I thank Dr K Sankar, Director, SACON for his kind support and facilitation for submitting the thesis. I am thankful to Dr P Balasubramanian, Research Coordinator & Nodal Officer, SACON for all the helps rendered thorough out the study and for completing PhD formalities.

The study is part of the project entitled, “Real time air pollution modeling and dissemination of location based information using mobile devices” funded by Natural Resource Data Management System (NRDMS), National Spatial Data Infrastructure (NSDI), Department of Science and technology (DST), New Delhi. I am grateful to Dr PS Acharya (Scientist G, NRDMS & NSDI Division) for all supports. I am also obliged to DST (Government of India) for financial assistance through Rajat Jayanti Vigayan Sancharak Fellowship Award-2011. The helps rendered as a Co-PI to the project and guidance of Dr. R. Mohanraj (Bharathidasan University) is greatly treasured; his insights into air pollution studies were stimulating.

I am thankful to Drs. K. S. Anoop Das, Goldin Quadros, Mathew K Sebastian, PR Arun, Ram Pratap Singh, P P Nikhilraj, P Rajan, Mohd Zeeshan and Sarath Guttikunda for their kind support; that was great help in time, in need. I also acknowledge their valuable promptings and

discussions on the study. I am thankful to Dr. K. S. Anoop Das for encouraging me to venture into my doctoral program.

I am thankful to *Thinking with Data* workshop organized by Dr. Rajesh Kasturirangan and Mr. S. Anand at National Institute of Advanced Studies during mid 2012. The workshop was a great motivation for pursuing this study and most of the techniques used in the study were introduced to me in that series of classes. I am thankful to Simple Labs, Chennai for introductory workshop on *Do It Yourself* hardware development. I am thankful to organizers of workshops on Open Geospatial technologies held at Foundation for Ecological Research, Advocacy and Learning during 2011 and International Institute of Information Technology, Hyderabad during 2014. These workshops introduced me on to open source tools at large and specifically to geospatial analysis.

I am thankful to Mrs Aswathy P V for her help during the study. I am thankful to Dr Sachin S Gunthe and Mr Tabish Umar Ansari for their help in WRF-CHEM evaluation study. I am thankful to Mr C V Jayaprakash (PSG College, Coimbatore), Mr K. Mohan Raj (Coimbatore wetlands, Coimbatore) and Mr Kalidas (Osai, Coimbatore) for their help in various stages of the study especially during field studies. I am grateful to Mr. Chandra Sekaran, Engineer at Tamil Nadu Pollution Control Board for his help during the study. I am thankful to the open source community behind all the software applications used in the study.

I express my gratitude to scientists at SACON Drs S. Muralidharan, (Late) S Bhupathy, P Pramod, Shomita Mukherjee, Karunakaran, Rajah Jaipal, H N Kumara, S Manchi Shirish, Mahindaran Mylswamy and S Babu for enriching my stay at SACON with in the backdrop of scientific temper and curiosity. I am thankful to them for discussion relating to the studies; it helped improving my study. I thank other staff at SACON for their help rendered in various stages of the study. I am thankful to Mr R Jaykumar (Administrative Officer), Mr K Aneesh Abraham (Finance Officer), Mr V Vaidyanathan (PA to Director), Mr. Manoharan, (Librarian I/c), Lt Col (retired) N Sundhar Raj (Site Engineer), Mr. S. Patturajan, Mr. M.Eanamuthu, Mrs. R. Rajalakshmi, Mrs. M. Jayageetha, Mrs. V. Santhalakshmi, Mr A Devaraj, Mr. R. Ravi, Mr. P. Subramanian, and Mr Vijaykant for their helps and making my association with SACON

memorable. I thank Mr A. Srinivasan, Computer assistant at SACON for his helps during the study.

I am thankful to Drs Balakrishnan, Dhanya, Ranjini, Sheeba, Srinivas, Nehru, Ganesan, Jayakumar, Ramesh Kumar, Anbarasu, Zaibin, Vinoth, Chandran, Ramesh, Joseph, Muhil, Madhumita, Jins, Akshaya, Avadhoot, Nambi, Kirubha, Divya, Shanta Kumar, Sony, Chaithra, Suresh, Rajesh, Maharajan, Manikandan, Prakash, Santhosh, Arijit, Partha, Anoop Pankaj-my seniors, friends and fellow researchers in SACON for their kind help and discussions throughout the study.

I am indeed indebted to my parents, brothers, sisters-in-law and wife for love and enduring support bestowed on me and I acknowledge them for bearing with me during my PhD work.

Abstract

Particulate air pollution is a major health burden and environmental concern in urban areas. As a serious health problem in urban areas, current intervention measures has to be sufficiently refined for urgent and sustainable management. Data intensive approach can gives tools to integrate diverse data sources for deriving decision-making information and improved applications for adaptive management of pollution. However lack of spatio-temporally relevant and reliable data on particulate pollution and the data existing in non-interoperable formats to a great extent hampers knowledge generation for effective control of pollution and management of air quality.

The current study focused on developing basic tools for data intensive approach in a second tier urban centre of India. The study intends to explore an affordable real time air quality information systems focusing on Coimbatore, a fast growing and second tier urban center in the state of Tamil Nadu, India and its surroundings as the study area. The major objectives of the study were (1) to develop a real time particulate pollution monitoring system using low cost commodity sensors and assess its effectiveness in the study area, (2) attempt a real time particulate pollution modeling system for the study area using WRF-CHEM, addressing its computational requirements, and (3) demonstrate application of interoperability measures on real time particulate pollution data.

To address the first objective, a real time particulate monitor was developed by integrating off-the-shelf indoor dust sensors with an appropriately customized data communication system. To address the objective two, as an essential data requirement for WRF-CHEM modeling, particulate matter ($PM_{2.5}$ and PM_{10}) emission inventory was prepared for the study area.

Programming tools (codes) were developed for remote computing based real time execution and evaluation of model performance over the study area using the developed emission inventory. Objective three of the study was addressed using sensor web enablement specification and its application. Web based data dissemination and application of statistical analysis tools were used to demonstrate the advantages of interoperability measures on real time particulate pollution data in the study area.

Table of contents

List of figures	xiii
List of tables	xv
Abbreviations	xvi
1 General Introduction	1
1.1 Urban particulate matter air pollution	1
1.2 Issues with current intervention measures	3
1.2.1 Reductionistic view	5
1.2.2 Authoritative approach	6
1.2.3 Incompatibility	7
1.3 Data Intensive approach	7
1.3.1 Complex context	9
1.3.2 Collective action through openness	9
1.3.3 Compatibility through data-interoperability	10
1.4 Study objectives and thesis organization	11
2 Study area: Coimbatore	13
3 Real time particulate pollution monitoring	23
3.1 Introduction	23
3.2 Methodology	26
3.2.1 Monitor development and field deployment	26
3.2.2 Real time data communication and management	27
3.2.3 Data conversion and <i>collocation</i> sampling	27
3.3 Results and Discussion	28
3.3.1 Ambient testing	31

3.3.2	Data validity	35
3.4	Conclusion	38
4	Particulate pollution emission inventory	40
4.1	Introduction	40
4.2	Methodology	42
4.2.1	Study area:	42
4.2.2	Emission source and activity data	43
4.2.3	Emission factors, emission calculation and spatial allocation	47
4.3	Results	49
4.3.1	Programming tools	49
4.3.2	$PM_{2.5}$ emission	53
4.3.3	PM_{10} emission	54
4.4	Discussion	54
4.5	Conclusion	61
5	Real time particulate pollution modeling	62
5.1	Introduction	62
5.2	Methodology	64
5.2.1	WRF-CHEM model and its components	64
5.2.2	Real time automatic execution	66
5.2.3	Evaluation of computing resource requirement	68
5.2.4	Evaluation of model performance for Coimbatore domain	70
5.3	Results	72
5.3.1	Real time automatic execution	72
5.3.2	Evaluation of computing resource requirement	74
5.3.3	Evaluation of model performance for Coimbatore domain	76
5.4	Discussion	83
5.5	Conclusion	87
6	Data-Interoperability measures	89
6.1	Introduction	89
6.2	Methodology	91
6.2.1	Sensor Observation Service: <i>IstSOS</i>	91
6.2.2	SOS for real time monitoring	94
6.2.3	SOS for real time modeling	94

6.2.4	Advantages of Sensor Observation Service	95
6.3	Results	95
6.3.1	SOS for real time monitoring and modeling data	95
6.3.2	Advantages of Sensor Observation Service (SOS)	104
6.4	Discussion	108
6.5	Conclusion	111
7	Conclusion	112
	References	119
	Appendix A	142

List of figures

1.1	Relationship of urban development and air pollution evolution	4
1.2	Technological challenges of sensor web enablement	11
2.1	The population increase in Tamil Nadu during 1901-2011	14
2.2	The population increase in Coimbatore urban during 1901-2011	15
2.3	The number of vehicular registration in Tamil Nadu during 1993-2014	19
2.4	The number of vehicular registration in Coimbatore during 2000-2016	20
2.5	Ambient particulate matter concentration in different stations of Coimbatore during 2004-2012	21
3.1	Real time particulate pollution monitor with casing for field deployment . . .	29
3.2	Software and hardware components of real time particulate pollution monitor	30
3.3	Real time particulate pollution monitor field deployment stations in Coimbatore	32
3.4	Hourly $PM_{2.5}$ concentrations in the stations where the monitors were deployed	33
3.5	Hourly PM_{10} concentrations in the stations where the monitors were deployed	33
3.6	Linear regression for $PM_{2.5}$ readings	36
3.7	Linear regression for PM_{10} readings	37
3.8	Hourly $PM_{2.5}$ concentration in custom real time monitor and Aerocet 531S during collocation sampling	38
3.9	Hourly PM_{10} concentration in custom real time monitor and Aerocet 531S during collocation sampling	38
4.1	The study area map with extent of area considered for emission inventory calculations(50x60km)	43
4.2	Road network in Coimbatore region	45
4.3	Steps involved in residential sector emission inventory estimation	50
4.4	Steps involved in Transport sector emission inventory estimation	51
4.5	Steps involved in Industrial sector emission inventory estimation	53
4.6	Total $PM_{2.5}$ emission from Coimbatore	55

4.7	$PM_{2.5}$ emission from residential sector in Coimbatore	55
4.8	$PM_{2.5}$ emission from transport sector in Coimbatore	55
4.9	$PM_{2.5}$ emission from wind blown dust in Coimbatore	56
4.10	$PM_{2.5}$ emission from Industrial sector in Coimbatore	56
4.11	Total PM_{10} emission from Coimbatore	57
4.12	PM_{10} emission from residential sector in Coimbatore	57
4.13	PM_{10} emission from Transport sector in Coimbatore	57
4.14	PM_{10} emission from wind blown dust in Coimbatore	58
4.15	PM_{10} emission from Industrial sector in Coimbatore	58
5.1	Components of WRF-CHEM model system	67
5.2	Executable programs of WRF-CHEM model system	68
5.3	Simulation domain for computing resource evaluation study and for model performance evaluation	69
5.4	Steps involved in WRF CHEM simulation addressed by automatic execution script	73
5.5	Steps involved in integrating the Coimbatore local EI over global EDGAR HTAP EI	77
5.6	Hourly temporal variation of $PM_{2.5}$ and PM_{10} concentration WRF CHEM simulated using CBE EI and observation by Aerocet 531S monitor	80
5.7	Daily average spatial distribution of WRF CHEM simulated $PM_{2.5}$ by using CBE EI	81
5.8	Daily average spatial distribution of WRF CHEM simulated $PM_{2.5}$ by using EDGAR EI	81
5.9	Daily average spatial distribution of WRF CHEM simulated PM_{10} by using CBE EI	82
5.10	Daily average spatial distribution of WRF CHEM simulated PM_{10} by using EDGAR EI	82
5.11	Wind rose diagram of the automatic weather station wind profile during the simulation period	83
6.1	Components of Sensor Web Enablement(SWE) specification	92
6.2	Key objects of Sensor Observation Services(SOS)	93
6.3	Server status page of <i>IstSOS</i>	96
6.4	Service description page of <i>IstSOS</i>	97
6.5	Various process involved in importing data into <i>IstSOS</i>	100
6.6	Components and its interaction in web map application based on SOS	105
6.7	Interfaces in web application based on SOS	107

List of tables

1.1	Characteristics of various knowledge systems	6
3.1	Cost of the materials used in real time particulate pollution monitor	30
3.2	Price range of some particulate matter monitors	30
3.3	The profile and period of sampling for Dylos™air quality monitor in different stations of Coimbatore	32
4.1	Per capita consumption of fuel types in Tamil Nadu	44
4.2	Vehicular population in Coimbatore district during 2012-2013	45
4.3	Fuel consumption in Coimbatore during 2007-2011	46
4.4	Emission factor for each sector and fuel type	47
5.1	Wall clock time taken for various steps of WRF-CHEM execution in different computer processors	75
5.2	Model parametrization scheme used for performance evaluation study	76
5.3	Performance measures for WRF CHEM simulation based on CBE EI and EDGAR EI	79
A.1	Linear regression Adjusted R^2 for collocation sampling	142

Abbreviations

<i>PM</i> ₁₀	Particulate Matter with aerodynamic diameter of 2.5-10 μm diameters
<i>PM</i> _{2.5}	Particulate Matter with aerodynamic diameter equal and less than 2.5 μm diameters
AMI	Amazon TM Machine Images
API	Application Programming Interface
AQI	Air Quality Index
AWS-EC2	Amazon TM Web Services Elastic Computer Cloud
BAM	Beta Attenuation Monitors
BDL	Below Detectable Limit
CBE EI	Coimbatore Emission Inventory
CEPI	Comprehensive Environmental Pollution Index
CHANS	Coupled Human Natural System
COE	Coefficient Of Efficiency
CPCB	Central Pollution Control Boards
CSV	Comma Separated Value
CTM	Chemical Transport Models
DISD	Data Intensive Scientific Discovery
EDGAR	Emission Database for Global Atmospheric Research
EF	Emission Factor
EI	Emission Inventory

GFS	Global Forecast System
GUI	Graphical User Interface
HPC	High Performance Computing
IOT	Internet Of Things
IPCC	Inter governmental Panel on Climate Change
JSON	Java Script Object Notation
MB	Mean Bias
MFB	Mean Fractional Bias
MGE	Mean Gross Error
MPI	Message Passing Interface
NCAR	National Center for Atmospheric Research, USA
nco	Netcdf Operator
NOAA	National Oceanic and Atmospheric Administration, USA
NSSO	National Sample Survey Organization, India
NWP	Numerical Weather Prediction
O&M	Observation and Measurement
OGC	Open Geospatial Consortium
PAH	Polycyclic Aromatic Hydrocarbons
PHEP	Pykara Hydro Electric Power Project, Tamil Nadu
PM	Particulate Matter
r	Correlation Coefficient
REST-API	Representation State Transfer-Application Programming Interface
RH	Relative Humidity
RMSE	Root Mean Square Error
RSPM	Respirable Suspended Particulate Matter
RT-AQF	Real Time Air Quality Forecast systems

RTO	Regional Transport Office, Tamil Nadu
SAS	Sensor Alert Service
SBC	Single Board Computer
SensorML	Sensor Model Language
SMS	Short Message Service
SOS	Sensor Observation Service
SPCB	State Pollution Control Boards
SPS	Sensor Planning Service
SWE	Sensor Web Enablement
TB	Tera Bytes
TEOM	Tapered Element Oscillating Micro balance
TML	Transducer and Model Language
TNPCB	Tamil Nadu Pollution Control Board
TNSTC	Tamil Nadu State Transport Corporation
TSPM	Total Suspended Particulate Matter
UFP	Ultra Fine Particulate matter with aerodynamic diameter less than 0.1 μm
UPS	Uninterrupted Power Supply
URI	Uniform Resource Identifier
VKT	Vehicular Kilometre Travel
WD	Wind Direction
WMS	Workflow Management System
WNS	Web Notification Service
WPS	WRF Pre-Processing System
WRF-ASF	WRF-Advanced Software Framework
WRF-CHEM	Weather Research and Forecast - Chemistry
WRF-EMS	WRF-Environmental Modeling System
WS	Wind Speed
XML	Extensible Markup Language

Chapter 1

General Introduction

1.1 Urban particulate matter air pollution

Urban air pollution is increasingly becoming daunting and pressing among the various environmental issues of 21st century [1]. Fast developing countries such as India face severe air pollution due to its current high priority on economic prosperity and rapid urbanization [2, 3]. By 2025, it is estimated that in India the states like Tamil Nadu would be crossing more than 55% urbanization levels [4]. That would mean, with no sustainable interventions to manage the air pollution without disrupting the current path of economic development, further worsening the situation [3]. Particulate matter (PM) is an important and major constituent of urban air pollution [5]. They are non-gaseous airborne substances in solid or liquid form commonly denoted as aerosol, a colloidal entity [1]. PM in ambient atmosphere is characterized by its total mass concentration, size distribution, and chemical composition [2]. Generally, PM is classified according to the particle size; coarse PM (PM_{10} - aerodynamic diameter less than $10\ \mu\text{m}$) and fine PM ($PM_{2.5}$ - aerodynamic diameter equal and less than $2.5\ \mu\text{m}$) and ultrafine particles (UFP, $< 0.1\ \mu\text{m}$) [1]. Larger attention is given to the lower size fractions, especially $PM_{2.5}$ and UFP, due to their immense hazardous potential to human health for their high capacity to intrude into the body system. $PM_{2.5}$ and lower fractions are found to be in size distribution well capable to penetrate deep into human respiratory tract and cause various health effects. Coarser

particulate matter, PM_{10} , has also proven causative linkage with diseases such as asthma and its ability to absorb heavy metals, Dioxin and other semi volatile organic compounds, which are highly toxic and carcinogenic [1]. PM in ambient air is emitted either by combustion related anthropogenic activities, natural processes (Primary sources) or by means of gas to particle conversion in atmosphere (secondary sources) [2, 6]. Urban PM is largely anthropogenic in origin [7, 3] and it is emitted most from burning hydrocarbons in industries, vehicles, housing [8], agriculture [9], and from sources such as diesel based power generator [10]. Of the secondary sources, gaseous molecules in atmosphere undergo physical transformation into solid or liquid particles through physical processes of absorption, nucleation and condensation [1]. It is said that secondary sources are the major portion of urban PM concentration [11]. Recent studies also have revealed that secondary sources are major contributors of PM in most urban centres where stringent restriction on primary sources is imposed. However, their proportions keep varying according to the season; for instance in winter direct combustion for heating purpose are extensive the regulatory interventions on primary sources gets into an impasse [2]. Understanding on the formation mechanism of urban fine particulates is lacking and it seriously hampering the pollution episode forecast and intervention measures to tackle the impact [2, 12, 13].

PM in air is a major health hazard and environmental concern in urban areas [2, 14, 15]. Health impacts, from mild allergies to serious chronic disease and premature death, are attributed to urban PM air pollution [5, 2]. A recent study [16] estimates that during 2011, $PM_{2.5}$ exposure caused about half a million premature deaths in India. About 6% increase in premature death due to long term exposure to fine PM [2, 17, 18], mainly through cardiovascular and respiratory morbidity and mortality [5, 19, 20], is also estimated. Exposure to particulate matter is linked with Coronary Artery Disease through oxidative stress and inflammation [21, 22]. Physical properties of PM such as large surface area to mass ratio, high redox capacity and formation of radical species are linked to the inflammatory disease and cellular damage [23]. Chemical constituents of PM such as heavy metals, polycyclic aromatic hydrocarbon, and elemental carbon are linked to DNA damage, reproductive defects and carcinogenic effects [24, 25]. Urban PM is also linked with a broad spectrum of impact on normal atmospheric

functionality [2], on solar radiative transfer, cloud formation [2, 26, 27], altering the duration and albedo of cloud, precipitation and lightening [2, 28]. It is reported that PM contribute to the global climate change by absorbing the solar radiation and reducing the radiation flux at the earth's surface [5]. It has direct effects on atmospheric aerosol and trace gases to act as short-lived climate forcers [5].

1.2 Issues with current intervention measures

The discussion on urban PM air pollution and its impact show the immense need of commitment and interventions to tackle the problem. Environmental legislation imposed to tackle the problem has shown considerable success in reducing the particulate pollution and health impacts, especially in the developed countries [29]. For example in United States of America, the Clean Air Act - 1970 and resultant public health programs were instrumental in considerably reducing morbidity and mortality in the country. The monetary benefit from the interventions is estimated to be 30 US\$ return per one US\$ investment [30]. Similarly, European commission's regulations to tackle Sulphur dioxide pollution is estimated to reap a benefit of notable pre-mature death postponement and mortality benefits of 191.6 million € in monetary terms every year from 2000 in 20 European cities [31]. Imposing restriction on primary emitters of PM has consequential overall improvement in air quality and manifold reduction in ambient concentration of PM in many cities of the developed world [29]. However, this achievement is seldom duplicated in developing and under developed countries, especially India, where legislation against air pollution was imposed almost at the same time as elsewhere [3, 32]. The air pollution in India is in dangerous level and top most polluted cities in the world are in India [33, 34]. The Figure 1.1 shows the relationship of air pollution evolution with respect to urbanization and the role of interventions measures.

As the industrial development caused peaking in air pollution woes, which led to intervention measures such as emission control, stabilization of air pollutant concentration, followed by improvement in air quality, further application of modern and costly technology is required to sustain or further improve the quality [29, 35]. While the developed countries have shown

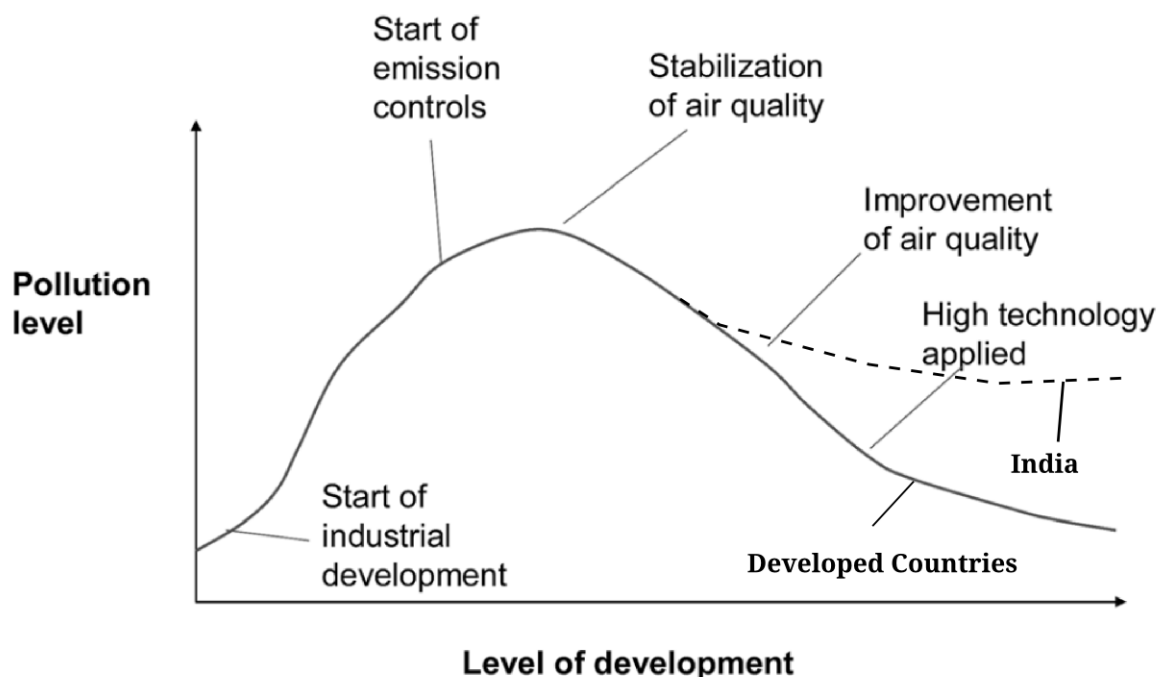


Fig. 1.1 Relationship of urban development and air pollution evolution, Adapted from Pandis et al. [29] with permission of The Royal Society of Chemistry.

greater drive in applying high cost technological interventions, countries like India is lagging much behind in striving for greater improvements in air quality or in application of technology to achieve or sustain it. In India, legislation such as Water Act -1974 and Air Act 1981 has greater role in environmental policies of India [36]. Such legislation gives provision for the formation of regulatory bodies such as Central Pollution Control Boards (CPCB) and State Pollution Control Boards (SPCB). However, weak institutional setups to enforce interventions that eventually leads to policy failure are important setbacks of effective implementation of regulatory provisions and interventions in countries like India [36–39]. Recent studies on implementation efficacy of environmental regulations with its desired outcome on health improvement such as on infant mortality reduction call for greater role of common public interest in influencing the regulatory bodies to enforce the regulations effectively through judicial enforcement and improvement especially in air quality aspect [36]. This warrants furthermore intensive attempts for identification of drawbacks of India's current intervention measures on tackling the air pollution problems. Some of the probable issues are discussed below. The

issues are not only reflective of the current regulatory interventions but also reflective of the societal and scientific approach to air pollution issues at large.

1.2.1 Reductionistic view

The urban level particulate pollution is being addressed mostly by regulating the sources such as industrial and transport sector [29]. Regulatory interventions such as imposing penalty on the polluters and temporally shutting down their operations are the widely followed measures. However, such interventions are in-sufficient considering the nature [29] and impact potential of particulate pollution. The multitude of particulate pollution sources, their interactions with interlinked actors in an urban setup makes the problem complex and causes hurdles in sustainable solution for it. Urban centres are complex social ecological system or a tightly coupled human natural system (CHANS) [40, 41], in which the particulate pollution issue is considered as disruptions of normal biogeochemical cycles of natural system [40], where human social and economic activities in particular production and consumption are both sources of particulate pollution and subjects of the resultant impacts [42]. Intervention measures currently followed are devoid of the linkages while they emphasis compartmentalized mitigations in the human natural system, which can be non-effective and unsustainable [43, 44]. Current intervention measures such as stipulation of emission standards are addressing the primary sources of air pollution and not its secondary sources. This is also the case for spatio-temporal variability, dynamics and local condition variation in pollutant emissions and their stability in the atmosphere. Particulate pollution rather than being considered as simple linear relationship between the emission and atmosphere, the intervention measures have to consider that as a resultant of dynamic and adaptive system. It has to be approached and managed taking into account the uncertainties in the system and its resultant outcomes of interventions [45, 46]. This requires integrated solution involving all stake-holders / components of the system with understanding on its complexity and participating components' interrelationships, uncertainty and dynamics [45].

Table 1.1 Characteristics of various knowledge systems

Characteristics	Folk	Authoritative	Scientific
Sources	Revelation	Authority	Empirical observation
Monopoly	Extensive	Extensive	Limited/ ideally absent
Scrutiny	Limited	Discouraged	Encouraged
Manipulation	Limited to substantial	Limited to substantial	Very limited to absent
Sound/Unsound	Moderate	Moderate	Very high
Rate of growth	Slow	Moderate	Very fast

1.2.2 Authoritative approach

Important characteristics of current interventions are its top down authoritative approach by regulatory organizations [47]. Federal systems as in India is largely vested on the powers of central elected government directing it to state and local level governing bodies [47]. The interventions are carried out through stipulating set of air quality standards and ensuring standard compliance of it by polluters [48]. The information necessary for this mode of mechanism is largely drawn from official sources, enjoying substantial monopoly and limits to scrutinize the veracity of information [49]. That can generally lead to detachment of multiple interacting components of socio ecological system especially the local community. The knowledge generated by this mode of approach is the basis of authoritative knowledge system. The Table 1.1 presents a comparison of such a system with other two systems [49]. The knowledge needed to address the air pollution has to be of scientific knowledge system [49–51]. The current authoritative approach can be made into more of scientific knowledge system by inculcating transparency and community involvement to achieve needed scrutiny and to resist the potential manipulations [52, 38, 53] from the authoritative sources to safeguard the interests of the authority. This aspect has to be addressed more for providing crucial knowledge-base for decision-making in air pollution management. In fact, the information uptake by decision makers should be determined by its credible, salient, and legitimate features [50, 51]. In the case of air pollution problems, the individuals' activity should have larger role in management; the information provided by authoritative agencies has to carry with it such traits.

1.2.3 Incompatibility

The interventions to address air pollution have to be a collective action of different organizations and based on heterogeneous information sources. Organizational and information source incompatibility results in ineffective co-ordination and thus the management [54, 55]. Such incompatibility is an obstacle in the management [56], in decision-making or follow-up monitoring on the intervention measures. Scientific interdisciplinary effort is essential for decision making in terms of spatial temporal risk identification and health impact outcomes. This requires improved data sources that can be better organized to derive the knowledge. The information incompatibility from its origin hampers the discovery, interfacing and deriving the decision support information from the heterogeneous sources [57] that collect or maintain the data largely for its own requirement without considering its interdisciplinary / inter-organizational utility / application and results in incompatible, non-usable data format, inadequacy in spatial and temporal data coverage etc.

1.3 Data Intensive approach

Massive technological developments have provided ample tools for high-throughput data collection and large scale computing [58], which leads to unprecedented transformation in production, collection and curation of the data especially regarding various components of socio ecological system [59]. This is a significant transformation from earlier state of expensive, scarce and limitedly accessible data to low cost, plenty with open accessible data source [60]. The resultant outcome is *Big data*, which is characterized by large volume, variety and velocity in data streams representative of various environmental phenomenon. The recent developments in internet communication made *Big data* a highly distributed and easily accessible entity [61]. For example the US national Climate Assessment (NASA project) 2013 data size reaches about 1,000 Tera bytes (TB) whereas the Intergovernmental Panel on Climate Change (IPCC) fifth assessment report is based on the data of about 2,500 TB [62], indicating the massiveness and scale of data pertaining to the earth system generated per year. This vouches for specific challenges pertaining to the access, process, analyse and disseminate such massive data [63].

That also offers greater opportunity in utilizing the data to manage better the socio-ecological system and its problems such as particulate pollution.

In scientific progression, there are three consecutive paradigms identified such as of empirical (observation on natural phenomenon), theoretical (using theoretical models, conjectures and generalization) and computational (using simulation) [64, 65]. Advent of *Big data* is termed as a fourth paradigm, which is gaining importance, outcome of large accumulation of data from observations, experimentation and simulation [64, 65]. This is data intensive scientific discovery (DISD), using data to gain insight into the various issues of society. In environmental sciences, the DISD is complementing the scientific understanding of the earth system in real world context to manage it by supporting the decision-making on socio-ecological system management [66]. Some of the characteristics of this approach are its priority towards societal need rather than the fundamental science's question driven approach, prioritizing the understanding of complex interactions between societal actions and earth system. In that prioritization, the approach is consequential and adaptive according to the dynamics and socio-ecological systems' complex interactions' uncertain outcome, which is constrained in terms of time and resource by society's requirement on knowledge for decision-making.

In air pollution management, this approach provides a set of technological tools to address the management questions. Such tools are tools to integrate the data on air pollution, monitoring and air quality forecast model with societal information to assess the alternative outcomes for applying various intervention measures. Moreover which enhances dissemination of spatio temporally relevant real time data for awareness creation and community involvement. This requires the data to be extensive in spatial and temporal context, scalable, robust and better shared for integrated assessment on various knowledge-bases on socio-ecological systems [67]. In this respect, the data intensive approach can address drawbacks in intervention measures on particulate pollution.

1.3.1 Complex context

Improving the information to understand better the complexity of socio-ecological system and enabling informed choice among common public are widely recommended measures to sustainability [68]. This instigate the data intensive approach based on real time observation networks in air pollution (both in-situ or ex-situ), simulation output and volunteer/common public information to integrate for better understanding and to ensure the effectiveness of air pollution intervention or innovating new measures [69]. To achieve that, it is essential to make the system adaptable to forces of change in air pollution management, which require better overview on uncertainty and dynamisms of the components' interactions in the system. The data intensive approach pursued in this aspect can generate dynamic perceptions, social consensus and adaptive strategy [70]. It gives required tools for understanding complex interaction between social ecological systems leading to air pollution. Through this approach, with the advent of internet communication enables sharing of experience acquired in different parts of world in addressing air pollution [1]. Analysis of this data with reflection on local conditions gives new intervention perspectives. It provides insight on scale dependent spatio temporal variability in air pollution concentration, characteristics of real world exposure, and disentangles the exposure variability based on socio-economic status and spatial variability. Another advantage of data intensive approach is to gain insight on alternative scenario assessment based on the different intervention measures and adopt most suitable and adaptive control technologies and its improvement. More over the real time data [71] provided through this approach gives opportunity to address the dynamics of air pollution concentration and time bound intervention measures.

1.3.2 Collective action through openness

The openness provided by data intensive approach generates community involvement. It instigate the generation of scientific knowledge system by encouraging scrutiny and evaluation of the collected data [49]. The data openness not only facilitates the wider access of the data for reuse but also the capabilities to collect, organize, analyze and draw inferences from the

data. This is based on three principles: openness, participation and collaboration [60, 72]. This is envisaged through inculcating transparency, sharing and working together to improve the value of data for societal applications [60]. In this respect, the urban particulate pollution can be addressed by data intensive approach, which provides necessary tools for networking the components of urban system [73], and generate necessary rationale for community involvement and interventions based on consensus and collective action.

1.3.3 Compatibility through data-interoperability

The data-interoperability addresses the incompatibility in information sources and for a certain extent in organizational level. Data-interoperability is one of the important technical challenges of data intensive approach [60]. Stipulating common standards for data and ensure the compatibility and integration of heterogeneous data is a step to address the challenge. Especially in addressing particulate pollution problem, require integrating wide variety of data sources such as from real time sensors, remote sensing platforms, simulation outputs and volunteered contributed pollution level details [67]. Sensor web enablement specification is a web standard to enable interoperability in data which is communicated through web platform [74]. It envisages transformation of internet with web of documents into web of sensor with data to enable the discovery, processing and analysis of the heterogeneous data seamlessly in web. It is an offshoot of open data movement, which looks for ways to make data capture and analysis workflow much simpler, without depending on external resources. The Figure 1.2 shows the technological challenges of sensor web [57]. The challenges comprised of integrating monitored, modelled, and volunteered information. The integration processes the automation, fusion of heterogeneous data source with inherent uncertainty information with event based architecture, and project the cyber infrastructure development to address the dissemination of collected data and devise knowledge for decision-making. This can enable the data intensive approach to address the particulate pollution problem by integrating diverse data sources and devising application and providing information for decision-making [67].

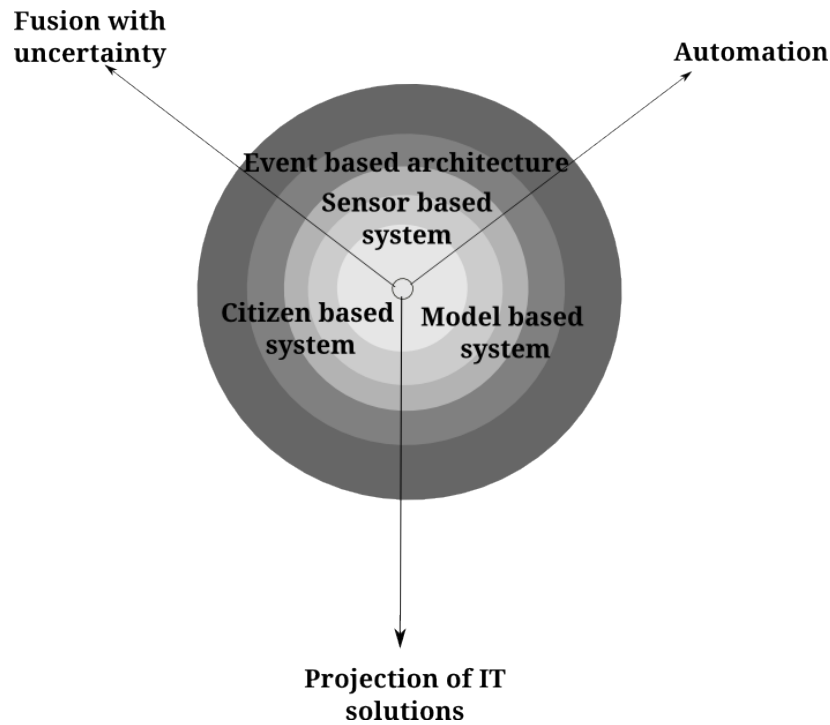


Fig. 1.2 Technological challenges of sensor web enablement[57]

1.4 Study objectives and thesis organization

Considering the need and advantages of data intensive approach for urban particulate pollution, an attempt was made for addressing it in Coimbatore, a second tier rapidly developing urban conglomerate in the state of Tamil Nadu, India. As an initial attempt, the study focused on developing basic infrastructure for data intensive approach. Lack of spatio-temporally relevant and reliable data on particulate levels is a major issue in managing particulate pollution in developing countries. The data existing in non-interoperable formats to a great extent further hampers knowledge generation utilizing the available data for effective control and management of air quality. These conditions prevailing mostly in Indian cities especially of second tier urban centres where the real time air pollution information is still lagging is significantly hampering data intensive approach to manage air pollution. Continuous monitoring by low-cost dust profilers and real time air quality modeling may be an alternative and relatively economical approach to the current practice of using expensive and intricate air quality monitoring and information systems. Which greatly avoid trickling of such system establishment in most of the

urban centres. Reliability of generated data can be ensured by optimal sensor calibration and model validation routines. Necessary data interoperability measures can be enabled by building the routines upon web data standards. In this context, the present study explores various ways for generating real time data for urban centres with interoperability measures for particulate pollution problem in an urban centre of developing country. The major objectives of the study were as follows.

1. Develop a real time particulate pollution monitoring system using low cost commodity sensors and assess its effectiveness in the study area
2. Attempt a real time particulate pollution modeling system for the study area using WRF-CHEM and address its computational requirements
3. Demonstrate application of interoperability measures on real time particulate pollution data

In the thesis, each objective is addressed as follows. The fourth chapter discusses about the development of real time particulate pollution monitoring system using low cost commodity sensors and assesses its effectiveness in Coimbatore area. The Chapter-5, discusses about development of emission inventory(EI), an important data input for numerical air quality modeling system. The sixth Chapter is an elaboration on real time particulate pollution modeling using Weather Research and Forecasting system enabled with Chemistry (WRF-CHEM), a numerical model for air quality forecast for Coimbatore region using the developed emission inventory. Chapter-7 discusses interoperability measures for real time particulate pollution data from monitoring and modeling with a brief demonstration on interoperability advantages of the data. Next chapter gives background of study area, Coimbatore

Chapter 2

Study area: Coimbatore

Population

Coimbatore is the second largest urban conglomerate in the state of Tamil Nadu, India. The city is a fast growing second tier urban centre in India. As early as 3rd century BC to 4th century AD, the area of current Coimbatore region is known to be ruled by semi-independent chieftains under the patronage of Chera kingdom of southern India [75]. Called as Kongunadu, the Coimbatore region was part of various kingdoms of southern India. The Coimbatore municipality was established in British colonial era in 1866. During the dawn of twentieth century, Coimbatore experienced a rapid expansion in textile and allied machinery industries after their decline in then erstwhile Bombay, then textile capital of India. Surplus and cheap electricity from the nearby hydroelectric power stations and flourishing cotton cultivation ushered the region's rapid industrialization and urbanization. The headquarters of the Coimbatore district, in 1981, Coimbatore was upgraded into a Municipal Corporation. Coimbatore city is the sixteenth largest urban agglomeration in India with total population of 2,136,916 with males constituting 50.08% of the population and females 49.92% as per the 2011 census [75]. The population within Coimbatore Corporation limits is 1,601,438 with a sex ratio of 997 females for every 1,000 males, which is above the national average of 929. As per the census of India 2011, Coimbatore district has 945,943 houses, of which 237,155 are rural and 708,788 urban. Among the houses, around 28 % of the houses in Coimbatore district were

using PM / air pollution emitting fuel sources such as fire wood, crop residue, cow dung cake, charcoal and kerosene for cooking [75]. The period after 1991 is the era of globalization and liberalization in India that ushered in wide-ranging changes. As elsewhere in the world, Tamil Nadu recorded increase in the proportion of people living in urban areas, especially during the last two decades [76]. The Figure 2.1 and 2.2 shows the population increase in different census period in Tamil Nadu and Coimbatore during 1901-2011 respectively. During the same period, Coimbatore also expanded in terms of size and scale of urban agglomeration. Current urban population of Tamil Nadu is 49%, the highest among other Indian states. In Coimbatore district predominant part of population (74.9 %) are urban.

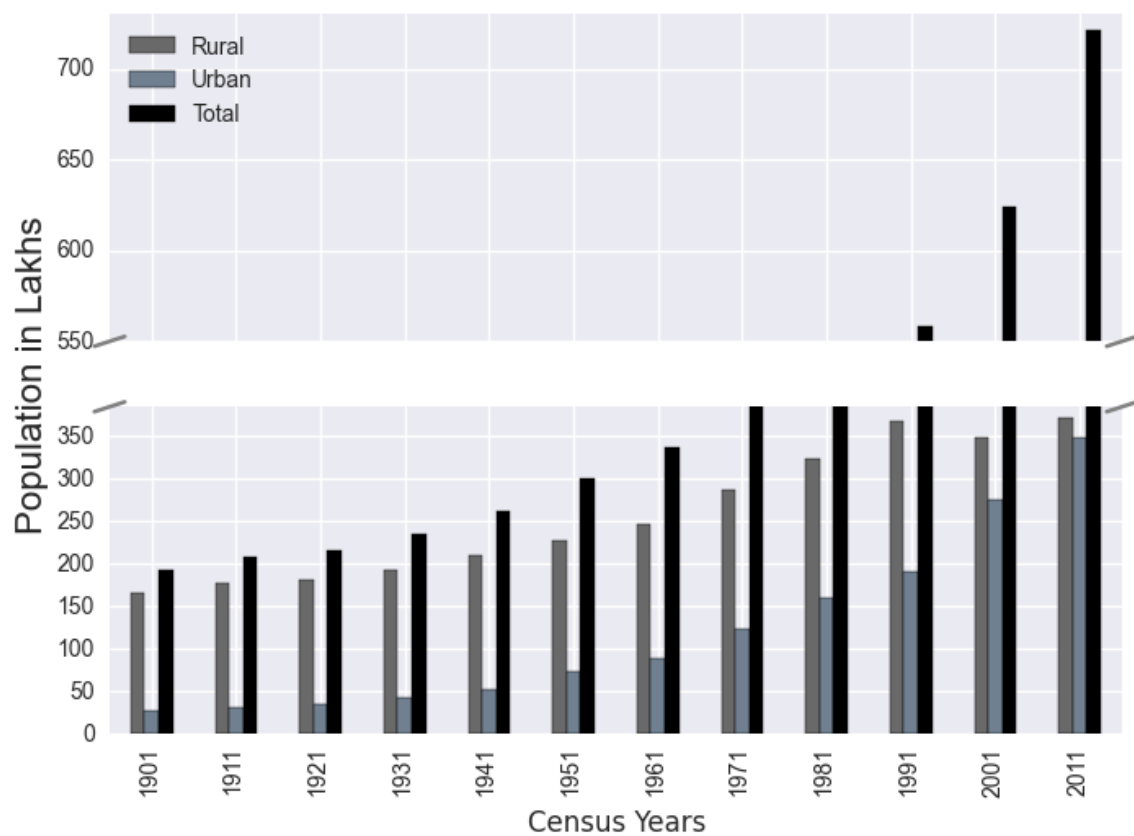


Fig. 2.1 The population increase in Tamil Nadu during 1901-2011 [76]

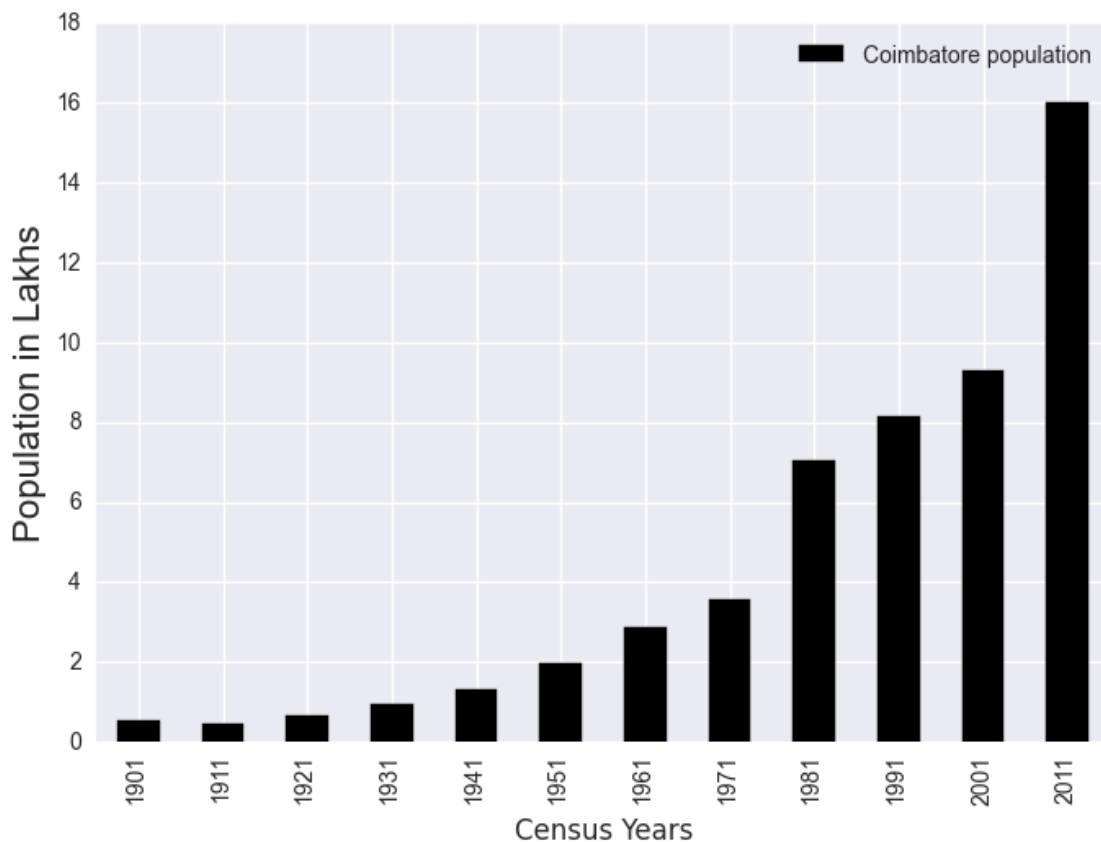


Fig. 2.2 The population increase in Coimbatore urban during 1901-2011 [76]

Geography and Climate

Coimbatore is located adjoining the Western Ghats mountain range of Indian subcontinent on its south west. It lies at about 411 meters (1349 ft) above sea level. The river Noyyal, originating in the Velingiri hills in Western Ghats, courses through the city. The river once perennial has turned to a seasonal one in recent decades. The urban agglomeration of Coimbatore covers an area of 246.75 km^2 while the district in total spreads over 4723 km^2 , of which 1052 km^2 is reserve forests. In the current study an extent of area covering the urban agglomeration between $10^\circ 48' 25''\text{N}$, $76^\circ 43' 21''\text{E}$ South West and $11^\circ 18' 27''\text{N}$, $77^\circ 19' 21''\text{E}$ North East was considered as study area. The city is located in the eastern opening of the Palghat gap that is the widest one among the three major mountain gaps in the Western Ghats. This position made the city significantly important for historical trade links. Geographically the Coimbatore can be divided into west hilly forested region and east plain dry region where the majority of urban region lies.

The region has predominantly black soil, most suitable for cotton cultivation that perchance is the single most important factor that led to industrialization of the city inviting the moniker the *Manchester of the south India*. Northwestern parts of the region with red loamy soil is suitable for agriculture and currently holds large number of brick kilns that draw upon the surface soils for brick-making catering the huge needs of the growing city.

The geographical location of Coimbatore in Palghat gap also have significant role on the region's weather and climate. The region has a tropical wet and dry climate with higher portion of rainfall recorded during the northeast monsoon from October to December. Due to the Palghat gap, the region also receives the southwest monsoon during June to August. The region receives an average annual rainfall of around 700 mm, in which 47% during the northeast monsoons and 28% during the southwest monsoons. During 2014, Coimbatore region recorded mean maximum temperature ranging from 35.2 °C to 36.3 °C and the mean minimum temperature ranging from 17.9 °C to 21.0 °C. The region experiences high humidity with most of the month's average relative humidity for late night and early morning exceeding 75 %.

Agriculture

One of the major factors in industrialization and urbanization of Coimbatore region is plenty of Cotton cultivation. The suitable black soil, tropical dry climate and cheap labor ensured flourishing cotton crops and allied works such as cotton ginning and spinning, foundries and industries manufacturing machinery for the booming textile industries. Currently cotton cultivation has considerably reduced in the region owing to various reasons; similarly associated industries. The red loamy soil in North Western part of the Coimbatore is suitable for horticulture crops, cereals and other crops, and still around 48% of geographical area of the district is under agriculture. During 2014-2015, major crops cultivated were Coconut (83789 Ha), Corn (28458 Ha), Banana (8115 Ha) and pulses (4933 Ha). The productivity of crops in the region is highest in the case of Sugarcane (107430 Kg/Ha) followed by vegetables and fruits such as Cabbage, Banana, Grapes and tuber crops such as Tapioca (22659-47000 Kg/Ha) [77]. Crop residue generation is directly proportional to agriculture productivity. Burning agriculture crop residue is a major source of air pollution emission, especially of particulate pollutant, in India

[78]. Burning agriculture crop residue is a commonly followed waste management technique in India. This is especially in the case of sugar cane cultivation. Large quantities of residues generated in the fields are generally burned, in various parts of Coimbatore district, after the harvest to initiate the next cycle of cultivation. This is also seen in the case of corn and cereal crop fields as well. The forests surrounding the northwestern part of Coimbatore region also acts as a source of air pollution from biomass burning due to summer forest fires that are mostly set by humans.

Industries

Burgeoning urbanization in Coimbatore region is primarily due to industrialization in early times and later by commercial establishments, real-estate ventures and large number of education institutions. The early major drive for industrialization and ensuing urbanization was the construction of Pykara Hydro electric power project (PHEP) situated towards the western part of Coimbatore, now in Nilgiri district, during 1932. The flourishing cotton production and the power from PHEP boosted development of textile mills and textile accessory machine manufacturing units. Thus, Coimbatore evolved into a cluster of large textile mills and related industries in due course. Currently as per the Coimbatore district profile handbook-2015 [79], there are 4225 working factories, 413 medium scale industries and 12873 small scale industries employing around 186820 workers. Of these, the major ones are textile and allied industries, information technology, manufacturing automotive parts, motor pumps, wet grinders, jewellery making, and cement and stone crushing units. India's oldest cement manufacturing unit is located in the southwestern part of the Coimbatore district, the ACC limited, formerly Associated Cement Company, established in 1936. Nearness to the Western Ghats offers necessary raw material for stone crushing units and another major cement-manufacturing unit bordering the administrative boundary of Coimbatore district. The northwestern parts of Coimbatore are home to hundreds of brick kilns, and other small scale and cottage industries having poor facilities for controlling emission of polluted air from their chimneys. Coimbatore has large-scale flourmills with altogether grinding capacity of 50,000 MT per month catering the needs of southern states of India. Thus, the major air polluting industries in Coimbatore are

Cement manufacturing units, foundries, stone crushing and brick kilns. Coimbatore host special economic zones and industrial hubs in which Kurichi industrial cluster (SIDCO) is considered as a critically polluted area by CPCB [80], based on the recently developed Comprehensive Environmental Pollution Index (CEPI) that weighs the pollution potential of the factories located in industrial clusters. Foundries located in clusters are large air pollutant emitters. There are 722 foundries in the Coimbatore district catering the needs of various manufacturing industries and others.

Transport

Coimbatore is a major transportation hub in southern India, well connected in terms of Road, railway and air. The corporation limit is bestowed with six major arterial roads. Coimbatore district has a total road length of 13904 km. Of this, major section is village roads (54.9%). That is followed by Corporation (15.2%), town (13.2%), state highways (8.4%), municipality (4.1%), national highways (2.6 %) and other forest roads (1.6%). Most of the roads are un-paved and it is a major source for fugitive dust emission from transportation. The vehicular growth and ensuing particulate pollution emission is a major concern in Coimbatore. Tamil Nadu is reportedly having the top and fastest vehicular growth of all the Indian states; during 2001-2011, vehicle number has increased by 202.9 % whereas the population of Tamil Nadu has grown by 16.13% only. This trend is reflected in Coimbatore as well; during 2000 - 2016, the vehicular growth is 386.6%. The Figure 2.3 shows the number of vehicular registration in Tamil Nadu during 1993-2014. The Figure 2.4 shows the number of vehicular registration in Coimbatore during 2000-2016. It is observed that the larger portion of the vehicular population in 2016 is composed of two wheelers(81.3%) followed by cars/jeep(13.8%), goods vehicles(2.4%), 3 wheelers(1.1%), other category vehicles(1.0%) and buses(0.4%).

Particulate air pollution in Coimbatore

Industrialization and resultant rapid urbanization is the main cause of PM pollution in Coimbatore. Emission from foundries, cement, stone crushing and brick kiln are major sources. Transport related sources, the growing vehicular population with stagnant infrastructure devel-

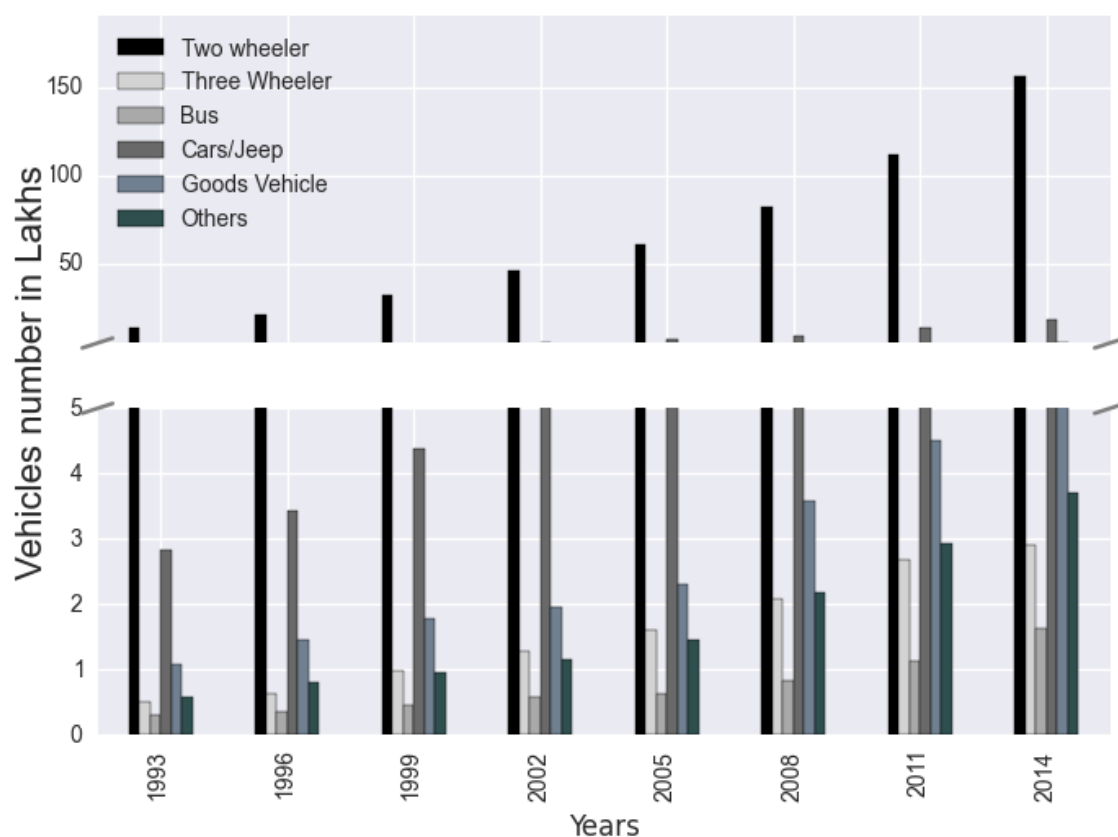


Fig. 2.3 The number of vehicular registration in Tamil Nadu during 1993-2014 [81]

opment, are the second major emitter of PM. That is followed by use of particulate emitting fuels for cooking and lighting purposes, and crop residue burning. Major studies on particulate pollution in Coimbatore region was carried out during 1999-2001 by Mohanraj and Azeez [82]. The study reports respirable suspended particulate matter (RSPM) $30\text{--}149\text{ }\mu\text{g}/\text{m}^3$ with an average of $71.3\pm22.26\text{ }\mu\text{g}/\text{m}^3$. The total suspended particulate matter (TSPM) was $24.4\text{--}460\text{ }\mu\text{g}/\text{m}^3$ averaging $110.8\pm69.15\text{ }\mu\text{g}/\text{m}^3$. High concentration of RSPM and TSPM were observed in urban stations during 1999-2001. Follow-up studies at the same stations reported PM_{2.5} ranging between $27.85\text{ to }165.75\text{ }\mu\text{g}/\text{m}^3$ with an average of $76.28\text{ }\mu\text{g}/\text{m}^3$ [83]. These studies were carried out using high volume air sampler with filtration method and fine particulate monitor. Mobile monitoring using a hand-held particulate counter was carried out by us in August to September 2015. Around 75 locations in various land use and land cover class of Coimbatore urban and rural region were sampled for PM during 09:00AM to 06:00PM. The

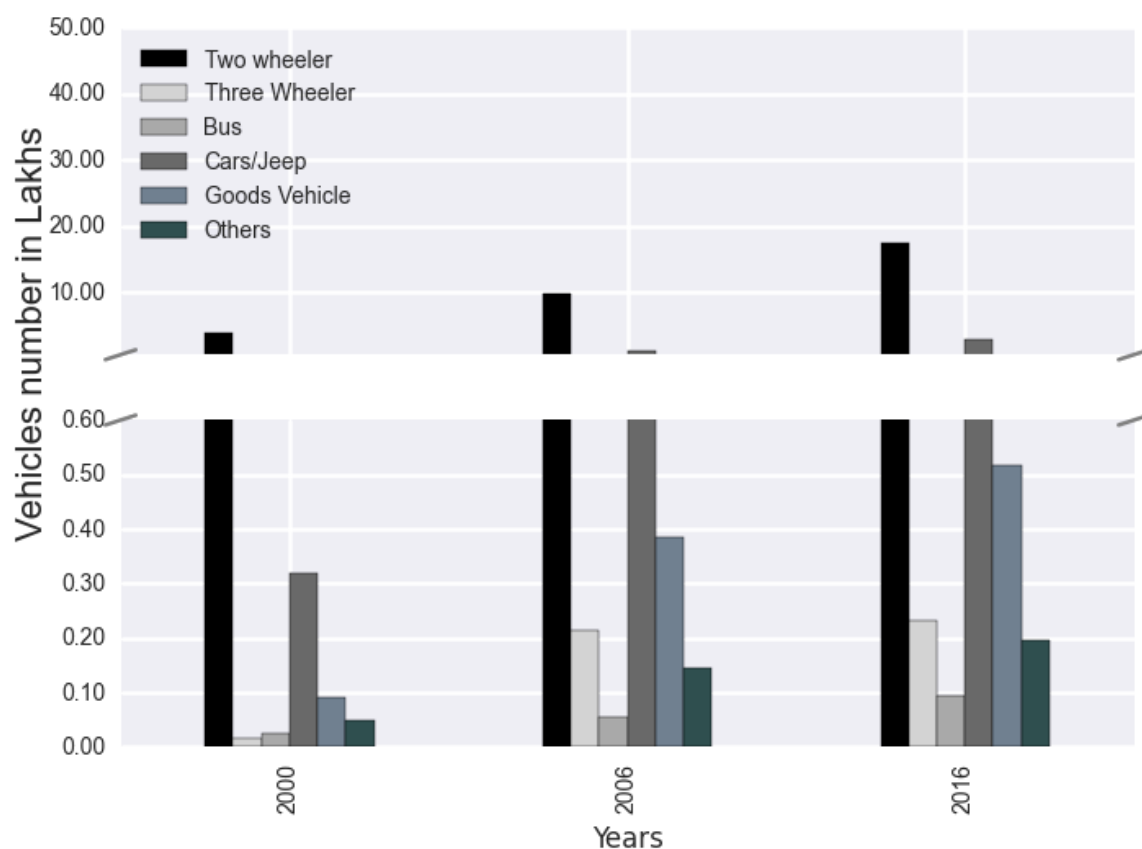


Fig. 2.4 The number of vehicular registration in Coimbatore during 2000-2016 [81]

study reveals the PM_{2.5} concentration varying between 3 to 1013.2 $\mu\text{g}/\text{m}^3$ with an average of $53.69 \pm 44.42 \mu\text{g}/\text{m}^3$. PM₁₀ varied between 20.60 to 6931.60 $\mu\text{g}/\text{m}^3$ with an average of $473.42 \pm 629.13 \mu\text{g}/\text{m}^3$. Long-term monitoring of temporal variability of TSPM and RSPM is being carried out by the regulatory body, Tamil Nadu State Pollution Control Board (TNPCB). The Figure 2.5 shows the temporal variability of TSPM and RSPM during 2003-2012 in three classes of stations, namely residential, mixed and industrial area in Coimbatore [84]. Studies on Polycyclic aromatic hydrocarbons (PAH) and heavy metals bounded with PM shows varying concentration in Coimbatore urban region. The study [85] carried out during 1999-2001 reported 20-172 ng/m^3 total PAH with an average of 90.37 ng/m^3 in PM₁₀ samples. The study carried out during 2009-2010 [83] reported total PAH concentration ranging from 27.85 to 165.75 ng/m^3 in PM_{2.5} samples. In RSPM, quantity of heavy metals such as Zinc was highest followed by copper, lead, Nickel, Chromium and Cadmium [86]. The concentrations

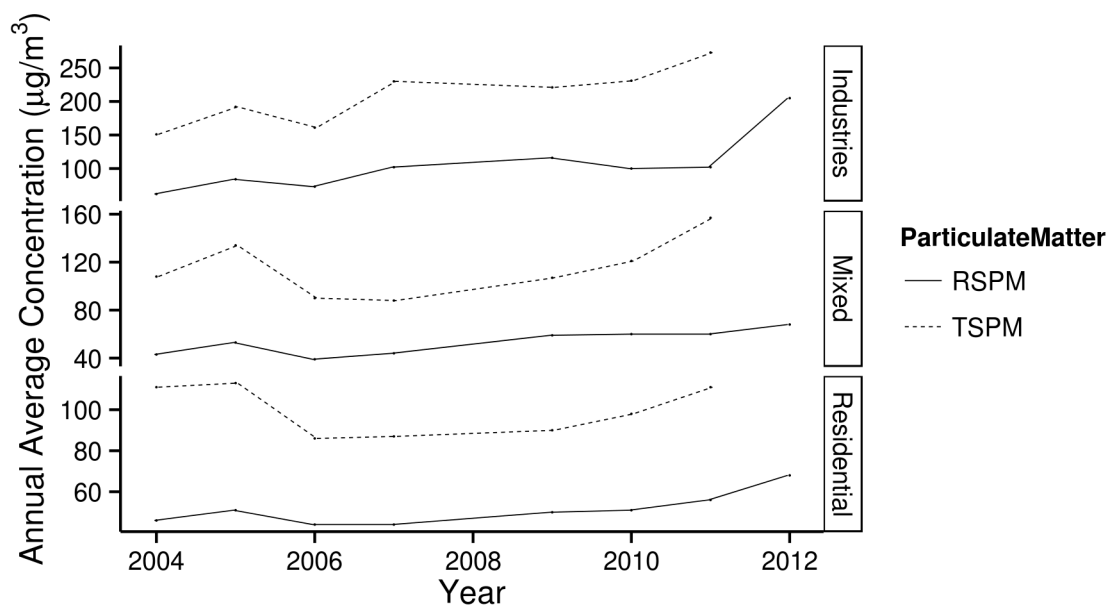


Fig. 2.5 Ambient particulate matter concentration in different stations of Coimbatore during 2004-2012 [84]

were varying between below detectable limit (BDL) to 2147 ng/m^3 with highest average of $481\text{--}544.3 \text{ ng/m}^3$ in Industrial areas. The studies reveal rising PM in temporal scale with potentially hazardous heavy metals and PAH bound on to the particulates.

As in any other urban centres in India, interventions against particulate pollution in Coimbatore are largely issuing the emission standards for various pollutants such as for industries and vehicles. These standards are tracked by ambient monitoring stations, stack monitoring or periodical vehicular emission checking for general compliance by TNPCB. Recent comparison with 41 cities (of one plus million population urban centres) in India indicated that Coimbatore is relatively better in air quality in terms of air quality index (AQI) under National Ambient Air Monitoring Program of CPCB. Coimbatore recorded 99% of good AQI days for a 157 days ambient monitoring in 2015 and it is highest in the country [87]. However, being a second tier urban centre the effectiveness and sustainability of regulatory interventions for air quality management are much to be developed. There is historical under investment in institutional building in urban India indicated by serious infrastructure and service deficits especially in the case of second and lower tier urban centres [88–90]. This hampers the management capacity of

local bodies to address the air quality management in urban centres, which is currently carried out by top-down approach of regulatory body interventions. Moreover, the local bodies are widely criticized for poor governance process, corruption and public exclusion [91]. This devalues the local governance and accountability, crucial for sustainable urban transformation [92] much needed especially for the particulate pollution type that involves complex management issues.

Chapter 3

Real time particulate pollution monitoring

3.1 Introduction

Particulate pollution and its resultant health effects are considered one of the large health burdens in the world [93]. An estimate of 3.2 million premature deaths and over 76 million lost disability-adjusted life years are caused due to outdoor fine particulate matter pollution [93–95]. It is causative for short term [96–98] and long term health effects [96, 99–101]. In India only in 2013, ambient air pollution caused 587,000 deaths [102]. Major source of the pollution is attributed to combustion process and its resultant particulate matter emission [102]. Particulate pollutants (or suspended Particulate Matter - PM) are particles with low aerodynamic diameter. The most commonly discussed categories of such particulates are $2.5\mu\text{m}(PM_{2.5})$ and $10\mu\text{m}(PM_{10})$. The small size particles especially those less than $PM_{2.5}$ is capable of penetrating deep into respiratory systems area, where gaseous Oxygen / Carbon Dioxide (O_2/CO_2) exchange happens. That causes oxidative stress, including mitochondrial death, which leads to multiple health effects [103] and morbidity of various types. Recent studies have shown wide range of human activities are causative for particulate matter emission into the air making that (PM) an important indicator of air quality related health effects [104–

106]. It is observed that PM exposure is higher than the level suggested by earlier understanding [105, 107]. Particulate pollution monitoring is an important tool in air quality assessments, and for identifying measures for management of pollution. Conjoined with pollution emission inventory and concentration forecast, it forms the basic components for human exposures and health effect assessments [108]. However, current infrastructure for particulate pollution monitoring especially $PM_{2.5}$ is found to be inadequate to address its concentration variations with respect to spatio-temporal variability, human time-activity patterns and to figure out personal exposure levels [93, 109].

The size distribution of the PM is one crucial aspects of monitoring and the environmental and health implications of the particulates. Hence, PM is monitored primarily with objective of finding the concentration or size distribution [110], the analysis of chemical nature and related aspects being only next in priority. Direct methods of mass ($\mu\text{g}/\text{m}^3$) or count (number of particles/ m^3) are generally adopted [110–113] in measuring the concentrations. However, most instruments, employed to study health effects and to comply with regulatory monitoring, are based on mass concentration methods. It is monitored in terms of standard measures in two size bins of $PM_{2.5}$ and PM_{10} [114, 115], by manual air filtration, optical methods or by continuous and integrated instruments such as Beta Attenuation Monitors (BAM or β -gauge) and Tapered Element Oscillating Micro balance (TEOM) [116, 107, 117]. However, high investment and running cost for automatic real time instruments such as of BAM and TEOM largely hinders their wider deployment [112]. For example in India, most of the large cities, invariably the second tier urban centres are yet to receive any real-time monitoring facilities; instead, they are relying on time-consuming manual methods. In Coimbatore, a fast growing second tier city, particulate pollution is monitored at three representative land cover / land use locations using high volume air sampler. The operation is carried out manually using a high volume air samplers fixed with glass microfibre filter paper followed by gravimetric method. The data frequency is only three to six times per day and based on availability of manual operator. This mode of particulate pollution monitoring suffers from a couple of serious limitations. Firstly, the data thus generated face temporal and spatial constraints that reduce its utility in understanding the health effects, and secondly that method cannot amply cope up with the timely requirement for

regulatory measurements and in issuing well-timed advisories. Thirdly, with off-line operations, it has an inherent time lag and opacity in data processing and updating. Such constraints limit the utility of the outputs in terms of required temporal resolution and real time dissemination and consumption for the internet era. Particulate pollution monitoring by particle concentration is lately gaining importance as increasingly its essentiality and advantages [96, 118, 119] are being recognized. It is observed that along with mass concentrations, concentration in terms of number is also an important indicator for deducing health effects and for comprehensive understanding of urban air quality [96, 120]. Furthermore, the particle counters for monitoring is being widely tested for their explicit advantages over mass concentration methods in terms of low cost, ease of use and very short response time in measuring particle concentration in the ambient air.

In recent years, there is a rapid proliferation of low cost open source hardware tools such as micro controllers, single-board computers in the market. Simultaneously there is a growing online community interest, forum discussion and publication of tutorial and blogs using these tools. This trend equips common public or amateurs to extend their capability to develop instruments that can collect research grade data using commodity (commercial and off the shelf) sensors [121–123]. Particle counters based on light scattering principles are used in industrial clean room service operations and their mass production, for their increasing demands, ensures their wider availability and low cost [124]. Monitor based on these counters are gaining acceptance as a supplemental resource to compensate for the inadequate PM measurement data [116, 125]. There are several recent studies showing promising data validity of monitors developed using such particle counters and open source hardware comparable to high-end research grade particulate monitors [124, 93, 126, 127] that are exorbitantly expensive and ill-affordable. However, those studies are emphasizing the requirements of characterization of such monitors in a wide variety of geographical / environmental setups and particulate matter sources [128, 93, 122, 127]. In this context, the current study aim to characterize the effectiveness of a customized particulate matter monitor for Indian scenario. It is based on a low cost commodity monitor fabricated with locally available open-source hardware tools and

expertise. This chapter discusses development of the real time particulate matter monitor and evaluates its data validity and field deployment effectiveness.

3.2 Methodology

3.2.1 Monitor development and field deployment

Dylos™ model number DC1100 is a patented air quality monitor commercialized by Dylos Corporation (USA). Its primary market is indoor air quality service industry such as clean air room testing and workspace particulate pollution assessment. The monitor has a laser based dust sensor that provides a continuous reading of particulate counts in two size bins of $PM_{0.5}$ and $PM_{2.5}$. The particle count readings are displayed in the screen attached to the monitor. It has nine pin communication (COM) port, which can connect to a computer to store or transmit the data to where it is required. In addition, the monitor has inbuilt memory of limited data storage. The bin $PM_{0.5}$ measures particles of size ranging from $0.5 \mu\text{m}$ to $2.5 \mu\text{m}$, while $PM_{2.5}$ bin measures particles of size ranging from $2.5 \mu\text{m}$ to $10 \mu\text{m}$. The measurement is carried out on the air, drawn by a small fan, which enters through a moulded funnel and baffle structure with sample volume size of 0.01 ft^3 , in a continuous process. The particle count in the airflow is done by a 650 nm wavelength laser beam. The rays scattered by the particles are captured by a photo diode positioned at an angle to capture the diffracted lights. To reduce the cost of instrument, there are no optical lenses or other focusing optics, particle size selector and airflow sensors in the monitor. The particle size selection is effected by an algorithm. The algorithm, which can differentiate the signals from scattered lights and the differing peaks width of the signals [126], also compensates for the possible fluctuations in the flow rate.

To customize Dylos™ into a real time ambient particulate monitor, it was equipped with a data collection, storage and real time data communication setup. To capture the raw readings from Dylos™ and transfer it into a data storage system, a serial USB (RS232) COM port plug was used. This data, as particles count number per 0.01 standard cubic feet (ft^3 or scf), is from the photo diode post processed signals. Raspberry pi Single Board Computer (SBC) was used to store, filter and communicate the data in real time. The SBC is programmed with

Python code to control the time resolution of data collection from the DylosTM. For the field deployment, the whole setup of monitor with data storage and communication was housed in an all weather steel box, specially designed for the purpose. To make the monitor design simple, an in-line power connection was used. To avoid the long power outages, the in-line power was backed with a 24 AH battery and 65 W inverter. To evaluate the functionality of the developed / customized real time particulate matter monitor, four monitors was deployed in various parts of Coimbatore city. Considering the safety of the monitors, it was placed mostly in terrace of buildings, at height ranging between 1.5 meters and 7 meters. The field-testing was carried out from 24 January 2014 to 20 July 2014. The field evaluation focused on assessment of data usability and functionality of the monitor.

3.2.2 Real time data communication and management

The data from the monitor was communicated to the server by the mode of SMS (Short Message Service) or Representation State Transfer Application Programming Interface (REST-API). The SMS mode was adopted for the stations that had poor reception for Internet data. The SBC of the monitor was attached with data card and equipped with SMS send/receive software to send the data from the DylosTM at an interval of every 15 minutes. For REST-API, the SBC with data card was setup with Internet mode to transfer the data in every 2 minutes, and free ThingSpeakTM[129] *Internet of Things* service was used for the purpose. The SBC of the monitor was installed with a Debian Linux operating system and it was programmed with Python based script to control the data collection from DylosTM monitor and to communicate with central base station/server by means of SMS or REST API. The central base station was an Ubuntu Linux head-less server running on IBM M4 machine equipped with a data-card to receive the SMS and REST-API.

3.2.3 Data conversion and *collocation* sampling

The particulate pollution levels are usually represented in mass based values as micro gram per meter cube, while the DylosTM monitor gives the particulates in counts per 0.01 standard cubic feet. To convert the DylosTM reading into mass values for the purpose of comparison,

a relationship was used based on Arling et al. [130] and a python script ¹ was used for this purpose.

$$PM\ Concentration(\mu g/m^3) = Number\ of\ Particles * 3531.5 * Particle\ Mass \quad (3.1)$$

The value 3531.5 in equation was used to convert the sample orifice size of DylosTM monitor 0.01 ft³ to 1 m³. The Particle mass for the $PM_{2.5}$ channel and PM_{10} channel were taken as $5.89 \times 10^{-7} \mu m$ and $1.21 \times 10^{-4} \mu m$ respectively. These values were derived based on approximations such as considering the shape of the major particles in sampling environment to be spherical and has a density of $1.65 \times 10^{-12} \mu g/m^3$ [96]. Similarly, the radii of particles in the $PM_{2.5}$ and PM_{10} channels were considered as 0.44 μm and 2.60 μm [132] respectively. Higher Relative Humidity (RH) in environment is known to force the free particulates in air to clump together and that can [130] wrongly influence the counts by laser based particulate matter monitors. To reduce such error readings under high humidity conditions, those readings either were discarded or adjusted using a correction factor. To assess the validity of data from the monitor, a collocation sampling (During 14 - 19 January 2015) was carried out concurrent with industrially calibrated portable particulate monitor MetoneTM Aerocet -531S. The monitor MetoneTM Aerocet- 531S was placed along with DylosTM in the same enclosure and readings were compared. Pair wise plots and linear regression adjusted R^2 value of hourly average samples were used to assess the relationship / collinearity between the readings from the collocated monitors.

3.3 Results and Discussion

The Figure 3.1 displays the monitor that was developed complete with its all weather casing for field deployment. Each unit cost around 48,700 INR, Table 3.1 lists the breakups (item costs) relating to the components used in equipping the monitor. It is observed that in terms of cost for customization, DylosTM monitor stands for around 70 % of the cost for

¹File named *DylosDataCountToMass.py* in Nishadh et al. [131]



(a)



(b)

Fig. 3.1 Real time particulate pollution monitor (a) with casing (b) for field deployment

customizing the real time particulate pollution monitor. The Table 3.2 lists the price ranges of commercially available particulate pollution monitors some of which can be customized into real time monitors. Studies have proved the effectiveness of lowest price range Shiney dust monitors for particulate pollution monitoring [93, 127]. Indicating the overall cost of real time monitor can be reduced considerably by appropriately changing the monitors engaged, enhancing its usability for mobile monitoring purposes [133] or augmenting sensing capability for particulate pollutant with indirect image based methods [134–136].

The Figure 3.2 shows software and hardware components of the real time particulate matter monitor. The software component consist of serial data reading from Dylos™monitor, a database to store the data records in SBC local storage and communicate on real-time basis either by SMS or REST API. Python based scripts were used to carry out SMS ² and REST API³ based data communication. The functions of these Python scripts are as follows.

²File named *DylosDataRecvSMS.py* in Nishadh et al. [131]

³File named *DylosDataRecvHttp.py* in Nishadh et al. [131]

Table 3.1 Cost of the materials used in real time particulate pollution monitor

S.No	particulars	Cost in INR	Source
1	Dylos™Monitor (kg)	34000	Dylos corporation,USA
2	Inverter and Battery	8000	Local purchase
3	Raspberry pi	3300	Crazy pi, Bangalore
4	All weather casing	1200	Local control box manufacturer
5	USB Data card	1200	Local electrical shop
6	Serial to USB converter	600	Local computer shop
7	Electrical accessories	400	Local electrical shop
	Total	48700	

Table 3.2 Price range of some particulate matter monitors

Hand-held/Indoor monitor	Vendor	Price range in ₹
Shineyi/Sharp dust sensor	Shineyi/Sharp	<2000
DC1100 Air Quality Monitor	Dylos Corporation	<34,000
Aerosol 531S	Metone	<3,00,000
DataRAM pDR-1500	Thermo Scientific	<7,00,000
MicroAeth AE51	AethLabs	<8,00,000

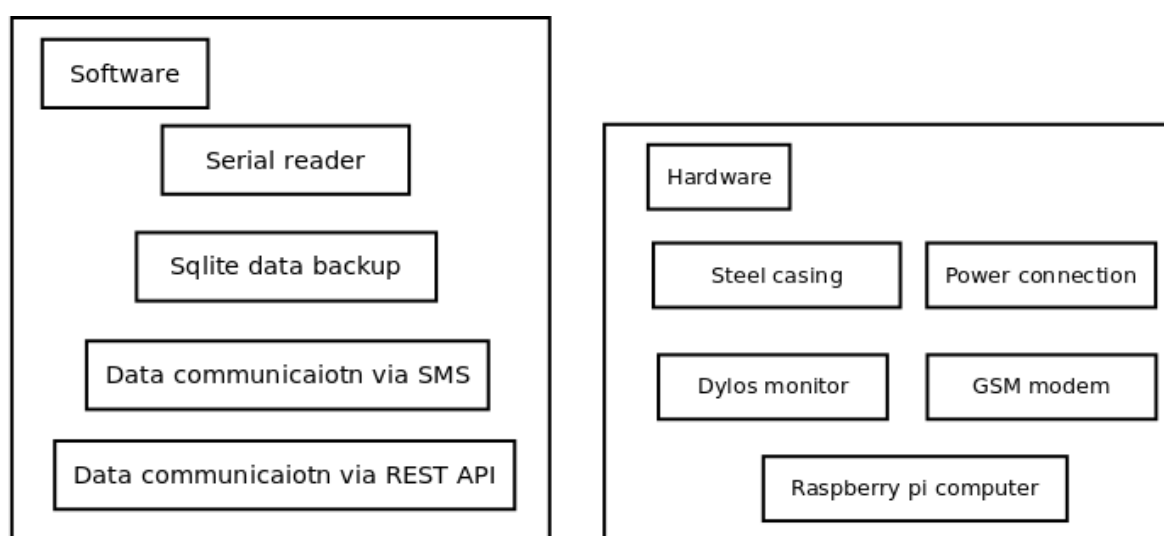


Fig. 3.2 Software and hardware components of real time particulate pollution monitor

1. Invoke the serial read in SBC using the serial library and set the time interval for data reading from Dylos™.
2. Store the data in SQLite database located in SBC, and act as data backup
3. Query the SQLite database, collect the most recent data and communicate to REST API in real time
4. Send the serial read data as SMS in real time

3.3.1 Ambient testing

The Figure 3.3 shows the location map of the field stations where the customized monitor was stationed for testing. The Table 3.3 gives details about those locations. Longest field-testing was carried out in Thadagam, hereafter referred as TDM. It is a village near to Coimbatore corporation limit and home to largest brick kiln cluster in Coimbatore. Due to its remote location, mobile internet connectivity was low in this area. In this situation, SMS based data transfer was only viable form for real time data communication and was followed for this station. Kuniyathur (KNMR) is in the southern part of Coimbatore Corporation, and the place is predominantly residential area bordered with agriculture fields crisscrossed with approach roads leading to a busy national high way. Ethernet connected REST- API was used for real time data communication in this station. Sivanandha Colony (SVNC) is a commercial and residential area in Coimbatore city. Here, the monitor was kept in an office building close to another busy road. REST-API based data communication was used in this station. GV residency (GVR) is a residential area located in the eastern part of Coimbatore Corporation. REST-API based data communication was intended to be used in this station utilizing the available Ethernet based local area network; however, since the monitor became dysfunctional within a day of deployment due electricity problem and failed in collecting the data from Dylos™. Hence, the data from this monitor was not used in further analysis.

Data collected from the field evaluation was assessed for general trend and reflection of the station's pollution profile. Due to logistic constraints, the monitors functioned in different

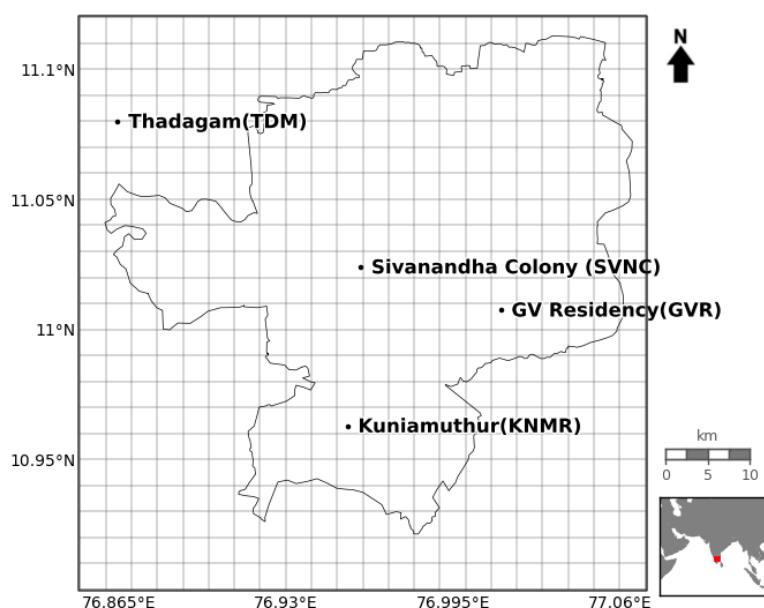


Fig. 3.3 Real time particulate pollution monitor field deployment stations in Coimbatore

Table 3.3 The profile and period of sampling for Dyls™ air quality monitor in different stations of Coimbatore

Station code	Station name	Monitor position height	Period of sampling
TDM	Thadagam	7 meter	24th January -28th March 2014
KNMR	Kuniamuthur	3 meter	6th-8th May & 14th-20th July 2014
SVNC	Sivanandha Colony	1.5 meter	17th April -05 May 2014
GVR	GV residency	7 meter	29-30 April 2014

time periods that hinders the across-site comparison of the pollution levels. However, for each monitor hourly trends were assessed. The average $PM_{2.5}$ concentration recorded by monitor placed in TDM was $72.27 \pm 14.33 \mu\text{g}/\text{m}^3$. At SVNC, the concentration was $30.34 \pm 10.65 \mu\text{g}/\text{m}^3$, and at KNMR, $24.76 \pm 12.37 \mu\text{g}/\text{m}^3$. In the case of PM_{10} concentration, SVNC recorded $414.02 \pm 205.17 \mu\text{g}/\text{m}^3$, TDM $313.72 \pm 148.52 \mu\text{g}/\text{m}^3$ and KNMR $221.31 \pm 92.05 \mu\text{g}/\text{m}^3$. It is observed that hourly $PM_{2.5}$ average value was gradually increasing towards evening except a single morning hour steepness in TDM monitor reading (Figure 3.4). The SVNC monitor shows gradual increase of $PM_{2.5}$ in morning hours and similar but lower value gradual increase in evening hours. In the case of KNMR, the monitor records the $PM_{2.5}$ concentration gradually increasing towards the evening time. The PM_{10} , in terms of (Figure 3.5) hourly concentrations,

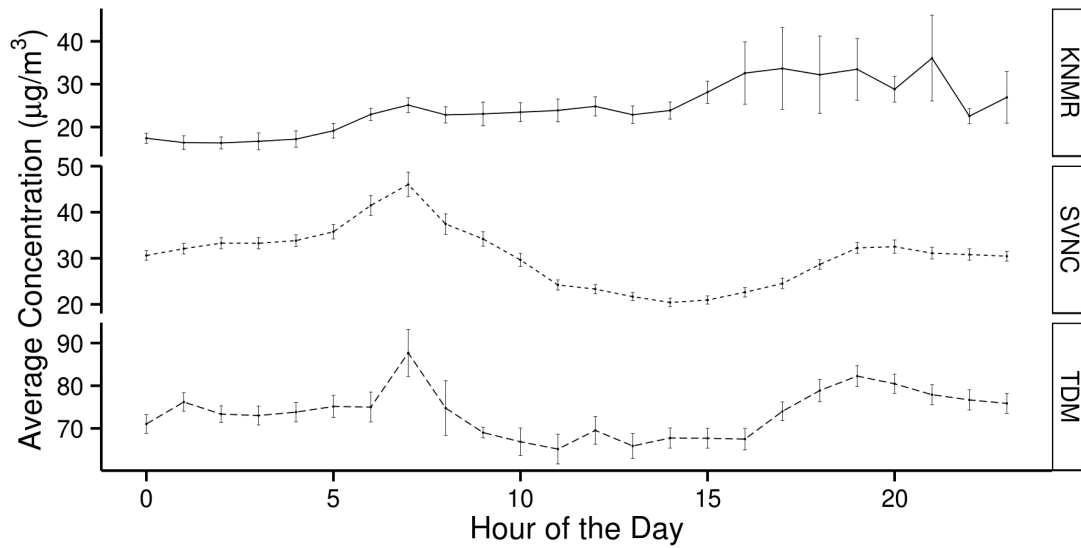


Fig. 3.4 Hourly $PM_{2.5}$ concentrations in the stations where the monitors were deployed, KNMR-Kuniamuthur , SVNC-Sivanandha Colony, TDM- Thadagam

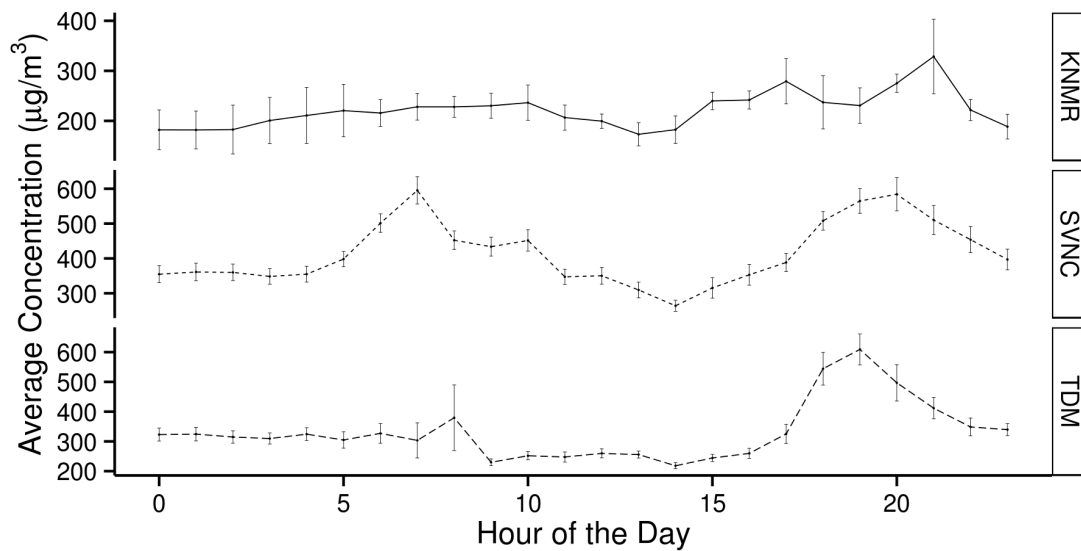


Fig. 3.5 Hourly PM_{10} concentrations in the stations where the monitors were deployed, KNMR-Kuniamuthur , SVNC-Sivanandha Colony, TDM- Thadagam

recorded by the monitors show a trend similar to that of $PM_{2.5}$. However, more gradual peaking during evening and morning time at SVNC, which may be indicative of vehicular traffic peaking in the nearby busy road. In the case of TDM and KNMR, the hourly trend may be indicative of diurnal activity related to brick kiln industry operation and road traffic respectively. The

data from monitors developed using low cost sensors are viewed as useful for information purpose and awareness creation, hotspot identification, supplementary network monitoring and personal exposure assessment [137]. In the current study, it is observed that the monitor reading can satisfactorily exhibit pollution trend necessary for information purpose. But, inability to conduct monitoring in all the stations with similar instruments simultaneously for logistic reasons disallowed the use of the data for hot spot identification and supplementary network monitoring.

As an individual node in wireless sensor network, the system effectiveness of real time particulate monitor was assessed in terms of system persistence and data communication. It is observed that the monitor was effective in continuous operation for data collection, storage and real time communication and long monitor operation. At TDM and SVNC, the monitor was in operation for 1490 and 200 hours respectively. The intermittent disruption in the monitors was mainly due to power supply failure, and resultant erroneous data recording by the DylosTM monitor and data communication. All these situations warranted operator intervention to revive the monitor. The current design has limitation such as using loosely coupled plugs and switches to power the components, connected with direct AC power, and backed with an Inverter. Providing power in this form by non sine wave based inverter Uninterrupted Power Supply (UPS) leads to disruption of the DylosTM monitor reading, at times leading to unrealistic high value recordings for particle counts. However, this error could be rectified using sine wave based UPS. Further, the collected data was scrutinized for the realistic nature of the readings and high unrealistic readings were filtered out. That indicates the need for improvement of design of the monitor with self-contained DC battery for power source and backup with embedded connectivity among the components. There are situations observed that the serial dysfunctional and delay in capturing the serial read by high level USB based cord used with SBC. It further hinders extension of the low-level sensors by addition of sensors for parameters such as Temperature and Humidity to the monitors. Probably, in such cases using low-level microcontrollers would be advisable. In spite of such possible drawback that could be fixed by uncomplicated design improvements, the monitors' functionality was in par with the expectation of real time particulate pollution monitors and wireless sensor nodes.

3.3.2 Data validity

Air pollution instruments generally are calibrated and data validated through either of two methods [137]. The first method is comparison of the reading from the instrument with a reference standard air pollutant concentration in a controlled environment or laboratory [138]. The second method compares the reading with a reference instrument that has been calibrated with the first method [139] and is known as *collocation* sampling. This is generally preferred for particulate matter monitor calibration due to its cost effectiveness. In the current study an aerosol calibrated Metone™Aerocet -531S monitor with accuracy level of ± 10 was used as the reference instrument. Considering the high humidity ($>80\%$) in late nights in Coimbatore, the study area, and safety of the instrument, *collocation* sampling was interrupted from 9PM to 7AM. The Figure 3.6 and 3.7 shows linear regressions conducted between the readings obtained by the two instruments for $PM_{2.5}$ and PM_{10} readings respectively.

The coefficients of determination (adjusted R^2) for separate sampling periods were ranging between 0.54 to 0.98 for $PM_{2.5}$ and 0.65 to 0.92 for PM_{10} . For the entire sampling period, the adjusted R^2 for $PM_{2.5}$ was 0.75 and for PM_{10} it was 0.78. It is observed that the coefficient of determination R^2 between the different particle size counts and non-averaged counts (see appendix A) varied widely. The highest R^2 was observed in the case of hourly mass concentration; in contrast, the R^2 was lower for counts. The Figure 3.8 and Figure 3.9 shows combined line plot of hourly averaged data from the Metone™Aerocet -531S and the collocated real time monitor. Concentration for $PM_{2.5}$, as per the real time monitor, tracks closely with Metone™Aerocet -531S, better than with the concentration of PM_{10} in most of times. It was observed that the PM_{10} concentration value from real time monitor was consistently higher than that from Aerocet- 531S monitor indicating the need for a correction factor for the readings obtained from the former.

Earlier study by Northcross et.al [126] carried out using PM_{10} based monitor has shown a range of R^2 values from 0.81 to 0.99 for $PM_{2.5}$. The collocated reference instruments were Dust-trak, an industrially calibrated particulate profiler and beta attenuation monitor (BAM or β -gauge). After calibrating Dust-trak with a 24-hour average value obtained from BAM, hourly

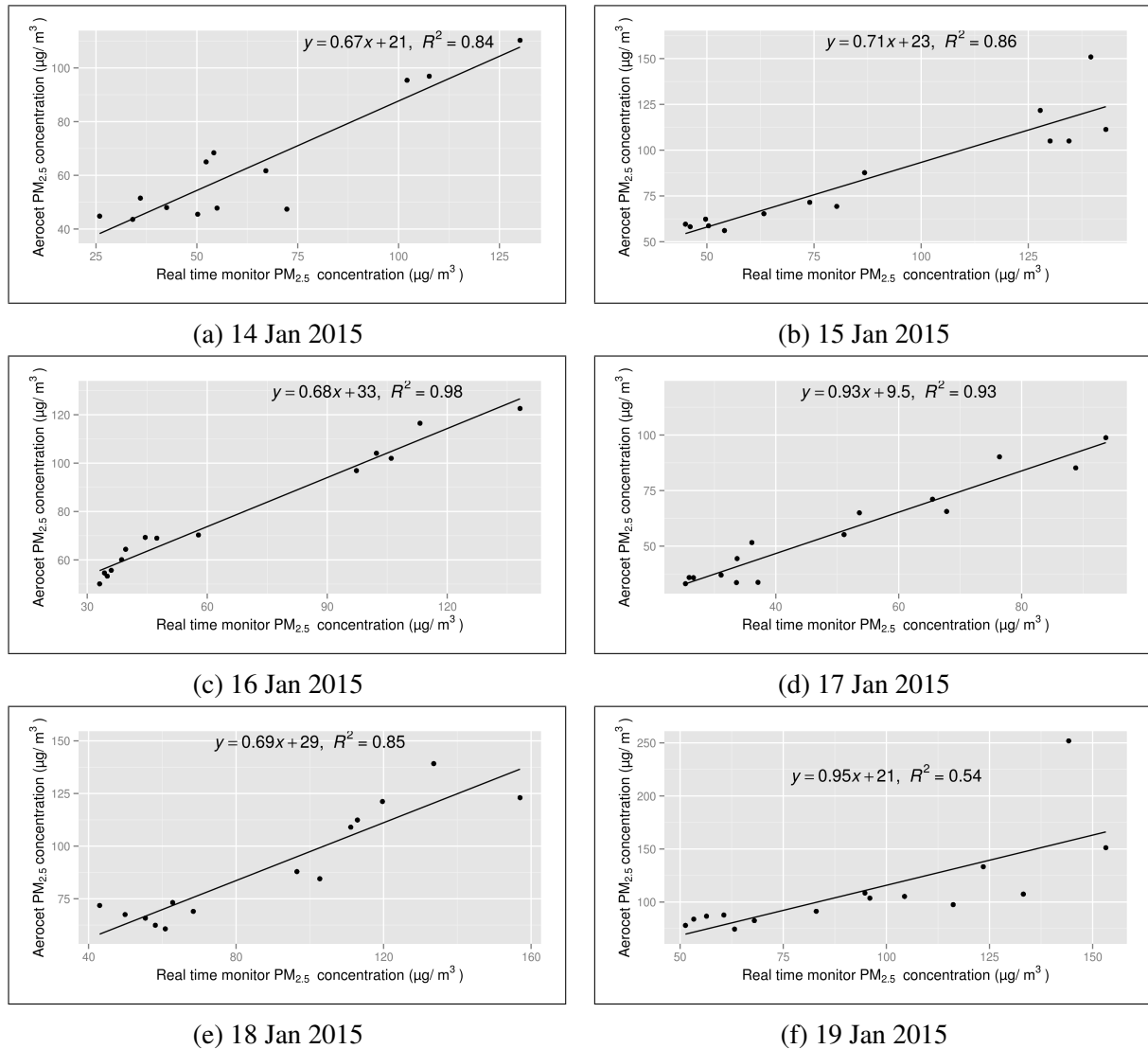


Fig. 3.6 Linear regression (LR) between real time particulate pollution monitor and Aerocet 531S in different periods for $PM_{2.5}$

average of Dust-trak readings was used for analysis. Light scattering based monitors readings are subject to varying composition of ambient aerosol, temperature and relative humidity [93]. In the current study, high variability (R^2) in agreement between the collocated instruments across different periods is probably due that phenomenon of variability in such parameters relating to ambient air quality. That would essentially means, the requirement for continuous calibration of low cost particulate matter monitors to have valid data [139]. The close tracking of mass concentration, which are a derived value from count reading, between the collocated

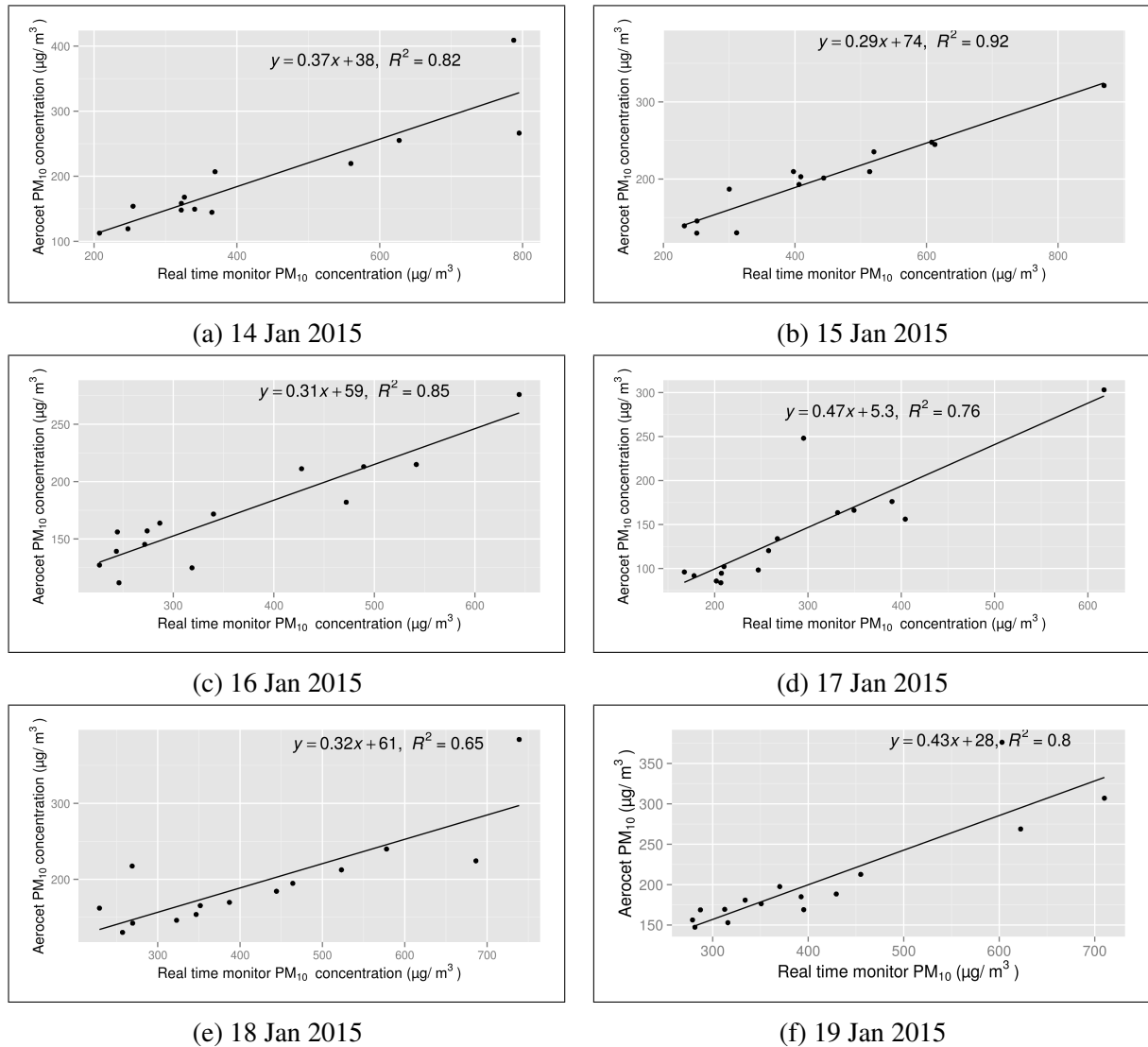


Fig. 3.7 Linear regression (LR) between real time particulate pollution monitor and Aerocet 531S in different periods for PM_{10}

instruments indicate the effectiveness of conversion using the relationship (see the formula 3.1), much simpler than that used in previous study [126]. Overall, the current study reiterates the usability of low cost monitors for particulate pollution monitoring purposes. The study also emphasizes the necessity of calibration and using it for augmentative purposes in the backdrop of regulatory compliant monitoring network.

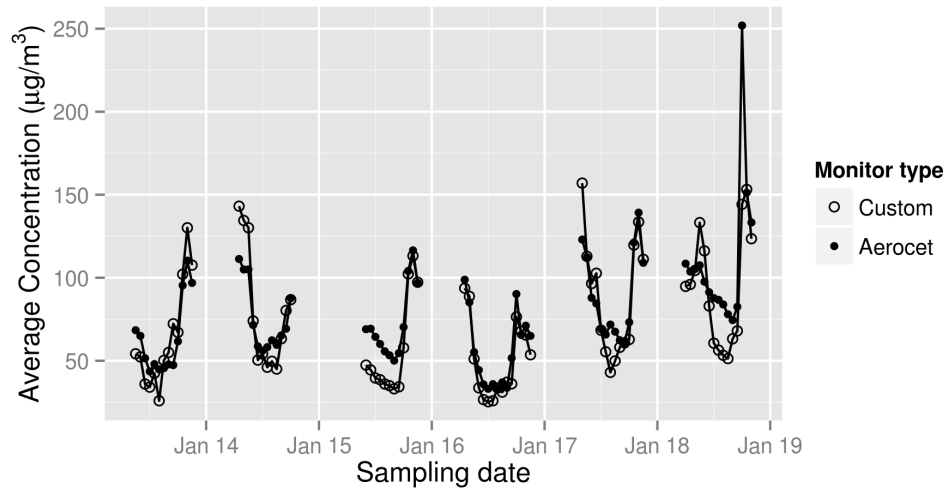


Fig. 3.8 Hourly $PM_{2.5}$ concentration in custom real time monitor and Aerocet 531S during collocation sampling

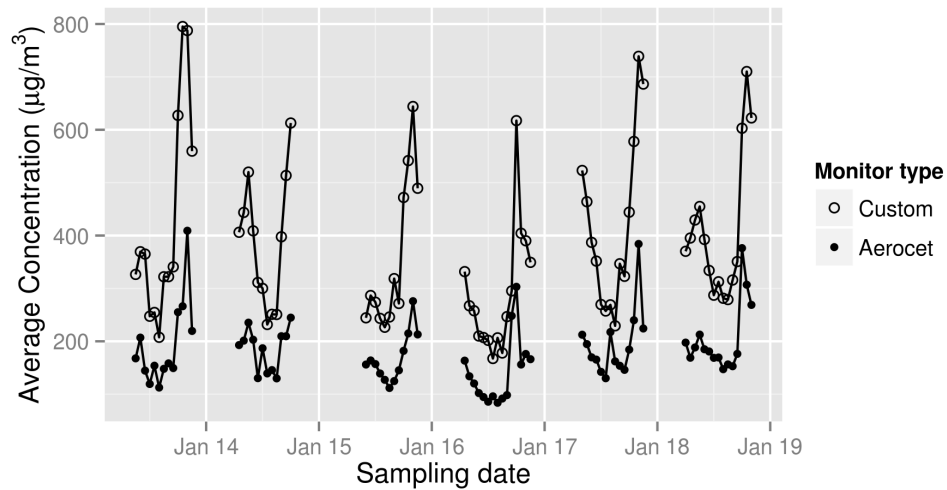


Fig. 3.9 Hourly PM_{10} concentration in custom real time monitor and Aerocet 531S during collocation sampling

3.4 Conclusion

1. This chapter discuss customization of a low cost indoor dust monitor into an outdoor real time particulate pollution monitor. Assessment of its functionality for ambient monitoring and its data validity was performed

2. Ambient testing of the developed monitor shows its effective working in field condition. The continuous operation of the monitor and real-time data communication were tested and found to be satisfactory
3. The system effectiveness of real-time particulate monitor was assessed in terms of persistence and data communication. It is observed that the monitor was effective in continuous operation for data collection, storage, real time communication and long operation
4. The customized monitor development using loosely coupled power source and single board computer were found to require intermittent manual interventions. That can be overcome by a design overhaul replacing with embedded boards, stable and permanent connectivity with battery power source
5. Data validation study shows that the developed monitor is correlating well with the industrially calibrated monitor. But the variation in the agreement with different calibration period indicates the requirement of continuous calibration of the monitor

Chapter 4

Particulate pollution emission inventory

4.1 Introduction

Emission inventory (EI) is breakup of air pollutant quantity emitted from anthropogenic or natural origin; it quantifies various pollutants generated from emission causing activities within a stipulated geographical area [140] and time span. Emissions of Particulates in an urban area would be mostly of anthropogenic origin; of course, natural forces and settings also would determine the distribution of emissions, but that would be of much wider geographical and temporal scales. Major anthropogenic urban emission source for particulate pollutants are industrial, transport, agriculture, residential, and construction sectors.

Emission inventory development is an important and primary task in air pollution management [141, 142]. It is a key input data for air quality forecasting and in planning management interventions [143, 144]. It gives comprehensive understanding on air pollution issues relevant to a geographical and time span context. Preparation of such a document mainly involve four facets; source identification, quantity of pollutant generated from it, pollution reduction potential and identification of information gap to carry out emission reduction measures effectively [145]. It helps in future projection of pollutant concentration in the context of possible scale-ups in emission activity such as demographic changes or increase in industrial activity and such other changes. It aids in designing and implementing various intervention

measures, with respect to present and future emission rates. Air quality models especially chemical transport and numerical weather prediction-based models require spatially allocated and gridded emission inventory [146] that provides initial condition values (seed values) for mathematical computation of spatio-temporal air pollution forecasts [147]. The quality of Simulation is largely influenced by the emission inventory resolution, completeness and accuracy [148]. Various studies have observed that emission inventory is the most uncertain input data in model simulation [149–151]. Accurate emission inventory is a great deal crucial for its important for functionality and advantages derivable from modeling. Limitations in it can significantly misguide or make non-effective the air pollution interventions based on it [149, 152, 143, 153, 154].

Emission inventory is scaled as per the extent of area considered for estimations, such as urban, regional, country or global level domains. The selected domain of spread is further divided into fine spatial grids and then emissions are estimated pertaining to each individual grid [155]. Sizes of the spatial grids are chosen for micro level computing and that acts as the resolution of the emission inventory. In each grid, sources of the pollutant are geographically represented as point, line or polygon features [156] based on the geographical extent or nature of the source. Point features are normally related to a locality (e.g. industrial), polygon features related to a group of similar kind of industrial or residential area and line features are related to features such as road networks.

EI is estimated taking a base year according to data availability [157]. The EI calculation requires data on quantity of fuel type used and its respective emission factor [158]. The basic data requirement for EI assessment are i) location of emission sources, ii) fuel consumption rates for the sources that are considered as pollutant emission activity, and iii) emission factors that are mostly laboratory estimated value of pollutants emitted from unit consumption of fuel by the pollutant emission causing activity. Based on the availability and specificity of data, two approaches, top-down or bottom-up, are adopted in EI preparation [157]. Top-down approach takes overall fuel consumption as surrogate for emission. Then the area level emissions are derived using related proxies for the emission activity and respective spatial segregates [158].

On the other hand, bottom- up approach considers every pollutant emitting sectors in an area based on their respective spatial extent. Overall emission rate from these sectors were estimated discretely and then they are combined together in the respective spatial grids where the emission source is located [159].

Due to non-availability of data required for bottom-up approach, Global level emission inventories for air quality models are largely developed using top-down approach and in coarse resolution grids [160–162]. These emission inventories though data ready for air quality models such as Weather Research and Forecast - Chemistry (WRF-CHEM), suffers from inadequacy due to the coarse geographical resolution (100 km-10 km) for its use in local scales (local level simulation). Hence, bottom-up is more appropriate for usage in air quality models for better realistic predictions in local level simulations. In India, several studies have looked into development of gridded emission inventory for large urban centres such as New Delhi for air quality modeling [159, 148, 163]. In view of the above, the current study attempted to develop particulate pollution emission inventory for Coimbatore region through a bottom-up approach. The study also used free and open source programming tools.

4.2 Methodology

4.2.1 Study area:

As per Census of India 2011, Coimbatore district has a population of 34.85 lakh and has a decadal (2001 to 2011) population growth of 19.06 % [164]. After Chennai, it is the largest industrial hub in Tamil Nadu mainly involved in Textile, Pump / Motor manufacturing, metal casting and several small scale engineering industries. An area of 3000 km^2 (with extent of area between 10°48'25"N, 76°43'21"E South West and 11°18'27"N, 77°19'21"E North East) surrounding Coimbatore Corporation and covering the nearby villages were included while preparing current emission inventory with a resolution of 1x1 km (Figure 4.1). The emission inventory was prepared taking 2012 as the base year.

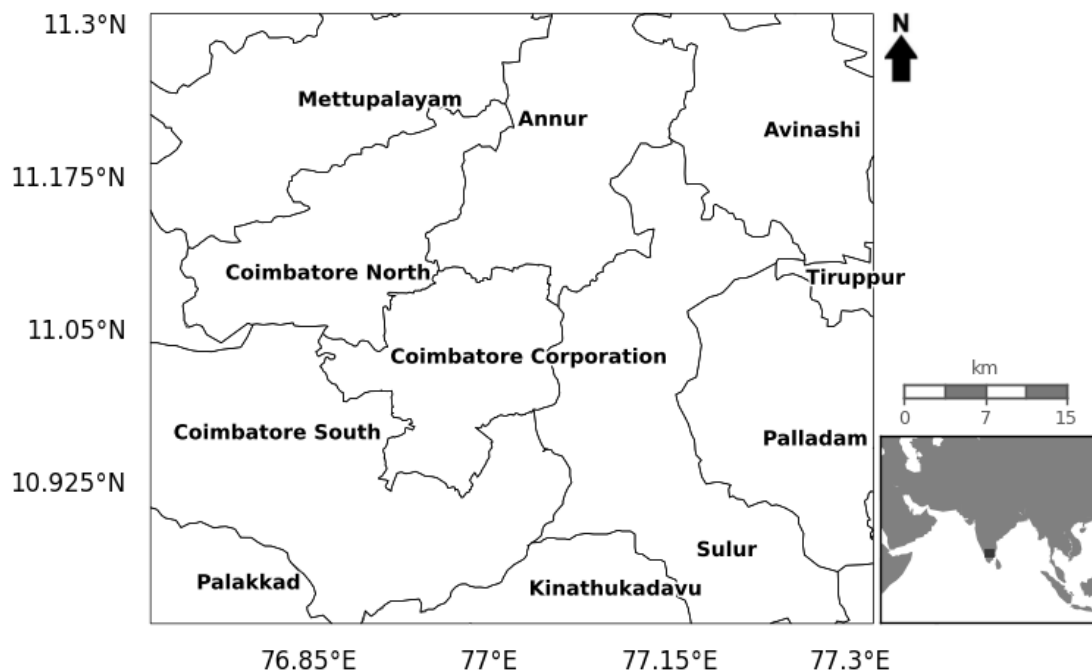


Fig. 4.1 The study area map with extent of area considered for emission inventory calculations(50x60km)

4.2.2 Emission source and activity data

Four major emission sources viz., sectors such as residential, industrial, transport and road-side windblown dust were considered for particulate pollution emission inventory. Relevant data from Census of India 2011, National Sample Survey Organization (NSSO), Statistical Handbook of Tamil Nadu, Tamil Nadu State Transport Corporation (TNSTC), Regional Transport Office (RTO), Tamil Nadu Pollution Control Board (TNPCB), and open street map shape files were used.

Residential sector Census of India 2011 data primarily on total population enumeration, house listing, and housing census were used [164] for the inventory. The population enumeration table enlists the number of houses at ward or village level. The house listing and housing census data table gives the number of persons per house, percentage of houses using different fuel types such as LPG, Kerosene and firewood for cooking activities at ward or village level. These two data tables were merged on ward or village level unique code and its respective

Table 4.1 Per capita consumption of fuel types in Tamil Nadu during 2012-2013 [165]

S.No	Fuel type	Rural	Urban
1	Firewood and chips (kg)	19.271	3.686
2	Kerosene (liter)	0.613	0.551
3	Electricity Kwh	17.016	36.426
4	LPG (kg)	0.948	2.17
5	Coal/charcoal (kg)	0.002	0.0

map on polygon boundary to get the source location. The National Sample Survey [165] data gives state level per capita consumption of each fuel type (Table 4.1). This table merged with house listing and housing census data gives per capita fuel consumption as emission activity in each ward or village. The 1x1 km emission inventory grid superimposed on ward/village map was selected based on the built-up area in each of it. Google™web map services with its high resolution satellite imagery were used for selecting the grid. Only these selected grids were subject to emission inventory calculation.

Transport sector The spatial spread of the road networks was considered as a source of emission in the emission inventory. For the purpose, the shape file data from Open street map containing line features (road network) were used. The Figure 4.2 shows road network within the study area categorized into different classes based on the level of importance i.e. intensity of road traffic. The vehicular emission activity was computed based on the data source from Statistical handbook of Tamil Nadu 2014 [166], which tabulates the district level vehicular population. Vehicular Kilometre Travel (VKT) for typical urban India was taken from Sahu et al [159]. The Table 4.2 summarizes these data. To derive vehicular emission source from the spatial spread of road network, road categories such as trunk, primary, secondary, tertiary and other roads was assigned with a proportion of vehicles passing through in a typical day [167].

Diesel buses are major means of public transportation in Coimbatore. It has a fleet of 3267 and 1724 unique routes catering the needs of Coimbatore and nearby districts of Nilgiris and Erode [166]. Large proportion of it is controlled by government owned TNSTC with a smaller proportion run by private operators. The emission activity from these buses were collected from

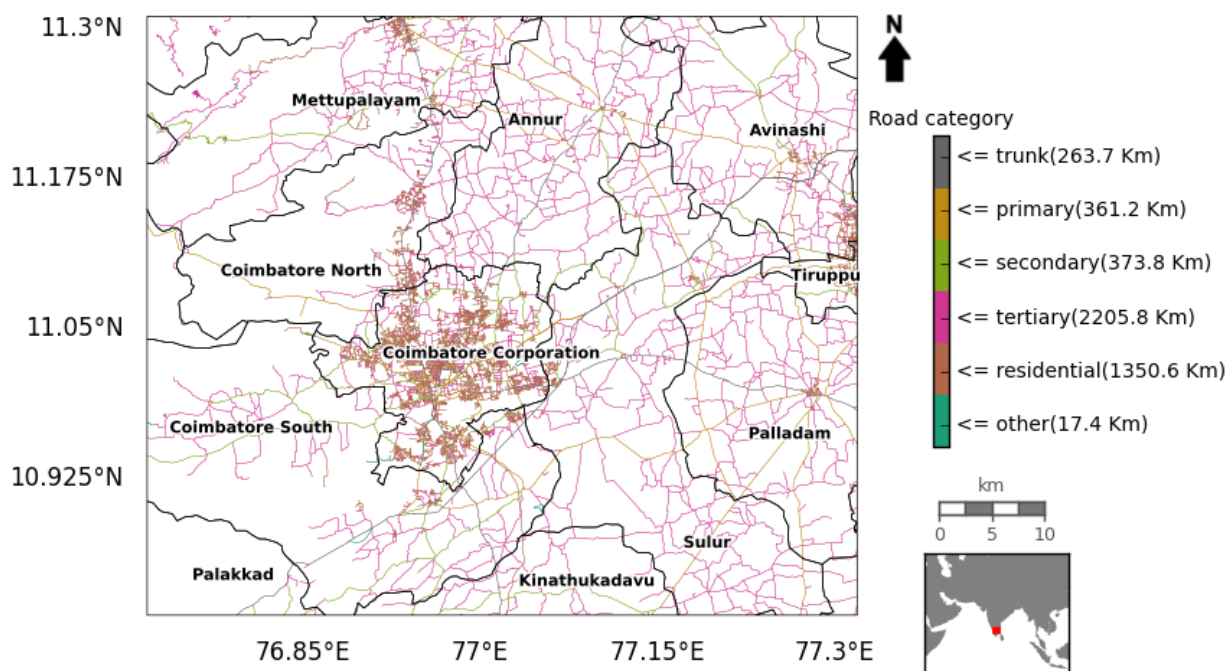


Fig. 4.2 Road network in Coimbatore region

Table 4.2 Vehicular population in Coimbatore district during 2012-2013

Type	No. of Vehicles[166]	PM EF(g/km)[108]	VKT [159]
Two wheelers	1423384	0.049	75
Four wheelers	244661	0.19	90
Goods vehicle	105414	1.965	150
Authorickshaw	11041	0.347	150
Bus	3766	1.075	210

daily trip sheet obtained from TNSTC, RTO and a travel suggestion website of Coimbatore city [168]. The VKT of each bus were also calculated from this. The emission sources from these buses were generated by converting the daily trip sheet of bus routes and bus-stop details into the line feature, road network.

Windblown dust The emission from road related dust was calculated by taking the road network as its source. The emission activity was computed by considering factors such as number of wet days in a year, proportion of paved or unpaved roads in the network, average silt loading and average weight of the vehicle run over the road. In 2012, Coimbatore urban region

Table 4.3 Fuel consumption in Coimbatore(kL/year) [169] during 2007-2011

Year	Petrol		Diesel		
	Transport	Industries	Transport	Commercial	Industries
2007-2008	61374	15344	117058.9	16722.7	33445.4
2008-2009	68558	17139	137474.4	19639.2	39278.4
2009-2010	58402	14601	105676.2	15096.6	30193.2
2010-2011	62802	15700	121689.4	17384.2	34768.4

experienced 28 wet days, 100 % of the road network is unpaved, average silt load 0.531 g/m^2 and average weight of the silt is 1.41 tons per year as per a surrogate urban Indian condition (an estimated values in Delhi conditions [159])

Industrial sector Industries such as Cement, foundry, brick kilns and stone crushers are major air polluting sectors in Coimbatore. For emission source data on industries, TNPCB, Coimbatore division was consulted for pollution category of the industry, address and annual production. The study area of the emission inventory has two mining and cement manufacturing industries in it. As per the TNPCB list and industry declaration, the annual production of these two industries range from 438000-1200000 Ton/year. The region has 722 foundry units functioning based on coal and furnace oil as their primary source of energy. The geo-coded industrial emission activity was estimated from the annual production and fuel required per ton of production. The emission from usage of electricity generators were also considered for particulate emission. The spatial sources of the emissions were geo-coded based on their location taken from the address details of educational institutions, hospitals, and commercial complexes that use diesel based generator to cope with long power outages. The emission activities were computed from the average daily electricity power outage and diesel consumption per hour of usage. Consumption of various fuel types by different sectors in Coimbatore region is provided in the Table 4.3.

Table 4.4 Emission factor for each sector and fuel type[159]

Sector	Pollutant	Unit	CNG/NG	LPG	Wood	Diesel	Coal /Coke	Kerosene	FO	Cow dung
Industries	PM_{10}	g unit ⁻¹	0	-	-	0.0009	-	-	-	-
	$PM_{2.5}$	g kg ⁻¹	0.34	0.31	-	0.97	1.36	0.34	0.65	-
Residential	PM_{10}	g kg ⁻¹	-	2.1	15.3	-	20.0	1.95d	-	-
	$PM_{2.5}$	g kg ⁻¹	-	0.33	1.5	-	12.2	1.9d	-	5.04

4.2.3 Emission factors, emission calculation and spatial allocation

Emission Factor (EF) constitutes an important part in emission inventory preparation and greatly influences its accuracy. In the present study, EF from the past studies were used [108, 170] for computational purposes. The values reproduced in Table 4.4 lists the emission factors used in the current study.

Following were the formulas used for particulate emission inventory calculation. The vehicular emission is calculated by formula 4.1.

$$E_t = \sum (Veh_l * D_l) * EF_{l,km} \quad (4.1)$$

where E_t = total emission of the compounds; Veh_l = number of vehicles of each type; D_l = distance travelled in a year per different vehicle type, derived from VKT per day; $EF_{l,km}$ = emission of compound, vehicle type per driven Kilometre.

The total emission from vehicular population is spatially allocated to each grid based on assumed volume proportion passing over roads of different category at a given time [167]. The equation 4.2 was used to allocate the total emission from vehicular population to each grid.

$$E_i = \left(\frac{LT_i}{\sum LT_i} * 0.38 + \frac{LP_i}{\sum LP_i} * 0.28 + \frac{LS_i}{\sum LS_i} * 0.15 + \frac{LT_i}{\sum LT_i} * 0.13 + \frac{LR_i}{\sum LR_i} * 0.06 \right) * C \quad (4.2)$$

Where, E_i = total emission in i^{th} grid, LT_i = total length of the trunk roads in the i^{th} grid, LP_i = total length of the primary roads in the i^{th} grid, LS_i = total length of the secondary roads in the i^{th} grid, LT_i = total length of the tertiary roads in the i^{th} grid, LR_i = total length of the other roads in the i^{th} grid, and C = total vehicular emission as per equation 4.1.

Emissions from unpaved roads are calculated using the formula 4.3.

$$E = [(K(S/12)^{0.8}(W/3)^{0.4})/(M/0.2)^{0.3}] * (N - P)/N \quad (4.3)$$

Where E = particulate EF ($g \text{ VKT}^{-1}$), S = surface material silt content, P = number of *wet* days with at least 0.254 mm (0.01 in) of precipitation for the averaging period (365 days), K = particle size multiplier for particle size range ($g \text{ VKT}^{-1}$), W = mean vehicle weight (tons), M = surface material moisture content (%).

The total emission is calculated using the formula based on works such as [159, 170–172]. Total emission from all the sectors is calculated as follows,

$$TE = \sum_a \sum_b FU_{a,b} \left[\sum_c EF_{a,b,c} U_{a,b,c} \right] \quad (4.4)$$

where a = sectors for which EI was prepared, b = fuel type, and c = technology, TE is total emission, FU is fuel utilization, EF is emission factor for each fuel type, U is the fraction of

fuel under a sector (transport, road, residential etc) with specific re-mediation technology in place to contain pollution emission. As there is no emission remediation technology considered in the present situation the respective factor for this aspect is taken as one.

Spatial allocation of emission inventory was carried out using free and open source geospatial tools of the *Python* programming language and its libraries such as *Numpy*[173], *Pandas*[174], *Fiona*[175], *Geopandas*[176], *Shapely*[177] and *Rtree* [178] index.

4.3 Results

4.3.1 Programming tools

The domain, which was gridded with 1x1 km grids, in total comprising of 3000 polygons in ESRI shape file format, was used for computation of emission inventory for all the sectors. The steps involved in EI computation are shown in the Figure 4.3, taking residential sector as an example. *Python* based programming codes⁴ were used for the computation. The code does four sequential steps, given below, in calculating the emission taking residential sector for demonstration.

1. Join Census of India data tables (such as population enumeration data, house listing) with village/ward boundary polygon feature in ESRI shape file.
2. Find the number of houses in each emission inventory grid (EI grid) from Census of India 2011 data with village/ward boundary details, this step involves sub steps below.
 - (a) Select out of 3000 odd EI grids those grids having built-up area within. This is carried out by performing a *within intersection* between the EI grid centroid and built-up area land cover polygon
 - (b) Find the EI grid centroid intersecting within village/ward polygon, and attach the village/ward polygon unique id to EI grids

⁴File named *EmissInvResidentailCalc.py* in Nishadh et al. [131]

- (c) Count the number of EI grid centroids intersecting within each ward/village polygon and get the proportionate number of houses in each grid from the total number of houses in each village/ward.
3. Join the National Sample Survey data and proportionate number of houses in each village using different type of fuel for their day-to-day needs, which is the data on residential sector emission activity.
 4. Join emission factor and execute the emission inventory computation.

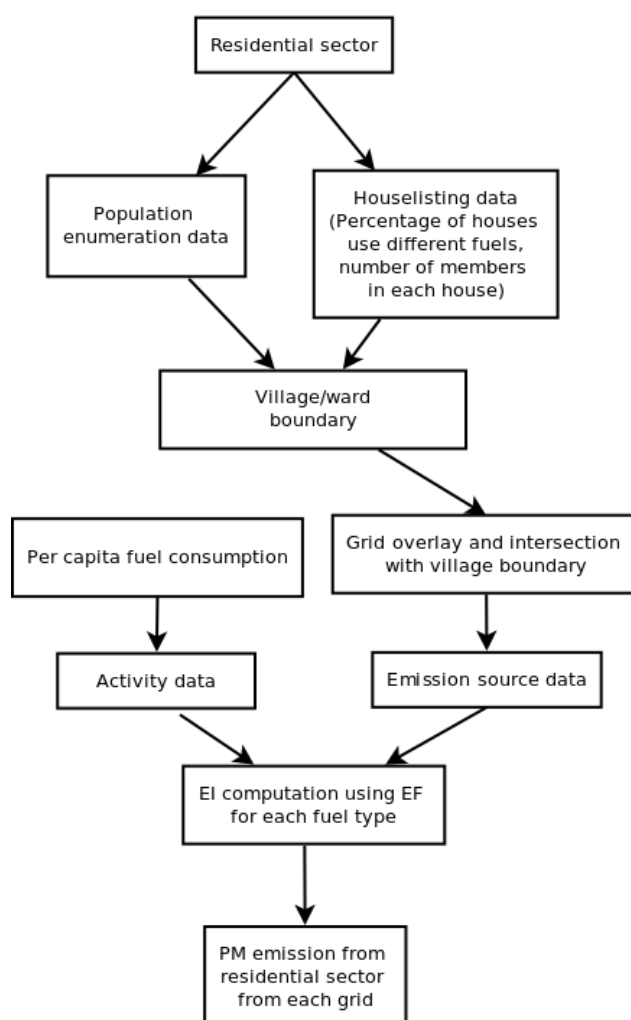


Fig. 4.3 Steps involved in residential sector emission inventory estimation

Python programming language library *Pandas* was used for joining the Excel™format tables. The joining is done using the unique attribute number given to each village or ward, followed by intersection within operation for grid centroid with two set of polygons. It is carried out by *Shapely* library function. First set of polygons is relating to the urban land cover area in the domain and the second polygon is Census of India 2011 village boundary polygons. This is carried out for considering only the built-up area and for avoiding the grids with completely barren land, for residential emission inventory compilation. The Figure 4.4 depicts the steps involved in estimating the inventory for transport sector emission. *Python* code⁵ implements five sequential steps in calculating the transport sector EI, as given below.

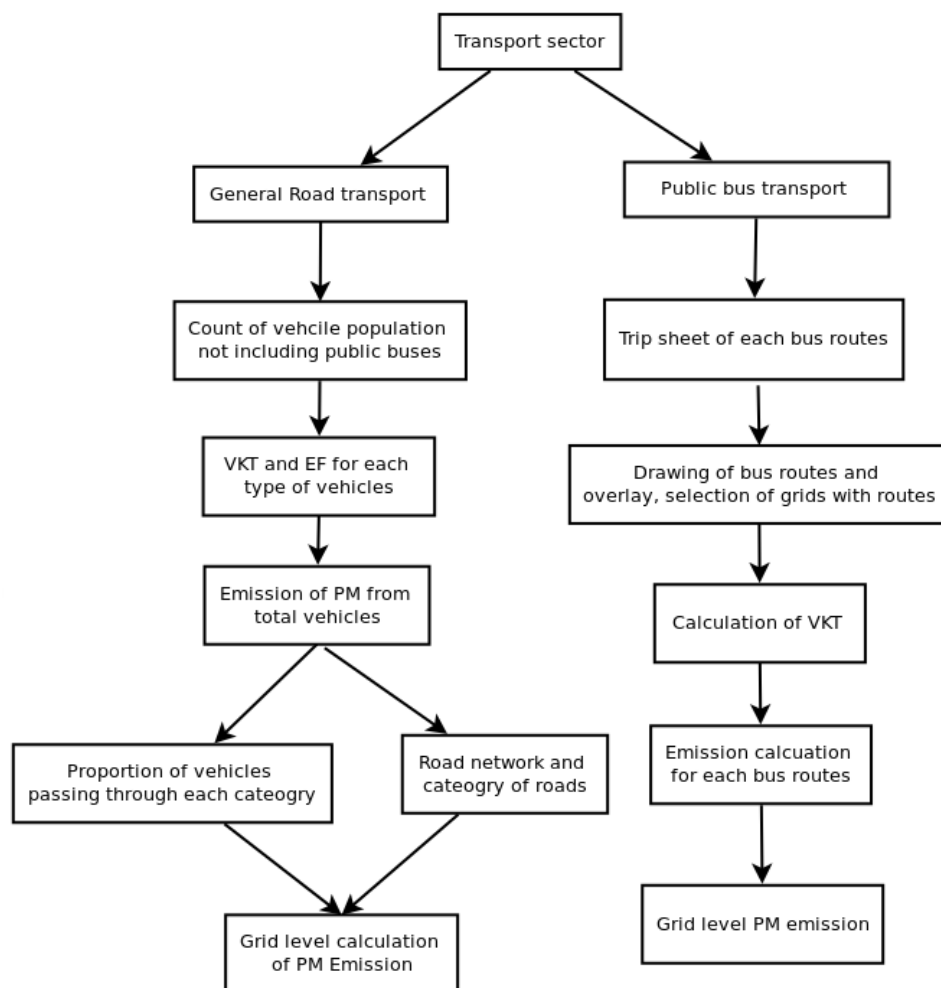


Fig. 4.4 Steps involved in Transport sector emission inventory estimation

⁵File named *EmissInvTransportCalc.py* in Nishadh et al. [131]

1. Determine category in which the road is falling, and its length in every EI grids. This involves following further sub steps
 - (a) Check the *intersection with-in* condition for each 3000 EI grids with road networks line feature data, which are in five different categories
 - (b) Since the last operation is computationally intense, *Rtree* index library was used for enhancing the computational performance
2. Apply the equation 4.2 to segregate the total vehicular emission into road network emission
3. The public bus transport passenger stops are located and incorporated into line map of each bus route
4. The line features were route network of each bus and is considered as road network. This followed by calculating transport sector emission inventory
5. Aggregate the total vehicular and public bus transport related emission inventory

Python code⁶ was used to calculate the particulate emission from windblown dust. The code was meant to execute two sequential steps involved in it.

1. Total Particulate matter emission from windblown dust was calculated based on the equation 4.3.
2. Spatial segregation of the emissions was carried out following the total vehicular EI steps and equation 4.2

For calculating industrial sector EI, *Python* based code⁷ that executed two subsequent steps. The Figure 4.5 depicts the steps involved in the calculation.

⁶File named *EmissInvWindblownCalc.py* in Nishadh et al. [131]

⁷File named *EmissInvIndustryCalc.py* in Nishadh et al. [131]

1. Determine the EI grids having industrial point location using *Intersection with-in* operation.
2. Calculate the EI for industries sector and aggregate it for all the industries located within the EI grid boundary.

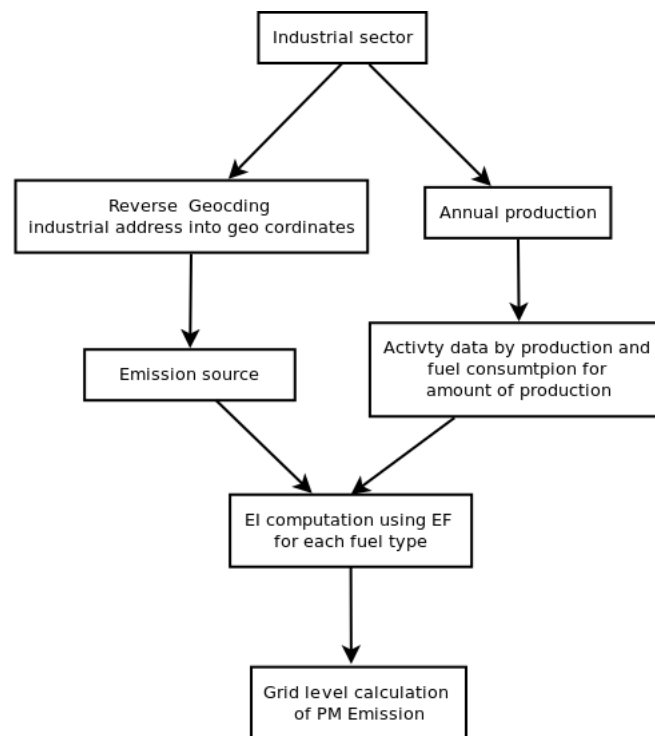


Fig. 4.5 Steps involved in Industrial sector emission inventory estimation

The overall EI for both $PM_{2.5}$ and PM_{10} for the whole study area, were derived using the sector wise EIs. It is also shown here that all EI computation steps can be carried out using *Python* and its geospatial libraries with no other external or proprietary software. The script form of the program enables generating a complete set of modules or graphical user interface (GUI) based stand-alone software for EI computation.

4.3.2 $PM_{2.5}$ emission

The Figure 4.6 depicts the spatial distribution of total $PM_{2.5}$ emission from different sectors of Coimbatore region for the year 2012. It is estimated that during 2012, a total of 12125.7

Ton $PM_{2.5}$ was emitted from different sources in Coimbatore region. In that, industrial sector is the major contributor (6549.4 Ton/year) followed by Transport sector (4488.6 Ton/year), Windblown dust (918.1 Ton/year) and lastly Residential sources (169.5 Ton/year). The relative contribution of industrial sector is 54.0%, transport sector 37.0%, windblown dust 7.6% and residential sector 1.4 %. The spatial level distribution of $PM_{2.5}$ from each of these sectors is given in the figures 4.7 to 4.10.

4.3.3 PM_{10} emission

The Figure 4.11 depicts the spatial distribution of total PM_{10} emission from different sectors in Coimbatore region for the year 2012. It is estimated that during 2012, a total of 13112.9 Ton PM_{10} was emitted from different sources in the region. Of the four sectors, industrial sector is the major contributor with 6631.7 Ton/year followed by windblown dust 3842.5 Ton/year, Residential sector 1608.1 Ton/year, and Transport sector 1030.4 Ton/year. In terms of relative contribution, industrial sector was 50.6 %, transport sector 7.9 %, windblown dust 29.3% and residential sector 12.3 %. The spatial level distribution of PM_{10} from each of these sectors is depicted in the figures 4.12 to 4.15.

4.4 Discussion

The study addressed localized data requirement and geospatial programming needs for urban level emission inventory development. For Coimbatore region, Data from four major sources were used for particulate pollutant emission inventory development. Data availability in dispersed and non-reusable multi data format was a challenging step in collating the data. Reusable data formats include CSV (Comma Separated Value format), while non-reusable formats includes those such as PDF (Portable Document Format). The data from the Census of India, NSSO, RTO and TNPCB were received through online or by personal visit to the respective offices. The data sets were available as word document, PDF, and excel sheets which are difficult to use, without re-entry, for further analysis and inventory computation. The geospatial data, such as the village maps, required were also available as PDF, making

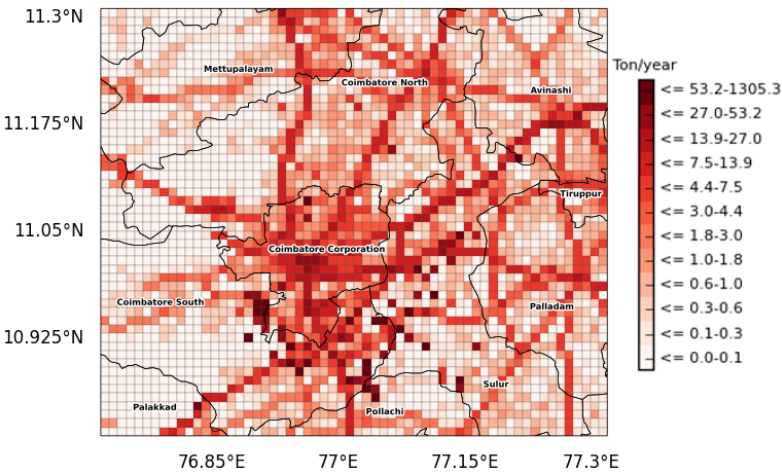


Fig. 4.6 Total $PM_{2.5}$ emission from Coimbatore

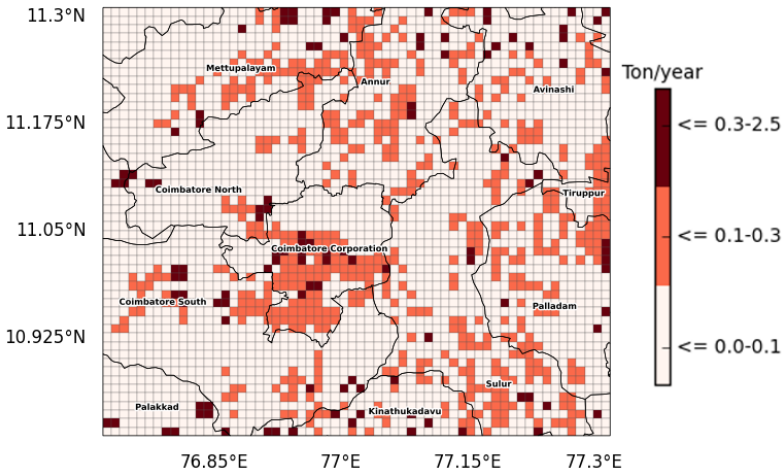


Fig. 4.7 $PM_{2.5}$ emission from residential sector in Coimbatore

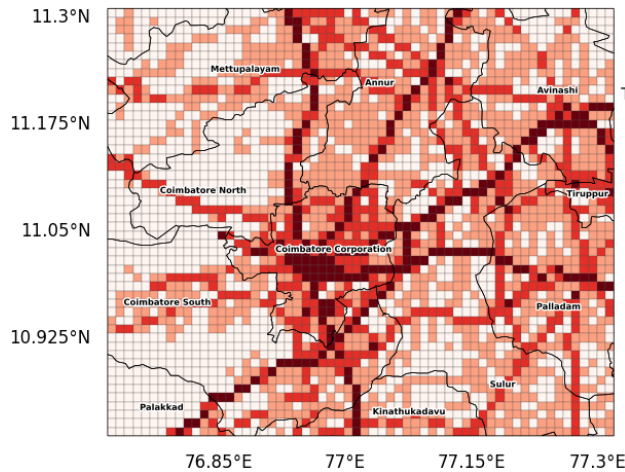


Fig. 4.8 $PM_{2.5}$ emission from transport sector in Coimbatore

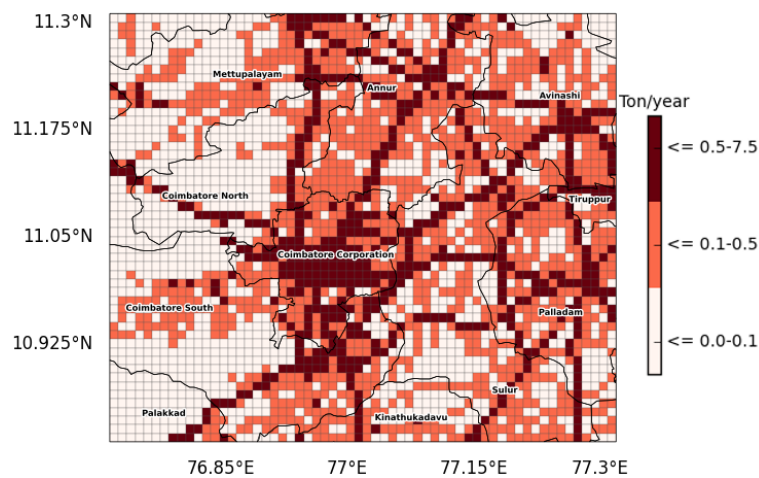


Fig. 4.9 $PM_{2.5}$ emission from wind blown dust in Coimbatore

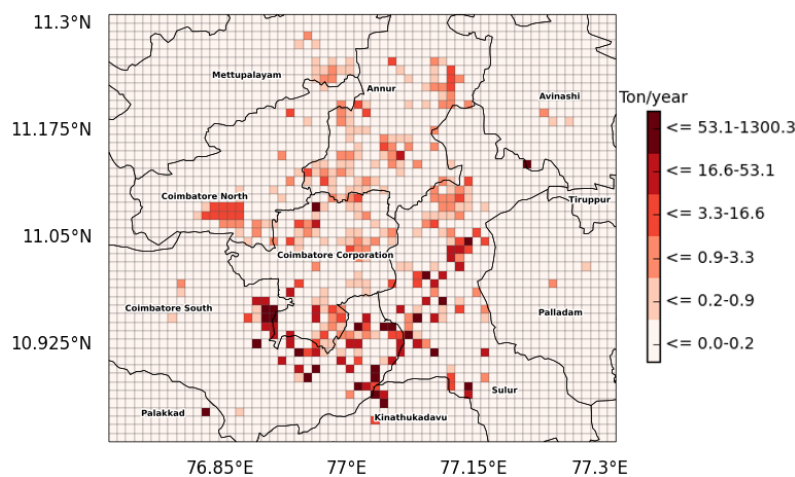


Fig. 4.10 $PM_{2.5}$ emission from Industrial sector in Coimbatore

it essential to redraw for further use for estimation of residential sector emission inventory. Online publication of data also has to be made the practice especially in the case of second tier urban centres. Recent policy level changes towards Application Programming Interface (API) [179] based dissemination of data from all government departments in India can be a solution for this data format issue and facilitate easy reuse of massive data.

In the current study, the data used for calculating public transport system emission was from an internal (to the specific department - TNSTC) fleet-management software. It would be advisable that transport related data especially for public transport system is made available in open data standards adopted in systems such as Global Transit Feed System (GTFS) [180].

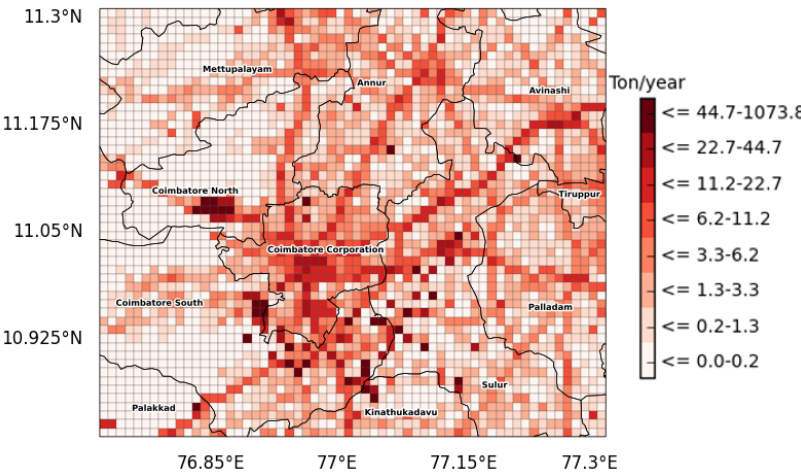


Fig. 4.11 Total PM_{10} emission from Coimbatore

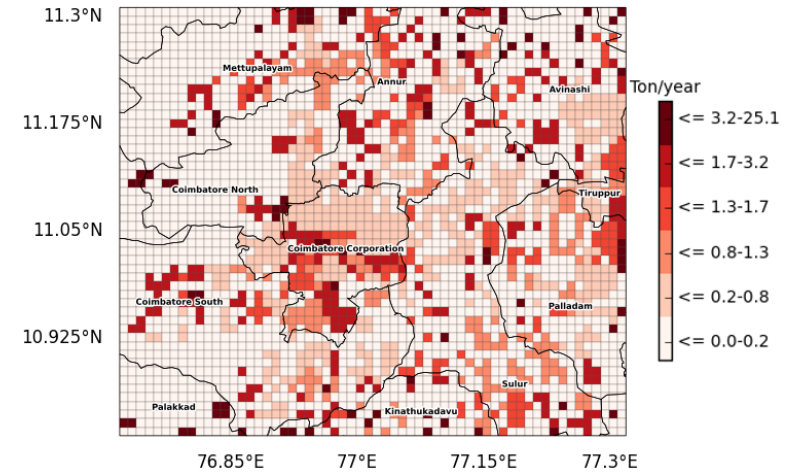


Fig. 4.12 PM_{10} emission from residential sector in Coimbatore

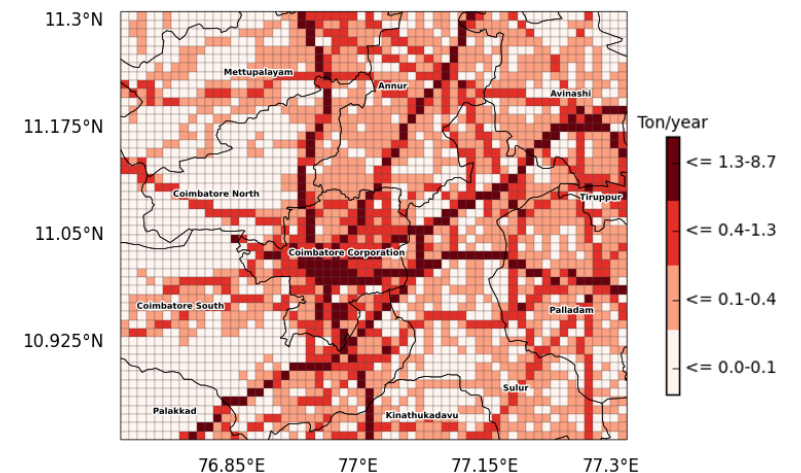
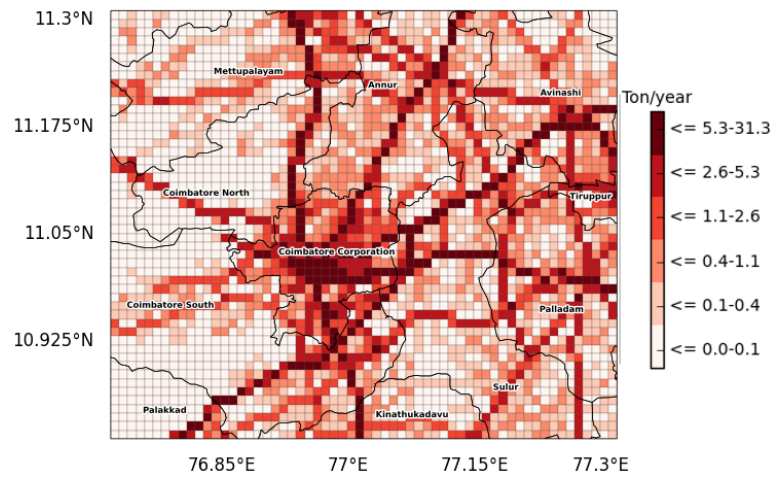
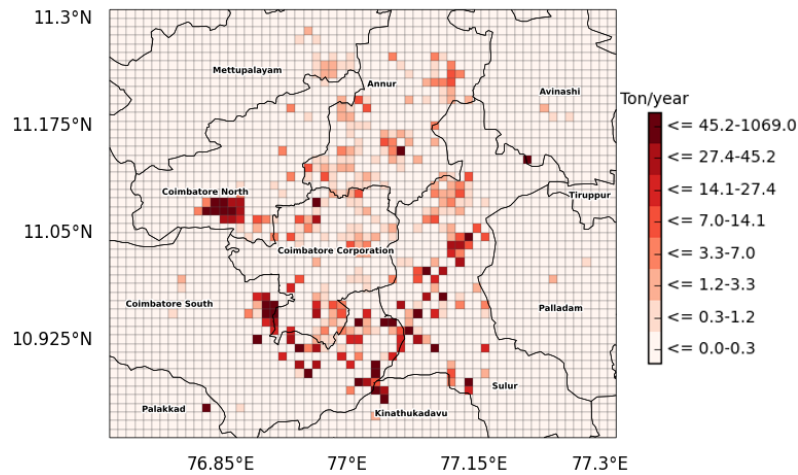


Fig. 4.13 PM_{10} emission from Transport sector in Coimbatore

Fig. 4.14 PM_{10} emission from wind blown dust in CoimbatoreFig. 4.15 PM_{10} emission from Industrial sector in Coimbatore

GTFS is a widely used data form for transport feed application and in the form usable to carry out EI calculations. Availability of up-to-date data is a major requirement for improving the accuracy of EI, crucial especially for transport sector EI calculations [181]. The fleet composition, size and age of different type of vehicles, have decisive role in emission composition and its temporal variability. Hence, recent vehicular traffic data or dedicated manual field data collections for count and composition are essential for computation of vehicular sector emission inventory [181, 163]. In the current study, relative proportions of vehicular traffic through different road categories were used; but it has limitation in terms of field verifiability and thus accuracy of estimating the road traffic. Despite being a second tier but fast growing

urban centre, Coimbatore lacks real time traffic feed data and it is essential that such important field data needs to be collected, documented for better EI development. Video based traffic analysis, such as web based video annotation tool *Vatic*[182] can be used for collecting and analysing the video records of road traffic to count and find the fleet composition. That would be handy for issues such as EI development as well as effectively planning for smooth traffic and also reducing the emission loads.

Open source tools based on *Python* programming language was used for data preprocessing and then emission inventory computation. Data organization and formatting could be made simpler by using various libraries of *Python* such as *regex*, *csv*. Integrating geospatial component with the collected data was carried out using *geopandas*[176] library of *Python*, which provides the tool for data-frame based organization of geospatial data. That is effective in organizing large geospatial data points such as thousands of industrial locations, road network line segments and village polygons. *Shapely* [177], a computational geometry library for *Python* was used for all the geometric operations. Grid level geometric operation such as *intersection within*, slicing line segments and polygons based on grid boundary could be carried out using the library. As these are computationally intensive operations with large geospatial data source, the operations were optimized by spatial indexing using *Rtree*[178] package in *Python*. That considerably reduces the computational time required for line segment operations and for applying the equation 4.2 for transport sector emission computation. From the study, it is observed that the application of *Python* based free and open source geospatial tool is much suited for emission inventory preparation. Furthermore, the used packages are scalable and thus will be usable to apply for large-scale region or country level emission inventory preparation. Another advantage of using *Python* based open source tools is its interactive publication of workflow pertaining to the emission inventory preparation. *Ipython notebooks* [183], a python and web based interactive platform for recording and reproducing the data analysis can be used for such purposes. *Ipython* is widely used in computationally intensive research fields such as bio-informatics to publish the codes or workflow used for data analysis [184]. This feature is widely viewed as an important aspect for reproducible research and communication of complex workflows in such fields. Publication of emission inventory calculation in such a platform can

significantly improve the community involvement and continuous improvement in emission inventory database. Regional data-backed stand-alone application can also be developed using *Python* and its extensive libraries for web and desktop application. Current regional applications (as seen in European Emission Inventory) for emission inventory computations can be improved in this regard [185].

The present study reports for the first time gridded high-resolution emission inventory for Coimbatore region, using open source software. It is observed that in the region, Industrial sector is largest emission source of particulate pollutant followed by transport and residential sector. The spatial distribution of the emission shows that southern part of the geographic domain is host to largest industrial emitters. The Southern West part of the Coimbatore is part of Palghat gap of the Western Ghats mountain chain and home to a large rock mining and crushing unit. In effect, this region also host two large cement manufacturing units. The emissions indicate that stringent regulatory measures and regular follow-ups directed towards foundries, crushing and cement manufacturing units can considerable reduce the particulate pollution emission in Coimbatore region. The emission estimates from public transport service indicate the extent and role of Diesel based buses on particulate pollution in Coimbatore region. The emissions estimated using trip sheet based accurate vehicle Kilometre travel indicate that to contain contribution particulate pollution fuel conversion from currently used diesel to CNG or electricity in public transport system may be a welcome change. The present exercise of developing Emission inventory for Coimbatore region brings out prevailing data gaps. It is observed that the residential emission inventory using census data has also issue. As the data is updated every decade, its extrapolation requires to be carried out carefully considering the spatial variation in the residential area development. In the case of transport sector, it is observed that emission from transport sector is largely dynamic and requires multiple data on actual vehicle population in the domain, engine technology used, age, Kilometre travelled etc [181, 163], which is largely not collected for second tier urban centres like Coimbatore. In the study, as a preliminary step the required data was collected from published reports from similar urban conditions. In the case of industrial sector, the data maintained by pollution control board is available for internal purposes only and is largely out of public domain. It hinders accurate

emission estimation for industrial sector. The lack of specific local emission factors for the fuels used and particulate pollutants especially relating to transport sector fuel consumption is another large data gap. In the current study, the accuracy of developed emission inventory was assessed by its application in air quality modeling and comparing with a global emission inventory (Chapter 5). Further study is warrant for reduce the data gaps and uncertainty in the emission inventory developed.

4.5 Conclusion

1. The study addressed localized data requirement and geospatial programming needs for urban level emission inventory development. Here, bottom-up approach for computing the emission inventory was adopted
2. As a measure to satisfy the requirement for real time particulate pollution modeling, $PM_{2.5}$ and PM_{10} emission inventory was developed for the Coimbatore region. The methodology for computation, and open source tools available for various geospatial operations necessary for the computation was discussed
3. *Python* based open source tools were used for various operations of emission inventory computations. Data processing and geometrical calculations were computed using various libraries in python programming language
4. The present study reports, for the first time gridded high-resolution emission inventory for Coimbatore region, using open source software. It is observed that in the region, Industrial sector is the largest emission source of particulate pollutant followed by transport and residential sector
5. The study indicated the data gaps prevailing in emission inventory calculation for second tier urban centres such as Coimbatore. The availability of data, data formats, outdated data and hurdles from various authorities pose limitation in emission inventory development

Chapter 5

Real time particulate pollution modeling

5.1 Introduction

Air quality is a combined effect of chemical and physical processes in atmosphere. The chemical and physical processes involved comprised of very dynamic process such as pollutant emissions, deposition, transformations, and transportation. Such processes are made more complex and dynamic by factors such as wind speed, direction, turbulence, radiation, clouds formation and precipitation. These processes are closely interlinked for manifestation of a certain air quality in an area [186]. As all these are motion related processes, their future state can be predicted by initial state, boundary conditions and employing mathematical equations of motion describing the flow of fluids [187], represented as a set of partial differential equations of motion and numerical approximations derived from a large set of observations on chemical constituents in atmosphere [187, 188]. Computer software programs are employed for solving the equations and that are termed as numerical models. The models helps in predicting future state of physical and chemical constituents such as aerosols or other atmospheric species, their distributions and levels with respect to the atmospheric motion. These can help in forecasting the pollutant concentration, its spatial distribution and thus the state of air quality in a particular time-frame. Computer programs solving these equations are grouped as Numerical Weather Prediction (NWP) models and Chemical Transport Models (CTM), which addresses

physical and chemical processes respectively, in atmosphere. Advancement in modern scientific understanding and advent of high-power computational capacity has considerably alleviated running numerical models. Consequently, these models are also gaining superior prediction accuracies and better reflection of the complexity of atmospheric processes in simulation [189]. In recent times, NWP and CTM are being integrated into unified models to consider the climate-chemistry-aerosol-cloud-radiation feedback mechanisms to further improve the understanding and forecast capacity [190].

Significant health and environmental effect of particulate pollution can be managed by pollutant concentration forecasts. That would help in issuing early warnings / alerts and taking precautionary measures against pollutants level hikes. For this, Real Time - Air Quality Forecast systems (RT-AQF) are established in some major urban centres [191] across the world. Numerical models are an important form of RT-AQF system, while empirical and parametric statistical models are the other two forms. Numerical model approach is advantageous over the other two approaches in terms of deterministic physics-based simulation (with time and space resolved) forecasting pollutants, precursors and possibly products (secondary pollutants). Numerical models also have lesser requirements for historical monitoring data as they are non-statistical models [191]. Weather Research and Forecast model with Chemistry (WRF-CHEM) is a coupled regional NWP and CTM widely used in RT-AQF system. In India WRF-CHEM was used for simulating South Asian air quality scenario [192], particulate pollutants PM_{10} [193], $PM_{2.5}$ [194], Aerosol, trace gases [195] and Primary Biological Aerosol Particles such as fungal spores [196]. It is also used for real time air quality forecasting in certain metro cities of India (System for Air Quality Forecasting and Research, SAFAR-India [197]). WRF-CHEM model is an open source application. Its development is through community involvement, encouraging wider adoption for atmospheric research and operational forecast purposes. Large user-base of the WRF-CHEM provides numerous write-ups, tutorials and user guides in Internet facilitating easy adoption of the model.

Being a typical numerical model, WRF-CHEM involves complex and numerous workflows and requires large computational power. This largely hinders its wide usage and operational air

quality modeling for second tier urban centres with less technical and financial resources. For easing the multi-component workflow involved in the model, automatic scripts are developed and used for real-time model execution. WRF-Environmental Modeling System (WRF-EMS) [198] was developed using Perl language to make automatic execution of the model with its numerous steps and data requirements. PyWRFscheduler [199] is similar to WRF-EMS, but is written in *Python* language and specifically directed to scheduled parallel execution in high performance local cluster computer environment. There are several *Python* [200, 201] based scripts to simplify the various operations of WRF simulation; but not many are known for WRF-CHEM simulation. Moreover, usage of WRF-CHEM in real time are required of high performance computing (HPC) environment [202], which can significantly restrict its usage for users with no access to HPC. In the case of model performance, it is limited by simulation errors arising due to inaccuracies in model treatment of emission, meteorology and understanding on local atmospheric conditions. It requires simulation domain specific evaluation studies to choose suitable model parameters scheme and sensitivity analysis [203, 204]. In this context, the current study was intended to addresses technical requirement for automatic real time execution of WRF-CHEM. To overcome the constraint of low access to HPC environment, we explored using the advantages of hireable on-line cloud computer services. WRF-CHEM model performance over Coimbatore region was assessed using 1 km resolution simulation and the output of the exercise was evaluated. The accuracy of emission inventory (EI) was evaluated by comparing model performance of the developed EI with global level EI.

5.2 Methodology

5.2.1 WRF-CHEM model and its components

The model WRF-CHEM [186] was developed by National Oceanic and Atmospheric Administration (NOAA), National Center for Atmospheric Research (NCAR), Department of Energy (DOE, USA) and several other independent contributors. The development is carried out through community approach [205] taking advantages of open source software development processes. This approach helps greatly in complex software development and

multidisciplinary collaboration in model development. WRF-CHEM is an extension of the widely used meteorological model WRF, with functionality to simulate chemical and meteorological variables' spatial and temporal distribution simultaneously. The on-line coupling of meteorological and chemical components facilitates identical transport scheme, model grids, time steps and physical parameters to use in simulation. This also enables feedback mechanisms between atmospheric chemistry and meteorology to be integrated in the simulation. *Fortran*, the language that was extensively used for scientific computations, was primarily used in WRF software development. Other than *Fortran*, programming languages such as *C* is used for parallel execution of the model and scripting languages such as *Perl* and *Shell* for model compilation in different computer architecture and *Netcdf* for model output. The WRF software architecture is based on Advanced Software Framework (ASF) [206], which provides necessary modularity and hierarchical organization for components of the model [207]. This organization helps in collaborative contributions and open source development of the model. In the case of WRF-CHEM, CHEM part is a contributed addition of CTM components to the WRF system. WRF-ASF is comprised of loosely coupled layers and components such as Driver Layer, Mediation Layer, Model Layer, Registry, and application program interfaces (API) [206]. Driver layer oversee the computational overall management and load distribution as per the model domain, nesting requirement, time loop, input /Output (I/O) operation in model domain, and act as an interface to the extension components. Mediation layer oversee the calls for advanced research WRF (ARW) dynamical core solver on model layer. Model layer oversee all computation routines for scientific codes involved in the model. Registry is an instruction set for code compilation of WRF system that determines the overall organization and functionality of the model. API oversees the input-output mechanism and extension of the model for various external models to interact with overall system in interoperable format.

The major components of the WRF-CHEM model are WRF pre-processing system (WPS), real data initialization and dynamical core solver advanced research WRF (ARW). The Figure 5.1 shows the components and its interaction in WRF-CHEM model system. The WPS carries out meteorological and static topological input data pre-processing according to the model domain for real data initialization. The necessary data on initial and lateral boundary conditions

is organized as per WRF domain grid in real data initialization. The WRF-ARW carries out model simulation in terms of time and spatial distribution using the supplied initial and lateral boundary conditions. To increase the resolution of simulation and reduce the computational cost, nesting option is followed in WRF-ARW to focus the simulation on area of interest. There are two types of nesting options in WRF-CHEM namely, one and two way nesting, which is based on the nature of interactions between the grids. In one-way nesting, initial and lateral boundary condition is passed in separate or concurrent simulation without any feedback mechanism between the nested grids. But in two-way nesting, feedback mechanism is allowed between fine and coarse grids.

Being a regional model, WRF-CHEM requires input data from global meteorological model such as Global Forecast System (GFS) for downscaling and simulating the atmospheric condition in the finite area delineated in model domain. Other than the meteorological data, the WPS requires emission information on model domain for chemistry (a sub-module of WRF). For initial condition preparation, emission information such as from biogenic, anthropogenic and biomass (burning) sources are required. Global and regional level emission inventories are used for such requirements. Chemical boundary conditions are included for accounting initial and background chemical concentration in the model domain. Variational data assimilation (a statistical method for including observational data) is included in WRF system for enhancing the validity of model simulation and optimizing the model estimate to real atmospheric conditions. The component for this purpose is 3D VAR, which assimilate observation data into the model during the integration stage. The output from WRF-CHEM model simulation is in *Netcdf* format and it is passed through several post-processing routines to get the required variables in time series or map representation. *Python* scripts are also available to post-process the model output in *Netcdf* format.

5.2.2 Real time automatic execution

On compilation, the WRF-CHEM system generates *Fortran* based executables for each major components. Those executables are depicted in Figure 5.2. The WPS comprised of executables namely *geogrid.exe*, *ungrib.exe* and *metgrid.exe*. The *geogrid.exe* defines the do-

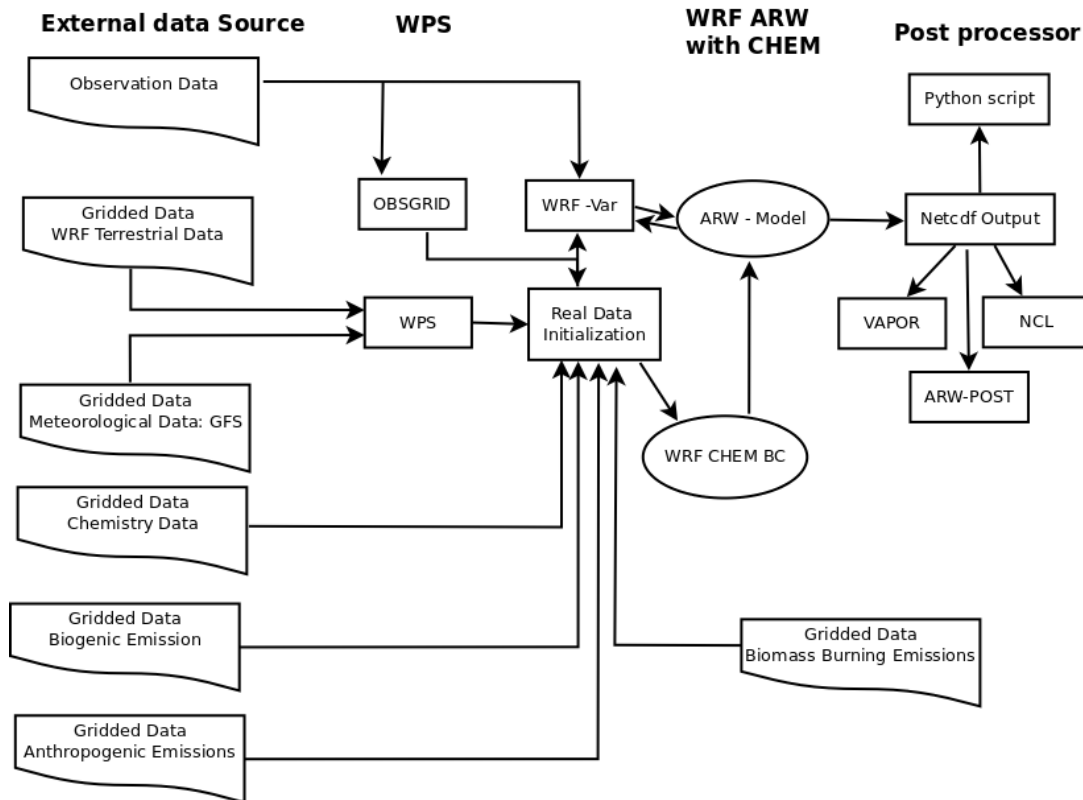


Fig. 5.1 Components of WRF-CHEM model system

main extent and interpolates topographical data for the domain. *Ungrib.exe* extracts the domain specific meteorological data from GFS output file. This output file has to be covering various time periods (temporal span for lateral boundary conditions) of the simulation. GFS output available in real-time every six hours is downloaded and included in the model simulation for this purpose. *Metgrid.exe* combines the output from the former two executables, makes the necessary horizontal interpolation for model domain and prepares the data for next model component, the *real.exe*. In real data initialization, the executables comprised of *convert_emiss.exe* and *real.exe*. The *convert_emiss.exe* processes the emission inventory data from various emission inventory processors such as *Prep_chem_Src.exe* [146] and *anthro_emiss* program. The *real.exe* carries out real data initialization. The WRF-ARW component, *wrf.exe*, carries out the simulation. It is aided with *ndown.exe* facilitating in one-way nesting option. As all these programs are Fortran executables, runtime execution is controlled by namelist files (input/wps files) that supply configurations on the operations of the model physics, girding, nesting and

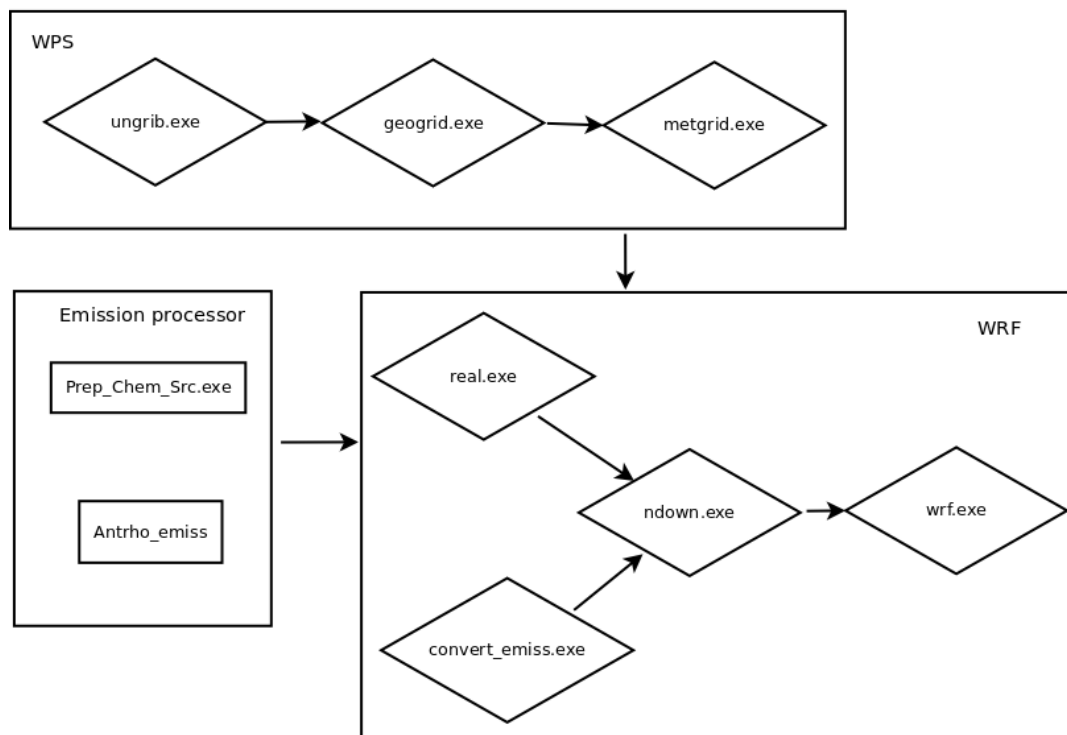


Fig. 5.2 Executable programs of WRF-CHEM model system

chemical parameterization scheme. The WPS execution is controlled by *namelist.wps* file, and *real.exe* and *wrf.exe* is controlled by *namelist.input* file. As running of these executables is sequential to editing of the namelist file for various functionalities (e.g. changing nested domain details), these steps are better to be in automatic script form to reduce the tediousness and manual interventions especially for real-time operational purposes. As explained earlier there are several programs available to execute automatically the WRF model, but to the best of our knowledge, none is available for WRF-CHEM. To develop an automatic script for WRF-CHEM, a *Python* based program [200] for WRF execution was modified to include the operation pertaining to the chemistry simulation.

5.2.3 Evaluation of computing resource requirement

The executables *wrf.exe* carries out all the computations pertaining to the atmospheric simulation. The other components of the WRF-CHEM system provide inputs regarding the required initial and lateral boundary conditions required for the computation. The component

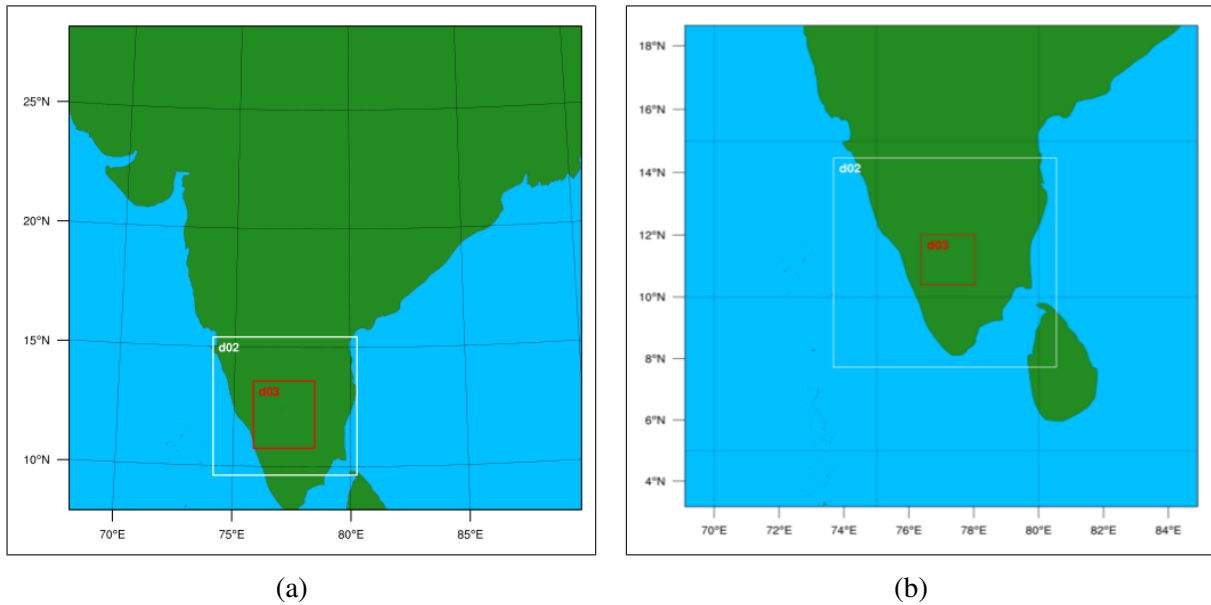


Fig. 5.3 Simulation domain for (a) computing resource evaluation study and for (b) model performance evaluation

wrf.exe derives the numerical solutions to the governing system of equations and physical process parameterizations used for atmospheric modeling. The numerical solution is subject to vertical coordinate and horizontal grid interpolation that represents the spatial distribution with steps for time integration for the model's specified timeframe. These steps are computationally intensive and necessitate the *wrf.exe* to be run in parallel execution. For this purpose, Message Passing Interface (MPI) program, which prompt utilization of multi-core processing power of computer for parallel execution, was used along with *wrf.exe*. For real time WRF-CHEM simulation, it is necessary to evaluate the computational requirement in terms of numbers of processor needed for time-bound completion of the simulation. Hence, the computational requirement for real time execution of WRF-CHEM simulation covering Coimbatore region was evaluated. Desktop computers and cloud computer clusters was used for the evaluation. We followed one-way nesting using *ndown.exe* for three nested domains chosen for Coimbatore region. The Figure 5.3a shows the domain used for computational resource evaluation. We used similar operating system, compilers and automatic execution script for the evaluation.

5.2.4 Evaluation of model performance for Coimbatore domain

Model performance evaluation is necessary to understand how much the simulation outputs represent the reality and satisfy the overall forecast quality. This is necessary in real-time air pollution predications to use the model output in objective decision support systems, in assessing the sensitivity of various physical and parameters used for forecast in specific area, seasonal variability and for further model development [208]. It is carried out by comparing the extent of similarity in model forecast against real world observations. Thus, atmospheric variables generated by model and ground observations by the sensors on identical spatio-temporal extent was compared using a set of performance statistics [209] to understand the extent of closeness between the predictions and real world observations. Some of the performance statistics used in the current study was as follows.

1. Fraction of prediction within a factor or two (FAC2): This gives measure of modeled value that falls within a factor or two from the observed values.

$$FAC2 = 0.5 \leq \frac{M_i}{O_i} \leq 2.0 \quad (5.1)$$

2. Mean bias (MB): Mean bias gives the state of the prediction mean, over or under estimate with respect to the observed value.

$$MB = \frac{1}{n} \sum_{i=1}^n M_i - O_i \quad (5.2)$$

3. Mean Gross Error (MGE): Mean Gross Error gives the indication of mean error regardless of the bias.

$$MGE = \frac{1}{n} \sum_{i=1}^n |M_i - O_i| \quad (5.3)$$

4. Root Mean Square Error (RMSE): RMSE gives the extent of closeness between the model and observation value.

$$RMSE = \left(\frac{\sum_{i=1}^n (M_i - O_i)^2}{n} \right)^{1/2} \quad (5.4)$$

5. Correlation coefficient (r): This gives the strength of linear relationship between model and observation value.

$$r = \frac{1}{(n-1)} \sum_{i=1}^n \left(\frac{M_i - \bar{M}}{\sigma_M} \right) \left(\frac{O_i - \bar{O}}{\sigma_O} \right) \quad (5.5)$$

6. Coefficient of efficiency, COE: Based on Legates and McCabe (1999) [210] and Legates and McCabe (2012) [211], this performance statistics gives the effectiveness of model output in predicting the variation in the observed value with respect to mean of observation.

$$COE = 1.0 - \frac{\sum_{i=1}^n |M_i - O_i|}{\sum_{i=1}^n |O_i - \bar{O}|} \quad (5.6)$$

7. Mean Fractional Bias, MFB

$$MFB = \frac{1}{n} \sum_{i=1}^n \frac{2(M_i - O_i)}{M_i + O_i} \times 100\% \quad (5.7)$$

8. Mean Fractional Error, MFE

$$MFE = \frac{1}{n} \sum_{i=1}^n \frac{2|M_i - O_i|}{M_i + O_i} \times 100\% \quad (5.8)$$

To carry out these performance statistics, R statistical programming package *open-air* [209] was used. The reading from Metone™Aerocet 531S particulate matter monitor was used for $PM_{2.5}$ and PM_{10} observations. Readings from Tamil Nadu Agriculture Weather

Network's automatic weather station was used for meteorological observations. The WRF-CHEM output using EDGAR emission inventory and Coimbatore EI described in previous chapter was separately evaluated with the observations / readings. For this purpose, two separate WRF-CHEM simulations using each of the emission inventories were carried out. The EDGAR(Emission Database for Global Atmospheric Research) is a global level emission inventory and available mostly in Netcdf file format. The Netcdf file operator library for *Python* [212] was used for accessing this global level emission inventory and integrating the Coimbatore EI with it for air quality modeling. The Netcdf operator tool *nco* [212] was used to process the variable and monthly data into a form, which can be read by the pre-processing system, WPS, for air quality modeling.

5.3 Results

5.3.1 Real time automatic execution

Automatic script for real time WRF-CHEM execution was developed by modifying an open source *Python* program written for WRF execution [200]. The features required for WRF-CHEM execution, such as emission inventory pre-processing and conversion, were included in the current study. The Figure 5.4 shows the steps involved in the simulation and addressed by the modified script⁸. In the Figure 5.4, executables are represented in diamond symbols, output files in document symbols and the processes are represented in predefined process symbols. The script largely uses OS module that assists in executing operating system specific and external programs inside *Python* programming environment. The script uses a list of modules to execute the major operations of the WRF-CHEM execution. The functions such as namelist editing, time bound downloading of input files, and automatic execution of various modules of WRF-CHEM was defined in the respective modules. The module *edit_namelist* carries out operations of changing the namelist files to suite execution date and time on real-time basis. This module is modified for editing other required namelist files, such as *namelist.input* files for real data initialization and *anthro_emis.inp*, *prep_chem_sources.inp* files for emission

⁸File named *WrfChemRTExecuter.py* in Nishadh et al. [131]

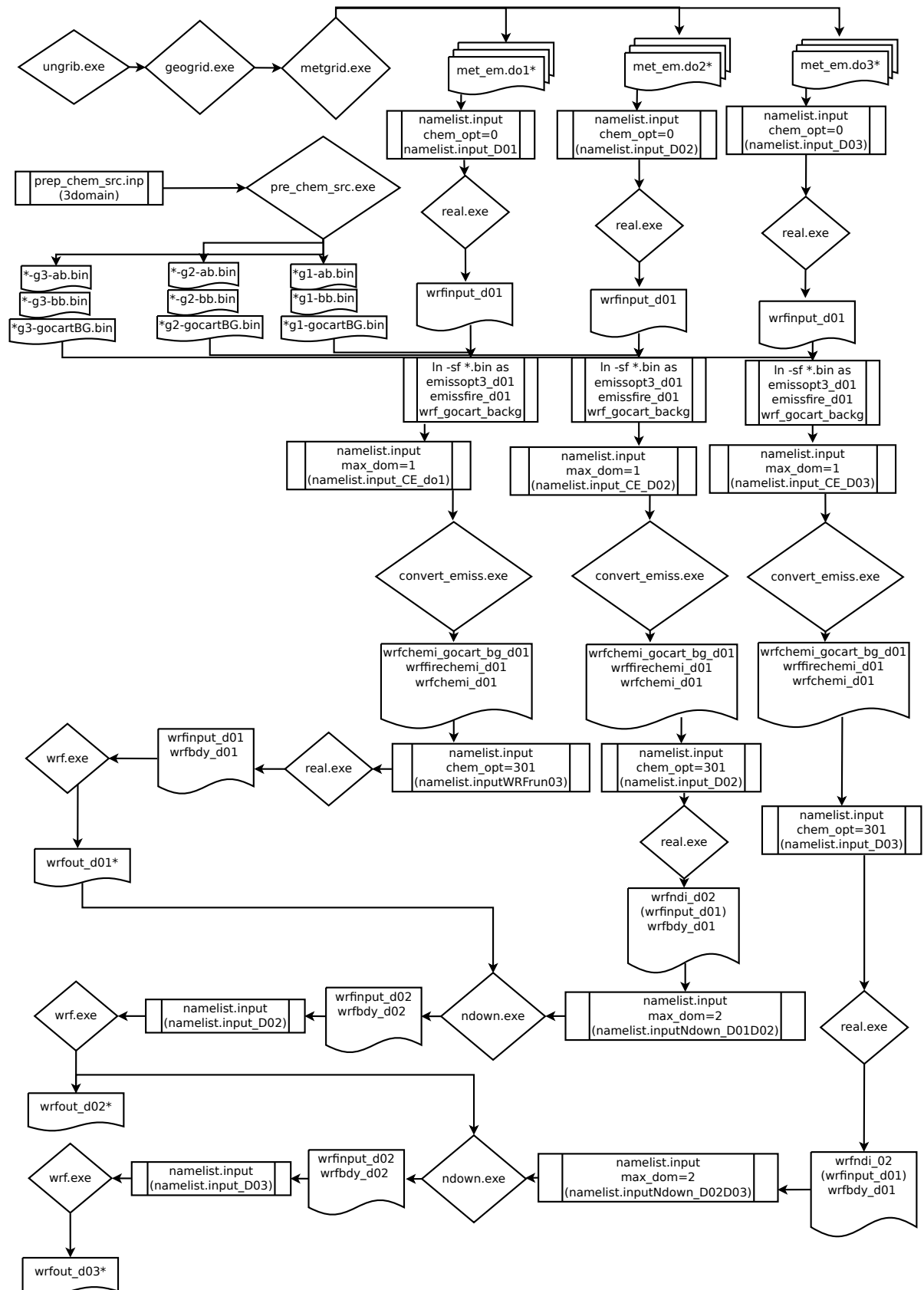


Fig. 5.4 Steps involved in WRF CHEM simulation addressed by automatic execution script

inventory processors. The module *get_init_files* carries out operations for downloading the input files to begin the simulation. It has functionality to ensure the availability of input data in the host computer (where simulation is in process) and download the required data from GFS web server, if not found in the host computer, based on time of execution. The GFS data is updated every six hours per day and the data has to be downloaded as per the simulation time-frame. The module *intelligent_run* operates the automatic execution of various sub modules of WRF-CHEM simulation. All operations are logged to store any error or issues reported during the simulation. The recording of the completion of the operations is stored in a separate file. Overall, these modules are used in serial manner to initiate the simulation and for one-way nesting of three domain levels. Each nesting operation was carried out one after the other (Figure 5.4). In each nesting stage, the script applies the module *edit_namelist* to change time period and domain grid details for the simulation, followed by *intelligent_run* module executing the WRF-CHEM simulation components. These steps are carried out after the script carries out input data download and start a log-record of the simulation time and other issue pertaining to the simulation. To understand further the issues with simulation as per the log records, the script provides facility to skip certain simulation steps. The log is also enabled for storing the time period taken for completing each simulation steps. This capability (logging the details) was essential for evaluating the computational requirement of WRF-CHEM simulation for Coimbatore region in the present study.

5.3.2 Evaluation of computing resource requirement

Using the logging facility in *Python* that records the time taken by each component to complete operation, total computational requirement of real-time WRF-CHEM simulation for Coimbatore region was assessed. It involved a comparison of the time taken for six-hour simulation by different computer processors. It was found that the cloud computer cluster Amazon™web Services Elastic Computer Cloud (AWS-EC2) takes the shortest time for one complete simulation (Table 5.1). In each computer, the WRF-CHEM model environment was setup by installing Linux based operating system, GNU based compiler such as *gfortran* and *CPP*. This is followed by installing required libraries such as *Netcdf*, *MPICH*, *Jasper*, *libpng*

Table 5.1 Wall clock time taken for various steps of WRF-CHEM execution in different computer processors

S.No	Processor	Compilation	WRF-CHEM six hour simulation			
			WPS	D01	D02	D03
1	Intel Core i5 3X core	<25 min	<1 min	3 hours 10 min	1 hour 17 min	6 hours 55 min
2	Xeon 3X core	<25 min	<1 min	23 min	46 min	4 hours 51 min
3	AWS EC2- 3X cluster	<25 min	<1 min	18 min	36 min	4 hours 10 min
4	AWS EC2- 30X cluster	<25 min	<1 min	4 min	9 min	58 min
5	AWS EC2- 60X cluster	<25 min	<1 min	3 min	7 min	50 min

WPS=WRF-Pre Processing, D01-Domain 01, D02-Domain 02, D03-Domain 03

and *zlib*. The library *Netcdf* is required for I/O (Input / Output) operations in WRF-CHEM simulation. *MPICH* is required for enabling parallel execution capability for the model. The libraries such as *Jasper*, *libpng* and *zlib* are required for WPS system to read the meteorological data from GFS, supplied in GRIB file format. The correctness of WRF-CHEM environment set-up was tested by a set of programs provided in Tutorial-WRF [213]. The tests involve checking the correctness of the installation and compatibility between compilers such as *gfortran* and *gpp*. This test is followed by checking the compatibility between *Netcdf* and *MPICH* libraries and building the various components of WRF-CHEM components. Three-level nested domain (Figure 5.3a) was chosen for simulation, which runs without any error in the range of computer processors chosen for the evaluation. The coarsest top domain grid points comprised of 90x85, followed by 76x73 and 97x106 with respective resolutions of 27 km, 9 km and 3 km.

The cloud computer cluster environment for AWS-EC2 was setup using *Starcluster* program. It is a *Python* based program to automate cluster creation, configuration and management. In AWS-EC2, the operating system and various configuration set-ups are supplied as individual self-contained and publicly shareable packages named as Amazon Machine Images (AMI). The AMI are shareable in the sense that, any AWS-EC2 user can use the available AMI and start using the system with pre-installed software and configuration. In the current study, a public AMI was configured with WRF-CHEM modeling system and *Starcluster* program. The AMI

Table 5.2 Model parametrization scheme used for performance evaluation study

Feature	Option	Description
Chemical Mechanism	MOZCART	MOZART Chemistry and GOCART aerosols (MOZCART) using KPP library
Microphysics	Lin et al. scheme	Scheme taking consideration of ice, snow and graupel processes
Longwave Radiation	Rapid Radiative Transfer Model (RRTM)	Scheme for multiple bands with trace gases and microphysics species
Shortwave Radiation	Goddard shortwave	Two-stream multi-band scheme with ozone from climatology and cloud effects
Surface Layer	MM5 Monin-Obukhov	Scheme having look up tables comprised of Monin-Obukhov with Carlsol-Boland viscous sub-layer and standard similarity functions
Anthropogenic Emissions	EDGAR and Coimbatore EI	Coimbatore Emission Inventory(Ei) with 1 km resolution and EDGAR EI with 10 km resolution
Biogenic Emissions	MEGAN	EI of gases and aerosols from nature based on MEGAN model

was designed using Paker.io, a web application for designing and developing specialized AMI. The *Starcluster* provides options to hire the cluster computer considering most economically viable option. AWS provides hiring options in terms of reserved or spot instances. In the spot instance option, AWS provides facility to bid the charges per hours competitively. The *Starcluster* program provides facility to process the bidding procedure automatically. This facility was used for cloud computer cluster creation and real time automatic execution in the present study.

5.3.3 Evaluation of model performance for Coimbatore domain

For evaluating the performance of WRF-CHEM simulation over Coimbatore region, a separate three-level nested domain with lowest resolution of 1 km was used (Figure 5.3b). The Table 5.2 lists the parameterization scheme used for the simulation. The simulation was separately carried out using the global and Coimbatore level emission inventories (EI). The EI that was developed for Coimbatore region was covering a total of 3000 km^2 surrounding the main urban and suburban regions. The WRF-CHEM, air quality modeling system used for the current study uses nested domain simulation. It carried out high-resolution simulation for the region starting from a coarser domain in the nested routine. The coarse domain takes

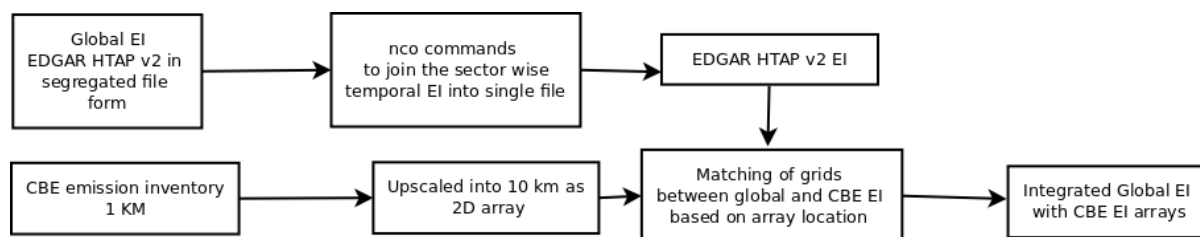


Fig. 5.5 Steps involved in integrating the Coimbatore local EI over global EDGAR HTAP EI

into account a larger region covering the whole of Indian sub-continent, and that needs EI in a global scale. This requires integration of the developed EI (for Coimbatore) with global level EI. *Prep_Chem_Src.exe* [146] is a widely used emission inventory pre-processing tool that plots the EI for air quality model domain such as the one we used, the WRF-CHEM system. *Prep_Chem_Src.exe* is used for plotting anthropogenic, biogenic, and biomass (burning) EI for the modeling domain. However, lack of documentation for using (with *Prep_Chem_Src.exe*) updated EI emission inventory available in Netcdf file format hinder its use in current study. Hence, we were constrained to use *Anthro_Emiss* program, a simpler preprocessing system for Netcdf based EI for anthropogenic emission inventory was used. The global level EI such as EDGAR HTAP v2 [155, 214] for the year 2010 with highest available resolution of 10 km are normally distributed in separate sector and month wise files, which are small in size. The small size files are most suitable for downloading, considering the difficulty in distributing large size file through web site. However, that would necessitate consolidation of sector and month wise global EI into single file which can be used with *Anthro_Emiss* program. Netcdf operator(*nco*) command line program was used for this consolidation. The Figure 5.5 depicts the steps involved in integrating Coimbatore EI with global EI.

The *nco* commands such as *nces*, *ncecat*, *ncap2*, and *ncrename* was used to sum the EI sector files, join months wise file for 2010 into a single file, attach temporal variable details and renaming the file. A *Python* script⁹ was used to run the *nco* commands internally to process individual files in EDGAR HTAP v2 global emission inventory. A separate *Python* script¹⁰

⁹File named *EdgarEIProcessorNCO.py* in Nishadh et al. [131]

¹⁰File named *CbeEIEdgarJoin.py* in Nishadh et al. [131]

was used to process the Coimbatore EI and integrate with the global EI. The script does four subsequent steps and are as follows,

1. Pre process the Coimbatore EI in shape file format for similar units ($\text{kg} / \text{m}^2 / \text{second}$) and spatial resolution of 10 km
2. Access the global EI file for grid position and determine the nearest grid points of the Coimbatore EI by taking EI grids as 2D array
3. Integrate the global EI and Coimabtoe EI based on most nearest grid centroid points
4. Output the integrated global EI with all the necessary metadata suitable for *anthro_emiss* program

The output file was used with *anthro_emiss* program to generate the anthropogenic emission inventory file for WRF-CHEM simulation. For every nested domain separate *anthro_emiss* execution was carried out following the above steps.

The Table 5.3 shows statistical performance measures of WRF-CHEM simulation carried out using Coimbatore EI and EDGAR EI. It was observed that the FAC2 was 1 for temperature and relative humidity in both simulation attempts. FAC2 value is lowest for PM_{10} , followed by $PM_{2.5}$, wind direction and wind speed. In the case of Mean bias (MB), Relative Humidity recorded positive value while other variables had negative values. In MB Temperature had the lowest negative value followed by wind speed, wind direction, $PM_{2.5}$ and PM_{10} , indicating high bias for PM_{10} and $PM_{2.5}$ in both the simulation attempts. Mean gross error (MGE) was high for PM_{10} followed by $PM_{2.5}$, wind direction, wind speed, relative humidity and least for temperature. Normalized Mean Bias (NMB) is positive for relative humidity, while other variables had negative values with highest negative value for $PM_{2.5}$. Highest value of Normalized Mean Gross Error (NMGE) was for wind direction in both the simulations followed by PM_{10} , $PM_{2.5}$, WS, RH and Temperature. Root Mean Square Error (RMSE) was highest for PM_{10} followed by wind direction, $PM_{2.5}$, wind speed, relative humidity and temperature. Highest correlation coefficient (r), was observed for relative humidity followed by Temperature,

Table 5.3 Performance measures for WRF CHEM simulation based on Coimbatore EI (CBE EI) and EDGAR EI

Variables		FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE	MFB	MFE
CBE EI	Temp	1.0	-1.3	2.0	-0.05	0.078	2.3	0.818	0.27	-5.29	8.22
	RH	1.0	4.2	9.5	0.079	0.179	12.1	0.832	0.16	4.74	16.93
	WS	0.495	-2.5	2.8	-0.486	0.528	3.3	0.17	-0.6	-56.42	65.5
	WD	0.381	-13.0	69.4	-0.168	0.901	113.3	0.047	0.1	45.65	87.65
	PM _{2.5}	0.157	-56.0	56.0	-0.674	0.674	65.6	0.277	-1.13	-96.95	96.95
	PM ₁₀	0.043	-145.0	145.0	-0.772	0.772	164.5	0.129	-1.59	-121.22	121.22
EDGAR EI	Temp	1.0	-1.3	2.0	-0.049	0.078	2.3	0.819	0.27	-5.26	8.23
	RH	1.0	4.1	9.5	0.077	0.178	12.0	0.833	0.16	4.5	16.82
	WS	0.533	-2.5	2.7	-0.483	0.523	3.3	0.2	-0.59	-55.98	64.65
	WD	0.39	-14.1	69.2	-0.183	0.898	113.9	0.016	0.11	43.55	87.1
	PM _{2.5}	0.114	-57.7	57.7	-0.694	0.694	67.0	0.29	-1.2	-101.07	101.07
	PM ₁₀	0.029	-148.5	148.5	-0.79	0.79	167.8	0.074	-1.66	-125.66	125.66

Temp-Temperature, RH-Relative Humidity, WS-Wind speed, WD-Wind direction

$PM_{2.5}$, wind speed, PM_{10} and wind direction. Highest value of Coefficient Of Efficiency (COE) was for Temperature followed by relative humidity and wind direction. For wind speed, $PM_{2.5}$ and PM_{10} , COE was negative values. In the case of Mean Fraction Error (MFE) and Mean Fractional Bias (MFB), lowest percentage was for temperature followed by relative humidity, wind speed, wind direction, $PM_{2.5}$ and PM_{10} .

Hourly temporal variation of $PM_{2.5}$ and PM_{10} compared with the readings from Aerocet-531S monitor is shown in the Figure 5.6. It is observed that in both the simulations, $PM_{2.5}$ and PM_{10} recorded a trend of variability similar to that of the field observations despite notable differences in the concentration values. The Figures 5.7, 5.8, 5.9, and 5.10 shows spatial variability of daily average $PM_{2.5}$ and PM_{10} concentrations, the outputs of two WRF-CHEM simulations using Coimbatore EI and EDGAR EI respectively. The spatial variations of $PM_{2.5}$ concentrations are in the range of 22.8-183.8 $\mu\text{g}/\text{m}^3$ in simulation using Coimbatore EI. The high concentrations predicted were exceeding 145 $\mu\text{g}/\text{m}^3$ in all the simulation days. In the case of simulation using EDGAR EI, $PM_{2.5}$ was in the range of 19.0-133.0 $\mu\text{g}/\text{m}^3$ with high concentration exceeding 50 $\mu\text{g}/\text{m}^3$ only on two days. The spatial variation of PM_{10} concentrations were the in range of 39.5-352.7 $\mu\text{g}/\text{m}^3$ in the simulation using Coimbatore EI and high concentration was exceeding 250 $\mu\text{g}/\text{m}^3$ in all the simulation days. In the case of simulations using EDGAR EI, PM_{10} was in the range 29.3-156.8 $\mu\text{g}/\text{m}^3$ with high

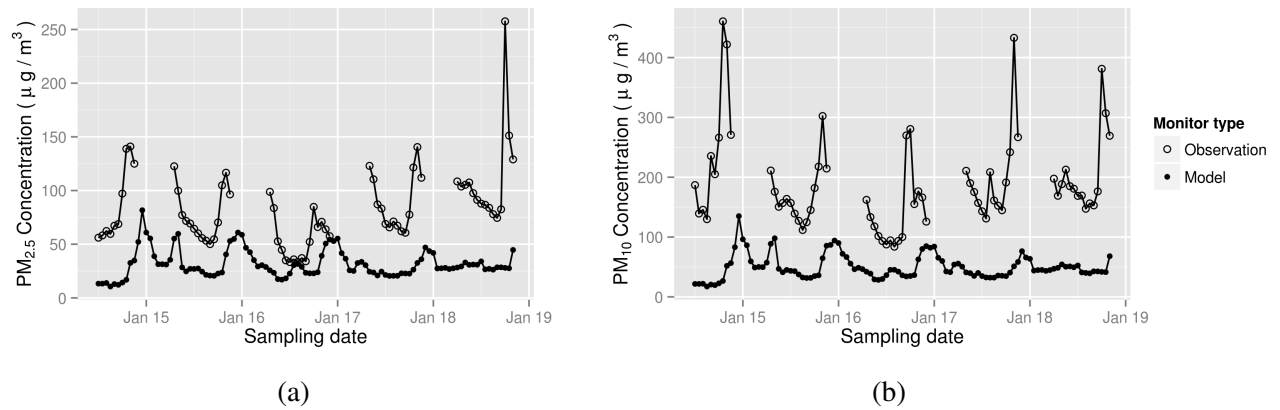


Fig. 5.6 Hourly temporal variation of $PM_{2.5}$ (a) and PM_{10} concentration WRF CHEM simulated using CBE EI and observation by Aerocet 531S monitor

concentration exceeding $100 \mu g/m^3$ only twice. The range of pollution concentration recorded by Aerocet-531S was $30.1-585.7 \mu g/m^3$ with mean of $81.16 \pm 39.56 \mu g/m^3$ for $PM_{2.5}$ and $30.1-1299.7 \mu g/m^3$ with mean of $192.58 \pm 113.04 \mu g/m^3$ for PM_{10} for all the days when simulation was conducted. This indicating the CBE EI simulation output was in close range with field observations. EDGAR EI used simulation was in fact underestimating the concentrations compared to field observations. The spatial distribution maps indicate $PM_{2.5}$ and PM_{10} to be in high concentration throughout the simulation period; highest in North eastern and Southwestern part of the Coimbatore region. The higher concentrations were predominantly seen outside the Coimbatore urban region (corporation limits). Lower concentrations throughout the simulation days were also the case for evaluation around the observation station (Kuniamuthur - KNMR - mentioned in Chapter 3). The wind rose (Figure 5.11) diagram covering the KNMR shows strong wind speed (8-9 km / hour) towards Northeastern direction during the simulation days. This is also reflected in the case of forecasted wind profile, output from the model simulation. This might be the reason for higher pollutant concentration forecasted in the North eastern part of the Coimbatore region.

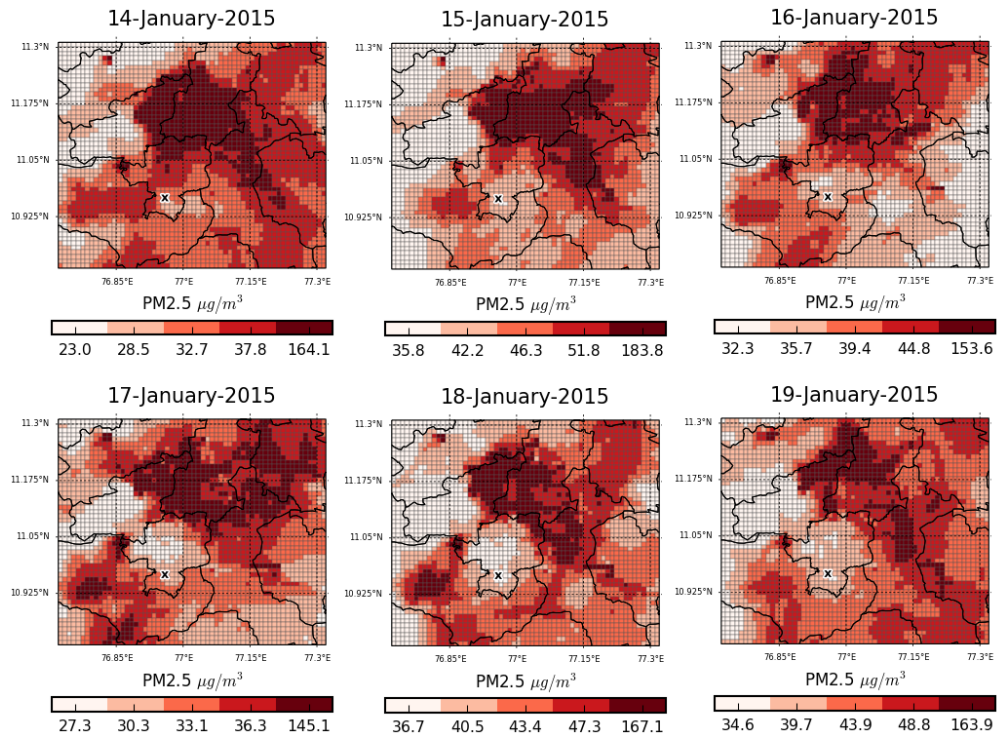


Fig. 5.7 Daily average spatial distribution of WRF CHEM simulated $PM_{2.5}$ by using CBE EI

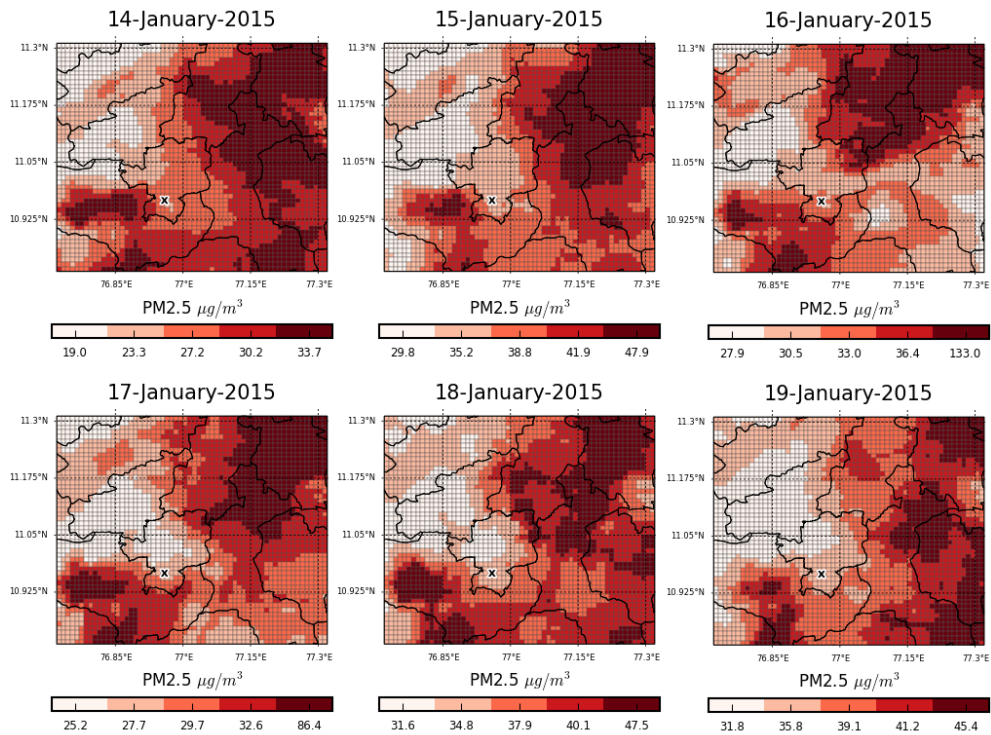


Fig. 5.8 Daily average spatial distribution of WRF CHEM simulated $PM_{2.5}$ by using EDGAR EI

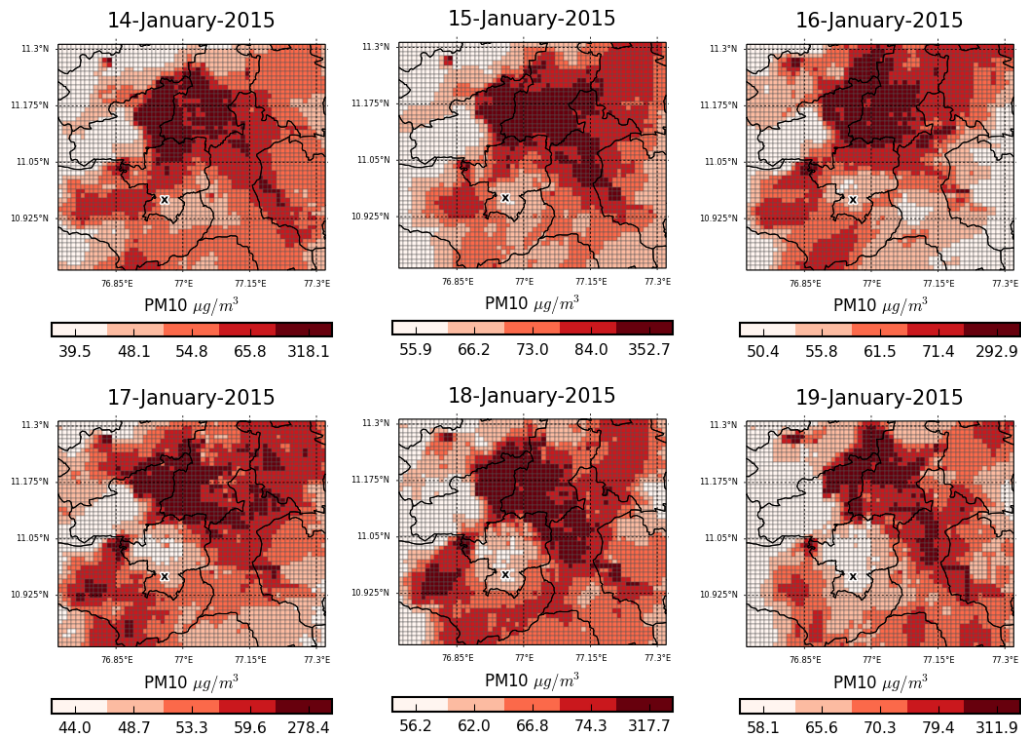


Fig. 5.9 Daily average spatial distribution of WRF CHEM simulated PM_{10} by using CBE EI

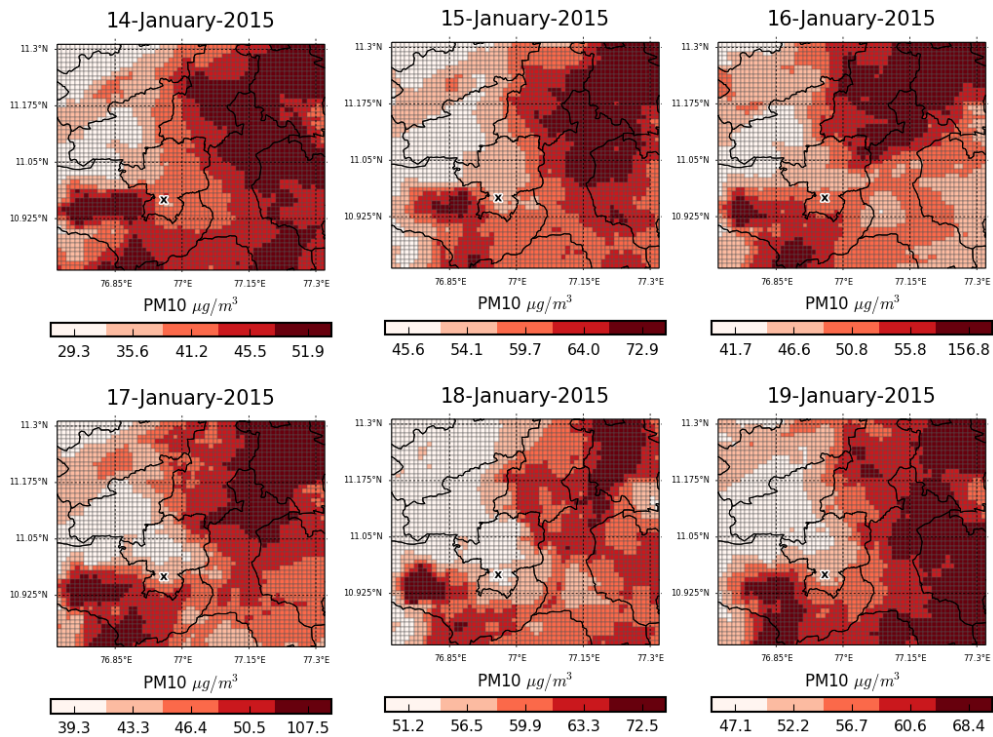


Fig. 5.10 Daily average spatial distribution of WRF CHEM simulated PM_{10} by using EDGAR EI

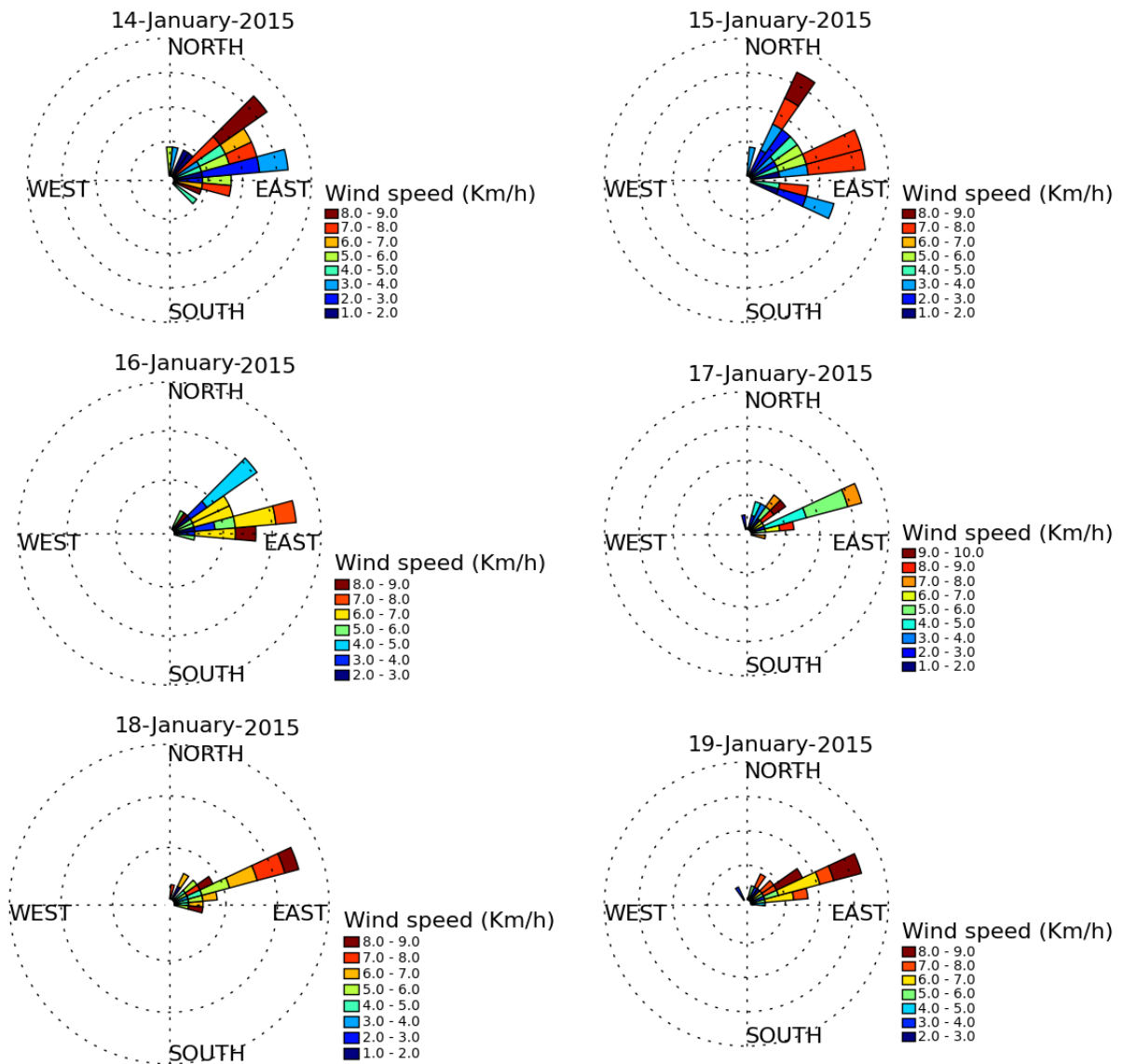


Fig. 5.11 Wind rose diagram of the automatic weather station wind profile during the simulation period

5.4 Discussion

Real time air quality modeling using WRF-CHEM requires time bound execution of various component of the model. This requires a means for automatic execution of the components and high-performance computing environment. The current study addresses these challenges with respect to use of the model for Coimbatore urban region simulation. According to Harrop et al.

[215], execution of atmospheric models is a ‘laborious process’; further, real-time execution of such model using high performance computing environment a ‘notoriously difficult’ task for researchers. To address these challenges the Harrop et al. [215] proposed a Workflow Management System (WMS), which addresses all the components of WRF simulation with ease of intuitive Graphical User Interface (GUI) named WRF-PORTAL [216]. However, an alternative to the WMS, a command line based application WRF-EMS [198] is widely used for automatic execution of WRF. This is developed using Perl and WRF portal using Java language. The tools available for automatic or simplification tools for executions of WRF are self-contained and complex with numerous operations and functions. They only provide facility for minimal intervention to understand the various components of WRF working behind the program. Moreover, the programming language used for the software such as Perl are largely considered as tough interpreted language [217, 218] and cryptic [218] for examining the processes carried out by the software program. As these are complex in entirety with myriads of operations, finding a separate functionality then tweaking or adding new functionality is difficult for a moderate programmer proficient only in respective language and it is virtually impossible for novice scientific background user [219]. That greatly restricts user contributed extension of facility in these programs such as extending the capability of WRF only simulation facility into WRF-CHEM.

On the contrary, *Python* programming language has standard syntax, simple and easy readability of code functions [220, 218]. This makes the *Python* based scripts easy to use and modify as per user requirements [221, 222]. There is gradual gain in usability of *Python* for scientific computing with field specific application and library development [223–228]. In atmospheric sciences, there are numerous libraries developed to address the unique data array and computational requirements of the field [222, 229]. In WRF simulation management, automatic scheduler [199] programs are available which has the functionality to oversee the execution queuing. It is much useful for managing the WRF simulation in a High-performance computing environment to optimize the computer resources. In the current study, a *Python* programming language based script [200] was used and modified to suite WRF-CHEM execution. This script is used for its simplicity and effective integration of extension module for

step involved with WRF-CHEM. Although not a complete system of WMS, the script provide essential functionality to manage the execution with all its input requirement. As it is an open source program, future development on the script can be carried out for optimizing the script and enabling the graphical user interface features to make the script a full-fledged WMS. The visualization libraries available with *Python* with seamless I/O operation on Netcdf file is an added advantage to use the language for Post-processing steps under WRF-CHEM simulation.

The current study showed the effectiveness of using cloud computer based clusters for WRF-CHEM simulation. It is observed that a four to six fold decrease in computational time while using cloud computer and increasing its cluster setup by 10 folds. However, further increase in number of clusters in cloud computer, does not proportionately raise the performance in terms of reduction in simulation time. Nevertheless, this observation has to be further explored with cluster setup optimization and WRF-CHEM parallel execution for gaining higher performance from cluster setup. In overall, the cloud computer instance set-up gives economical advantages; Energy consumption charges, capital and recurring maintenance expenditure due to local HPC setup. The present study addresses the scenario, of non-availability of HPC and budget constraints to setup a dedicated HPC, prevailing in most of the developing countries especially for second tier urban centres such as Coimbatore. It shows cloud computing with optimum number of clusters to be a solution. Considering the time taken for completion of the simulation, cluster environment is inevitable for real time air quality modeling using WRF-CHEM. In India, custom-built computer cluster [230, 231] or super computer based HPC [232] are used for operational and research related simulations. Advantages of using cloud computer services for scientific application are recognized for long for computationally intensive atmospheric modeling and simulations [233]. Hiring on demand and scalable computing resources offered by cloud computers are considered to be an economically viable option for HPC operation in scientific computing [234]. Cloud computing field knowledge organization [235] and related case studies on different cloud computer services performance and advantages [236–240] has widely undertaken. It is observed that cloud computer based cluster setup is most suitable for real time WRF-CHEM simulation under the present scenario. However, that requires further studies on optimizing the computing resources for better performance.

In real time WRF-CHEM modeling, it is required to start and complete the simulation based on input data availability, such as every six hour updated GFS meteorological data. In this case, on demand hiring of EC2 has to be automated for operational efficiency. The program *Starcluster* offers the facility to automate the complex process of hiring EC2 instances. As a *Python* based program it has option to access by an external script to schedule the on-demand hiring procedure. Moreover, it provides facility for on-spot bid based hiring of AWS-EC2, which is an economical option (50-70 % reduction in hourly charges compared to other reserved instance categories) for hiring. Using the *Starcluster* functionality to stop the instances, the cluster computer can be discharged from service, thus reducing the hourly charges based on completion of the simulation. Amazon Machine Image (AMI) developed with pre-build WRF-CHEM environment is an added advantage for on-demand hiring and utilization. Web application based on this can be used as an online WRF-CHEM modeling environment for educational and research purposes. Current facilities for WRF and other atmospheric modeling teaching and learning [241] can be carried out using these applications without the requirement of specialized software installation and HPC access.

Particulate pollutants $PM_{2.5}$ and PM_{10} in Coimbatore urban and surrounding region were simulated and evaluated based on six days of WRF-CHEM modeling. The WRF-CHEM modeling and investigation on using 1 km horizontal resolution is first of its kind applied for second tier cities in India and Coimbatore air pollution problem. Model performance and criteria goals for $PM_{2.5}$ and PM_{10} are well established [242, 243]. The goals categorized the level of simulation performance required to be achieved by the model to be applied for regulatory or other applications. According to the set performance and criteria goals, statistical performance measures such as MFB has to be in the range of less than 30 % (performance goals) and 60%, and MFE less than 50% (performance goals) and 75% (for criteria goals) to consider that the simulation has achieved the goals. In the current study, model simulation outputs evaluated with observation and assessed by MFB and MFE ranges were above the range set for performance and criteria goals. It indicate requirement of further improvement in the model performance. As per the current study, the model improvements can be approached in terms of emission inventory and model parameterizations used in the simulation and observation data for

model evaluation. The emission inventory is an important input for WRF-CHEM simulation. In the current study high-resolution emission inventory specific to Coimbatore region was developed and used in the simulation. The effectiveness of the emission inventory is assessed by comparing the output of the simulation (using Coimbatore EI) with that obtained from using global level EI. Even though the statistical measures are almost similar between the two outputs, it is observed that the spatial distribution and range of pollutant concentration is more similar with observation while using Coimbatore EI than using global EI, indicating the effectiveness of the locally developed EI for Coimbatore. Better agreement between meteorological simulation output and observation indicates the effectiveness of meteorological parameterizations scheme adopted for the simulation. However, the chemical parameterizations scheme selected has to be further scrutinized specifically with respect to Coimbatore. Improved understanding of the influence of parameterization scheme especially regarding Chemical settings has to be subjected to regional level sensitivity analysis to improve further the model performance. Model performance evaluation is usually carried out using more than two observation stations. In the current study due to non-availability of data from real-time pollution monitoring and hurdles in availing manual data (because of non-sharing by the concerned authorities), the evaluation is restricted to using one station data. Most evaluation studies also use the vertical profile of the atmosphere [192, 195]. However, in the present case the model performance evaluation is restricted due to the lack of related observations. It is imperative to have an accessible, accurate and updated data portal on atmospheric observation for performance evaluation of the model output. Despite being a fast growing urban centre and the second largest city of Tamil Nadu, Coimbatore lacks in such vital data and information resources. In this regard, extending the model domain to nearby cities such as Bangalore, Kochi or Chennai with real-time air pollution monitoring stations would help improving the evaluation of the model.

5.5 Conclusion

1. Current study addressed the technical acquirement of real time WRF-CHEM simulation of particulate pollutants in Coimbatore region. Automatic execution script, cloud based

computing resource for real-time execution and model performance with respect to emission inventory for Coimbatore region was evaluated

2. It is found that the *Python* based programming scripts are useful in automatic real time execution of WRF-CHEM
3. The utility of cloud based cluster computer for real-time execution of WRF-CHEM is also demonstrated in the study
4. In terms of WRF-CHEM performance for Coimbatore region, good agreement between observation and simulation for meteorological variables such as temperature, relative humidity was found; but it is below the required model performance and criteria goals for $PM_{2.5}$ and PM_{10} and indicates further improvement
5. The accuracy assessment of emission inventory shows the Coimbatore emission inventory compiled in current study is better while comparing with global level emission inventory. The Coimbatore emission inventory reflected the spatial variability and concentration well in the range similar to the field observations

Chapter 6

Data-Interoperability measures

6.1 Introduction

Management of air pollution requires accurate assessment of risks, prioritizing localities of high risks, and identifying and implementing mitigation steps. It comprised of a chain of processes starting from spatio-temporally relevant data collection, sharing data among regulatory authorities and concerned public, implementing and innovating interventions to reduce pollution, assessing the effectiveness of measures and generating public consensus for sustained interventions [244, 245, 69, 246]. That would also involve increasing reliance upon heterogeneous information sources such as real time data from wide type and array of sensor resources [116], atmospheric models [191, 247] and voluntary information [93, 127], many of which are currently not interoperable. Non-interoperability of these information sources is severely impeding the process of information integration to generate actionable knowledge and intervention measures [248] currently. In Information technology, interoperability is defined as “capability of two or more functional units to process data cooperatively”[249]. It is a necessary attribute of data, especially in multidisciplinary fields such as air pollution studies and management. Interoperability has crucial role in data discovery, open access, utilization and reuse [248]. Interoperability could be in four levels namely semantic, system, structure and syntax interoperability as per the heterogeneity of the information being accessed [250].

Semantic interoperability ensures the common understanding of the meanings of the exchanged information through metadata specifications. System interoperability addresses the intrinsic differences in hardware, software and communication components in a system. Structure interoperability deals with the compliance to uniform models to describe the data. Syntax interoperability addresses the data representation for automatic machine readability.

Environmental regulatory bodies such as national or state level Pollution Control Boards monitor real time air pollution in major urban centres of India [251]. However, the data collected and disseminated under these programs are not enabled to comply with the above said interoperability issues. This makes the data not open access, machine-readable without any metadata, and application programming interface to enable the capability to seamlessly process data [248]. These limitations hinder integration of these data with air quality modeling system and other volunteer information systems to address data gaps prevailing presently in air pollution management. Such constraints further results in inadequate dissemination of information for decision-making or for public awareness. Furthermore, the problem leads to limiting the scrutiny and free flow of information required by scientific knowledge system to address a complex environmental issue such as air pollution [252]. Data interoperability is enabled in two characteristic ways. Either through compliance with open standards [248] or by following most widely used data formats and conventions [253]. For example, in atmospheric and oceanic scientific research fields, data are often represented in grid forms for observations and simulation outputs. *Netcdf* is a software widely used for the purpose, to create, access and share 2D and 3D gridded data [253]. Its wide usage and applications made the *Netcdf* output data form a de-facto community developed data standard. On the contrary, there are stipulated open standards as in meteorology by World Meteorological Organization and in Climate science by Climate and Forecast Metadata Conventions to enable data interoperability [253].

Sensor Web Enablement (SWE) specification is a set of open standards developed by Open Geospatial Consortium (OGC) to discover, access, control, and process data from sensor networks in common internet based web interfaces [254]. Real-time air pollution monitoring

data has characteristics of high frequency updates, heterogeneous sensor origin and diverse background information in data collection methodology, geographic locality, and quality that makes it essential to adopt interoperability measures. Sensor Observation Service (SOS) [255] is an important component of SWE that is aimed to enable interoperability in sensor observations and its web based dissemination on real-time basis. SOS has conventions to impart various levels of interoperability, such as to address the typical time series data management for pollution monitoring and to effectively disseminate the data through a common machine readable Application Programming Interface (API). In the case of real-time air quality modeling, the model determines data format, its temporal and spatial representation, and data validity. This is especially for Numerical Weather Prediction based models such as WRF-CHEM used in the current study. Web based dissemination and processing is based on serial data. It requires model output to be converted into serial form from binary Netcdf format. Due to unique nature of model output the interoperability for environmental models are suggested through a set of web services oriented GEO model web initiatives [256], which are yet to gain wide utilization. On the other hand, OGC standard based model output interoperability measures for web application is being proposed recently taking ‘Model as sensors’ [257] that views the process of simulation, output, validation and visualization as de novo monitoring data. It uses Sensor Observation Service (SOS) to define variables and query the temporal and spatial relevant data from the model output. In this context, the present study explored the methods and advantages of implementing SOS for air pollution data collected from real-time particulate monitoring and modeling. It also demonstrates the advantages of interoperability measures of real-time data.

6.2 Methodology

6.2.1 Sensor Observation Service: *IstSOS*

Advancement in information technology has provided necessary tools to integrate and organize data accessibility, which ease web based dissemination of real time data. However, magnitude and agility in time and spatial bound data generation from real time monitoring or

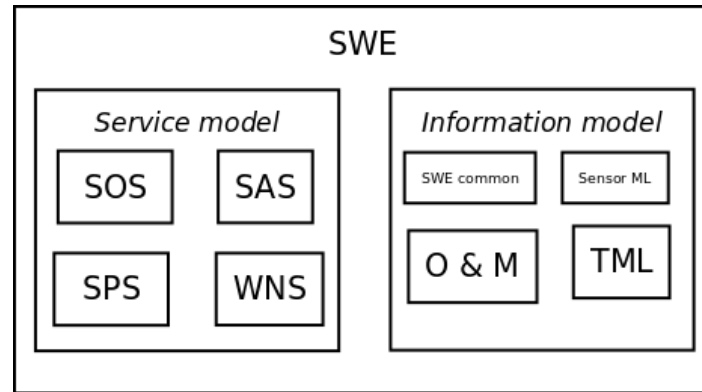


Fig. 6.1 Components of Sensor Web Enablement(SWE) specification

modeling data requires interoperable management [74]. Open Geospatial Consortium (OGC) is a consortium of more than 300 individual industrial and research organizations, which works towards interoperable data management in geospatial sector. Sensor Web Enablement (SWE) specification is an ongoing initiative of OGC to address the challenges of environmental sensor network data management. SWE specification comprised of several standards to define and control sensor resources, encode and publish the data in standard web form for Internet dissemination. SWE is broadly divided into information and service models based on the standards and specifications. SOS is an important part of the service model and the Figure 6.1 displays the components of the same.

Besides the SOS, the service model comprised of a Sensor Alert Service (SAS) that sends notifications (such as current state of sensor) from sensor resources. Sensor Planning Service (SPS) controls the access to sensor resources to manage their field deployment and observation collection routines. Web Notification Service (WNS) gives standard message interactions among / across the various service modules in SWE. Information model of SWE comprised of common set of encoding that describe the O&M (Observation and Measurement), the process and procedure operating in sensor resources to collect data on the environmental phenomenon by SensroML (Sensor Model Language) and communication function in sensor resource through Transducer and Model Language (TML). Setting of SOS in SWE extends its functionality to manage the network of sensor resources in an interoperable way. SOS achieves

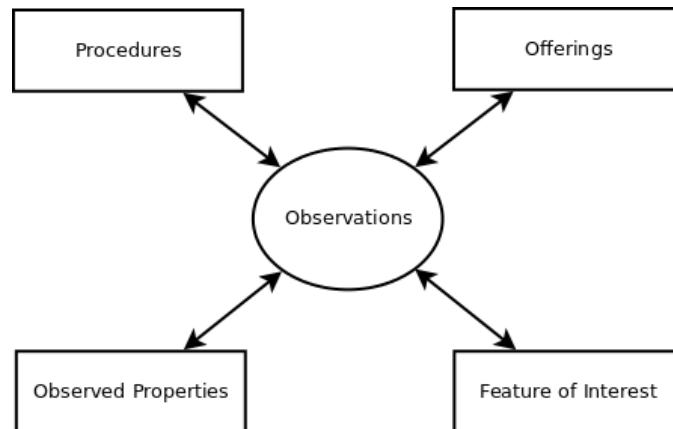


Fig. 6.2 Key objects of Sensor Observation Services(SOS)

it through a set of objects such as *procedures*, *offerings*, *observed properties* and *feature of interest* (Figure 6.2).

Procedure defines the sensor /method /device that perform the observation. *Offerings* define time period of observation supplied to SOS. *Observed properties* names set of parameters included in observations. *Feature of interest* defines the location features relating to the observed phenomena as geographical locality of the sensor resource. The SOS specification defined by OGC is developed into software programs for use in real world application. *IstSOS* [258] and 52°North *SOS* [259] are two widely used software programs that implement the SOS. In the current study, *IstSOS* was used due to its *Python* programming language based implementation and for being relatively simple and easy to use. The free and open source *IstSOS* program is developed by the Institute of Earth Sciences, University of Applied Sciences and Arts of Southern Switzerland. It is primarily developed for use in hydro-meteorological monitoring network for their flood risk management system. The software architecture comprised of three nested components namely *Istlib*, *walib* and *wainterface*. The *Istlib* acts as the innermost component that supports service overseeing the SOS standard implementation, *walib* is component of *wa* services that exposes / displays the web based management features, and *wainterface* oversee the interaction of former two components. The server side is programmed in *Python* [260] and the client side is programmed in *JavaScript* [261] language. *PostgreSQL* [262] database, enabled with *PostGIS* for geospatial data storage and retrieval, is used in the software *IstSOS*.

6.2.2 SOS for real time monitoring

Data from real-time monitoring (using the sensor) of particulate matter was imported into SOS system. Particulate pollution data from stationary monitoring, carried out in different parts of Coimbatore urban area during 2014-2015, was used for SOS enablement (See figure 3.3). The monitoring data is communicated to the central server (IBM-M4 server with Linux Ubuntu OS, maintained in the laboratory) in real time at a time interval of every 15 minutes. The communications were through SMS or mobile internet based HTTP transfer according to the location of the monitor station. Auxiliary scripts and programmatic request format available in *IstSOS* was used to define the sensor resources and import particulate pollution reading ($PM_{2.5}$ and PM_{10} size classes) into *IstSOS* system.

6.2.3 SOS for real time modeling

WRF-CHEM output data in *Netcdf* format is in binary form. For every one hour of simulation, WRF-CHEM generates one *Netcdf* file comprised of meteorological and chemistry variable outputs. The variables such as Temperature, Relative Humidity, $PM_{2.5}$ and PM_{10} were extracted from the hourly *Netcdf* file and were used for SOS enablement. The output from the simulation run carried out for assessing resource evaluation was used for the SOS as well. The variables from the hourly *Netcdf* file of the third finest grid with resolution of 3 Km were extracted for use in SOS. To reduce the number of grid points to be included in the SOS implementation, only the point of interest grid points located nearest to the real-time particulate pollution monitoring station was used. Each grid points were marked and the time series data for the variables were imported into the SOS system taking the grid point as sensor location. *Python* programming language based scripts using the *Netcdf* library was used to access the WRF-CHEM output and import the time series of required variable into the SOS system. The scripts and the programmatic request form used for monitoring data were used for modeling also.

6.2.4 Advantages of Sensor Observation Service

Main advantages of SOS implementation is its functionality to provide information regarding the sensor resources and observation data in Representation State Transfer Application Programming Interface (REST-API). This functionality enables to query and retrieve the data in widely recognized web format such as Java Script Object Notation (JSON). This format is a common data form optimized with web browser readability and exchange. Dynamic web applications were used to evaluate and demonstrate these interoperability advantages of SOS. The web applications dynamically querying key objects of SOS such as *procedures*, *offerings*, *observed properties* and *feature of interest* were used. Using R statistical program, platform independent interoperable querying of data were demonstrated.

6.3 Results

6.3.1 SOS for real time monitoring and modeling data

IstSOS, an open source *Python* programming language based SOS program, was used in the current study. The real time particulate pollution monitoring and modeling data from Coimbatore region was imported into *IstSOS* program. Various technical steps and characteristics of SOS implementation are discussed below in terms of four interoperability levels (discussed earlier) intended to achieve by importing the data into *IstSOS* program.

Semantic interoperability

Semantic interoperability is achieved by enabling meaning and background information on sensor resource and its observation data. It is done by encoding metadata specifications to describe background details of the sensor resources, characteristics of the sensor, and its method of data collection and observation units. *IstSOS* gives two modes of facility to describe the metadata of sensor resources imported into it; First mode is by Graphical User Interface (GUI) and second mode by HTTP request. In GUI, the metadata details are provided in server level and service level. The service provider and service identification forms are provided in

Coimbatore air | IstSOS webadmin develop... | localhost/istsos/admin/

Server | Services | Data viewer

IstSOS

About IstSOS | Status | Database | Service provider | Service identification | Coordinates system | Get Observation Configuration | Proxy Configuration | New service | Delete service

Server Status

Summary of istSOS instances and run-time status

Service	Features Of Int	Offerings	Procedures	Observed Prop	Availability	Database	GetCapabilities	DescribeSensor	GetObservation	GetFeatureOfInt	InsertObservation	RegisterSensor
awscbe	0	1	0	0	up	active	✓	✓	✓	✓	✓	✓
tnau_cbe	13	2	14	10	up	active	✓	✓	✓	✓	✓	✓
cosmcsm	0	0	0	4	up	active	✓	✓	✓	✓	✓	✓
dyloscbe	1	1	1	4	up	active	✓	✓	✓	✓	✓	✓
demo	0	0	0	0	up	active	✓	✓	✓	✓	✓	✓
cbed	2	1	2	2	up	active	✓	✓	✓	✓	✓	✓
tnau	3	2	4	4	up	active	✓	✓	✓	✓	✓	✓

Open Source Software by Institute of Earth Science - SUPSI

Fig. 6.3 Server status page of *IstSOS*

server level metadata, which is useful keywords for search engines. Moreover, this facility also provided function for overall management of *IstSOS* program. It has user interface for entering the details of the database, geographical coordinate system followed, and managing the REST-API query response. The REST-API response is managed with 'GetObservation' configuration of SOS that controls the number of API responses per second and exception responses provided by the SOS. The Figure 6.3 shows the server status page of *IstSOS* that provides links to forms for entering the details. *IstSOS* provides facility to host and manage multiple sensor resources, grouped as separate services. The facility to enter metadata details for each service is provided in a separate service page of *IstSOS* GUI. The Figure 6.4 shows the service page. The service page gives facility to describe the units of measurement and various properties observed by the sensor resources. As per the SOS specifications, complete details about the sensor resources are stored and grouped as *Procedure*. In the *Procedure* page of the *IstSOS* GUI, details such as general info about individual sensor, its location, type of sensor system (whether it is a in-situ fixed point or mobile sensor) and various output it generates are stored. The *Procedure* output is with *observed properties* and measurement units. The service interface also gives facility

Fig. 6.4 Service description page of *IstSOS*

to enter metadata details about service provider and service identification key specific to the sensor resource.

In HTTP request mode, *IstSOS* gives means to upload the details about sensor resources through programmatic path. The following shows the encoded *Procedure* in JSON format.

```
{
  "system_id": "1491",
  "system": "1491",
  "description": "1491",
  "keywords": "1491",
  "identification": [
    {
      "name": "uniqueID",
      "definition": "urn:ogc:def:identifier:OGC:uniqueID",
      "value": "urn:ogc:def:procedure:x-istsos:1.0:1491"
    }
  ],
  "classification": [
    {
      "name": "System Type",
      "definition": "urn:ogc:def:classifier:x-istsos:1.0:systemType",

```

```

        "value": "insitu-fixed-point"
    },
    {
        "name": "Sensor Type",
        "definition": "urn:ogc:def:classifier:x-istsos:1.0:sensorType",
        "value": "Active particlecounter"
    }
],
....
"location": {
    "type": "Feature",
    "geometry": {
        "type": "Point",
        "coordinates": [
            "78.937549",
            "12.994631",
            "300"
        ]
    },
    ....
    {
        "name": "particlecountofSizeclassPM2.5",
        "definition": "urn:ogc:def:parameter:x-istsos:1.0:cbedylos:pm25",
        "uom": "ug/m3",
        "description": "",
        "constraint": {

        }
    },
}

```

The above *Procedure*, a JSON file, was edited to store multiple sensor resources in *IstSOS*. A *Python*¹¹ based code was used to invoke an HTTP request for storing a new *Procedure* or multiple *Procedures*. The *Python* code edit the common procedure details encoded in JSON file with multiple sensor resources. The details about multiple sensor resources are stored in a Comma Separated Value (CSV) file and using *for-loop* function of *Python*, the JSON file was edited and HTTP request passed onto the *IstSOS* service. In this way, large number of multiple sensors details could be imported into *IstSOS* without following GUI which is time consuming and laborious.

¹¹File named *IstSosHttpReq.py* in Nishadh et al. [131]

System interoperability

System interoperability addresses variability in sensor resources such as type of sensors used in the network, and its mode of collection of data and communication. In the current study, four identical particulate pollution monitors were used, but with variable mode of real-time data communication either as SMS or HTTP request. As the monitors are deployed in different parts of Coimbatore, the data communication was chosen based on the signal reception of GSM modem at each location. In urban location where the GSM signal was good for Internet communication HTTP data communication was used, whereas in semi-urban and village locations where GSM signal was not adequately reliable SMS mode of communication was used. In the case of modeling data, the point of interest time series was collected from the cluster computers used for real time WRF-CHEM simulation and transferred to the central server where the *IstSOS* program was installed and configured. The Figure 6.5 shows the processes involved in importing the real-time data into SOS for achieving the system interoperability. In the central server, the collected data was converted into CSV format. The CSV file was then imported into *IstSOS* server using *csv2istsos.py*, a *Python* based script provided with *IstSOS*. The script requires the CSV data in specific format, bearing the details of sensor procedure and its Uniform Resource Identifier (URI). The *istSOS* also provides facility for the HTTP request based importing of data.

Structure interoperability

Structure interoperability deals with common encodings for real time data and its metadata. In SOS, it is defined as information models, which is a component of OGC Sensor Web Enablement specification. The Figure 6.1 shows the sub-components of information models. Extensible Markup Language (XML) was used for defining the common encoding. Following is an example of XML definition of insert observation routine in SOS.

```
<?xml version="1.0" encoding="UTF-8"?>
<sos:InsertObservation xmlns:xsi="http://www.w3.org/2001/..." SOS" version="1.0.0">
  <AssignedSensorId>urn:ogc:object:sensor:x-ist::??</AssignedSensorId>
  <om:Observation>
    <om:procedure xlink:href="urn:ogc:def:procedure:x-istsos:1.0:KNMR"/>
    <om:samplingTime>
```

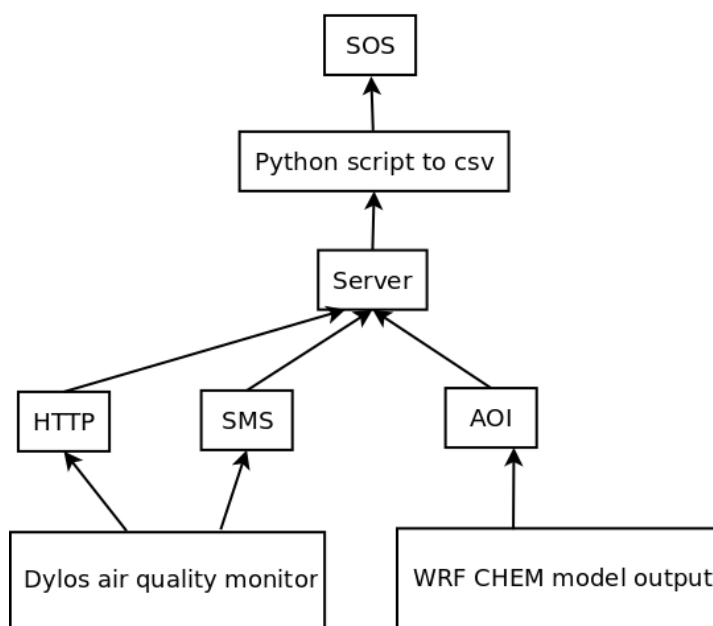


Fig. 6.5 Various process involved in importing data into *IstSOS*

```

<gml:TimePeriod>
  <gml:beginPosition>2014-07-29T22:35:03+05:30</gml:beginPosition>
  <gml:endPosition>2014-07-0T29:35:03+05:30</gml:endPosition>
</gml:TimePeriod>
</om:samplingTime>
<om:observedProperty>
  <swe:CompositPhenomenon dimension="3">
    <swe:component xlink:href="urn:ogc:def:parameter:x-istsos:1.0:time:iso8601"/>
    <swe:component xlink:href="urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np05mg"/>
    <swe:component xlink:href="urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np25mg"/>
  </swe:CompositPhenomenon>
</om:observedProperty>
<om:featureOfInterest xlink:href="urn:ogc:def:feature:x-istsos:1.0:Point:KNMR"/>
<om:result>
  <swe:DataArray>
    <swe:elementCount>
      <swe:Count>
        <swe:value>2</swe:value>
      </swe:Count>
    </swe:elementCount>
    <swe:elementType name="SimpleDataArray">
      <swe:DataRecord definition="http://mmiws.org/ont/x/timeSeries">
        <swe:field name="Time">
          <swe:Time definition="urn:ogc:def:parameter:x-istsos:1.0:time:iso8601"/>
        </swe:field>
      </swe:DataRecord>
    </swe:elementType>
  </swe:DataArray>
</om:result>

```

```

        </swe:field>
        <swe:field name="NoPG0.5">
            <swe:Quantity definition="urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np05mg">
                <swe:uom code="air -pm05"/>
            </swe:Quantity>
        </swe:field>
        <swe:field name="NoPG2.5">
            <swe:Quantity definition="urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np25mg">
                <swe:uom code="air -pm25"/>
            </swe:Quantity>
        </swe:field>
    </swe:DataRecord>
</swe:elementType>
<swe:encoding>
    <swe:TextBlock tokenSeparator="," blockSeparator="@" decimalSeparator="."/>
</swe:encoding>
<swe:values>
    2014-07-29T22:35:03+05:30,5197,270@
</swe:values>
</swe:DataArray>
</om:result>
</om:Observation>
</sos:InsertObservation>

```

As a part of OGC web services common specification, SOS server level data encoding is carried out by operations such as ‘Getcapabilities’ and exception reports on the current status of the server with respect to various functionalities. For service level sensor resource encodings, information model component *Observation and Measurement* is used. It is comprised of observation schema and sampling features. The observation schema describes *observed properties*, *measurement* and *collection*. The sampling features describe spatial summary of the sensor resource. To describe data in sensor observation services, *SWE common* is used. That supports data-frame to hold actual observed values with associated metadata by supporting a matrix like data structure.

Syntax interoperability

Software platform independent automatic machine readability is indicated by syntax interoperability. This is achieved by provisioning the data in automatically recognizable form

through REST-API. *IstSOS* provides a set of HTTP request methods such as GET and POST to query or import the data from *IstSOS*. The common data form used in *IstSOS* is JSON to encode the response. The following shows JSON response of ‘GetObservation’ from *IstSOS*. Java Script Object Notation (JSON), an open standard format, is of lightweight (small size) and language independent. It is widely used in dynamic web application as web service and REST-API response. The request and response (shown above in the XML schema of information models under structure interoperability) are provided in JSON format.

```
{
  "message": "GetObservation requested successfully executed",
  "total": 1,
  "data": [
    {
      "name": "TDM-CBE",
      "samplingTime": {
        "duration": "P30DT22H20M",
        "beginPosition": "2014-03-03T20:00:00+05:30",
        "endPosition": "2014-04-03T18:20:00+05:30"
      },
      "result": {
        "DataArray": {
          "elementCount": "9",
          "values": [
            [
              "2014-03-27T00:10:00+05:30",
              "3107.000000",
              "200",
              "170.000000",
              "200"
            ],
            [
              "2014-03-27T00:10:05+05:30",
              "5597.000000",
              "200",
              "320.000000",
              "200"
            ],
            [
              "2014-04-03T18:20:00+05:30",
              "6764.000000",

```



```

        "200",
        "389.000000",
        "200"
    ]
  ],
  "field": [
    {
      "definition": "urn:ogc:def:parameter:x-istsos:1.0:time:iso8601",
      "name": "Time"
    },
    {
      "definition": "urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np05mg",
      "name": "air-pm05",
      "uom": "ug/m3"
    },
    {
      "definition": "urn:ogc:def:parameter:x-istsos:1.0:cbe...:qualityIndex",
      "name": "air-pm05:qualityIndex",
      "uom": "-"
    },
    {
      "definition": "urn:ogc:def:parameter:x-istsos:1.0:cbe:dylos:np25mg",
      "name": "air-pm25",
      "uom": "ug/m3"
    },
    {
      "definition": "urn:ogc:def:parameter:x-istsos:1.0:...:qualityIndex",
      "name": "air-pm25:qualityIndex",
      "uom": "-"
    }
  ]
}

},
"featureOfInterest": {
  "geom": "<gml:Point srsName='EPSG:4326'...>",
  "name": "urn:ogc:def:feature:x-istsos:1.0:Point:TDM-CBE"
},
"observedProperty": {
  "component": [
    "urn:ogc:def:parameter:x-istsos:1.0:time:iso8601",
    "urn:ogc:def:parameter:x-istsos:1.0:meteo:air:pm05",
    "urn:ogc:def:parameter:x-istsos:1.0:meteo:air:pm25"
  ]
}

```

```

    ],
    "CompositePhenomenon": {
        "dimension": "9",
        "id": "comp_4",
        "name": "timeSeriesOfObservations"
    }
},
"procedure": "urn:ogc:def:procedure:x-istsos:1.0:BELLINZONA"
}
],
"success": true }

```

6.3.2 Advantages of Sensor Observation Service (SOS)

Advantages of sensor observation service provided are demonstrated by a dynamic web application and on-line statistical analysis routine. The software platform-independent and interoperable advantages of machine readable REST-API provided by SOS is demonstrated in those applications discussed below.

Dynamic web applications

Wide and time bound dissemination of information is an important aspect in real time air pollution monitoring and modeling. In this internet and smart phone era, dynamic web applications are vital in disseminating information. Conventionally, they are backed with its own custom database addressing the requirements of specific application. This hinders replicating the data sources for various information disseminating platforms such as social media and online news portals. REST-API provides interoperable data to be reused in different contexts. This is carried out by querying the REST-API by web applications. In that situation, the work of the information managers is considerably reduced and that would include only sustaining web applications' functionality. REST-API caters the needs of platforms such as social media, online news aggregation, and blog or web page tools which can show the real time pollution level. This is achieved by REST-API that providing functionality for multiple queries from different application sources. Basically dynamic web application provides programs to query the REST-API and visualize the data in required forms for different platforms.

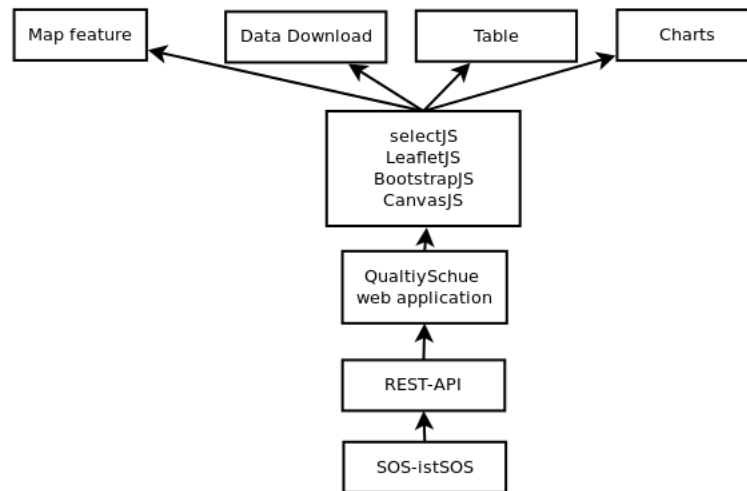


Fig. 6.6 Components and its interaction in web map application based on SOS

The present study demonstrates a web application that queries REST-API from SOS of real time air pollution monitoring and modeling carried out in Coimbatore region. The dynamic web application is an open source program named *qualitySCHUE* [263]. The application comprised of a web map of Coimbatore region depicting the location of sensor resources and model output for point of interest (location). The location mark provides for selecting specific location of interest and menu links for viewing the SOS data source in chart and table form. Various components and processes involved in the web application are shown in the Figure 6.6. *JavaScript*, a web browser based programming language widely used in modern dynamic web application, was used for the web application. Mainly two *JavaScript* files perform major functions of the web application. The *JavaScript* file *map.js* carries out the parsing of the SOS, queries the data based on the location, time and sensor observation parameters. That also oversees the mapping of the SOS sensor resources in web map page. The second *JavaScript* file *menu.js* oversees the REST-API call from the SOS instances with chosen parameters. The API call is a simple URL command as follows.

```

http://localhost/istsos/wa/istsos/services/+/selectedServiceToDownload+/operations/getobservation
/offerings/temporary/procedures/+/selectedProceduresToDownload+/observedproperties/
+selectedObservedProperties+/eventtime/+/selectedStartdateToDownload+/+/selectedEnddateToDownload+;
  
```

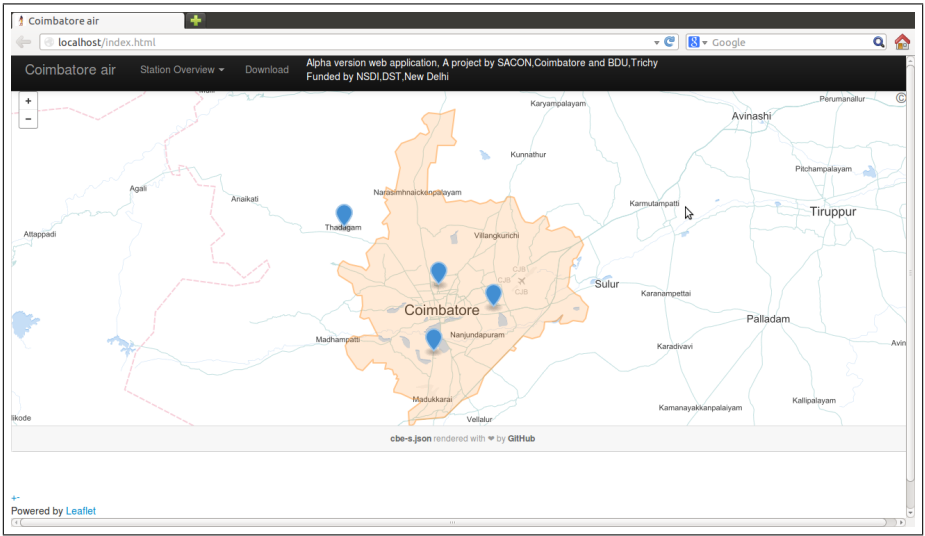
The Figure 6.7 below shows various pages available in the web application. The URL locates the data and the web application parses the SOS instance and downloads the required data in

interoperable format. The format used is Java Script Object Notation (JSON). This format is used for plotting the chart and the table. In background along with the above described *JavaScript* files, there are several libraries (*selectJS*, *LeafletJS*, *BootstrapJS*, *CanvasJS*) providing functions such as web mapping, selecting the SOS services, deciding the date range, and dynamically plotting the chart or table of SOS output in the web applications.

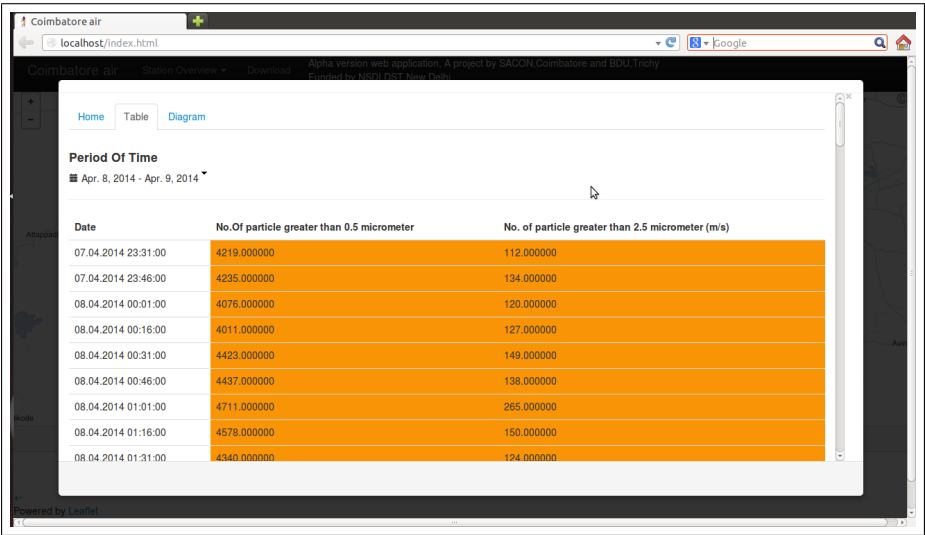
On-line statistical analysis

Statistically analysing and eliciting the hidden pattern in the real time pollution data is an important aspect in addressing air pollution problems. Conventionally statistical analysis was carried out separately by downloading and importing the data into specialized software and applying various statistical procedures on the data. Considering the dynamic nature and size of the data collected in real-time, statistical inference on the data has to be done *on the fly* to avoid prolonged information gaps. Quicker analysis can considerably improve the timing of the inferences, pollution management options and predictions of the potential risks from dangerous pollution levels. In this regard, REST-API provided by SOS can be an appropriate means for on-line statistical analysis. As in the case of dynamic web applications, the SOS query into statistical platform provides advantages in terms of *on the fly* and automatic execution of Statistical computing. In this study, open source *R* programming language was used for on-line statistical analysis. In the present study, SOS4R [264], an extension library written in *R* programming language, was used for connecting the SOS instance with *R* statistical platform. The SOS4R provides programming means to query the SOS-REST-API, collect the data in structured format and carry out appropriate statistical analysis. In the current study, the application of model output evaluation was done using REST-API. Interoperable real-time monitoring and modeling was queried from SOS and was compared for model evaluation. *Openair* [209], an extension library written in *R* programming language focusing the air pollution statistical analysis was used for model evaluation statistics. The process of integrating the SOS with *R* is carried out in four steps.

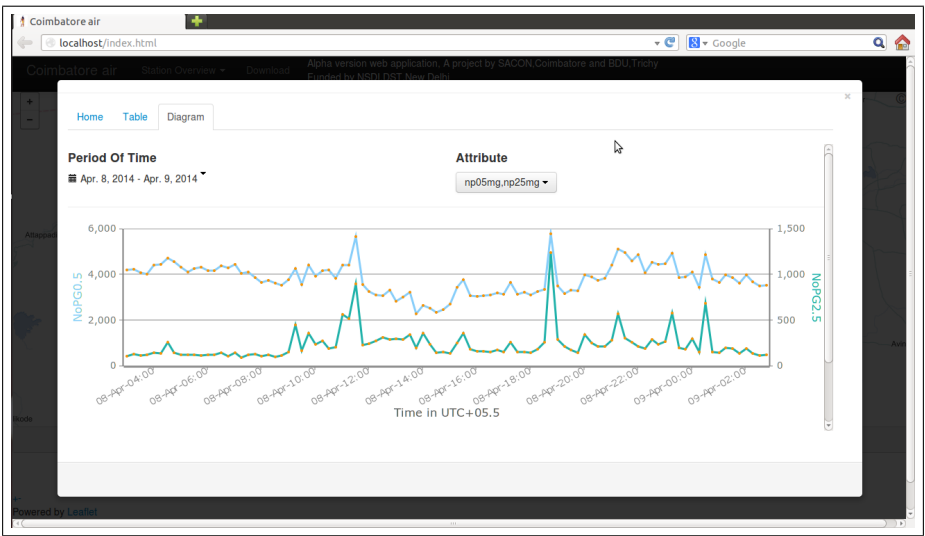
1. Connecting the SOS with *R* programming platform



(a) Web map interface



(b) Data in table form



(c) Data in Chart form

Fig. 6.7 Interfaces in web application based on SOS

2. Supplying the query parameters, which includes the type of SOS service, its *offerings*, *observed properties*, and time period of the data requested
3. Using the first and second steps, collect the structured data for model evaluation from monitoring stations and point of interest
4. Apply the model evaluation statistics on the structured data collected from SOS

R script¹² was written in this respect based on Nüst et al. [264], which performs the above-described steps. Technically, the steps involve generating a REST-API query request with all data requirements. Convert this query as an encoded form such as URL requests to interact with the SOS instances. Use the encoded request to query the SOS. Subsequently, process the response of SOS query that comprised of data and metadata of the service, and format the data for further analysis. This processes of integrating the SOS with statistical programming language such as *R* shows the advantages of interoperability measures for real time data. Moreover, this opens up a wide of array of statistical computing tool, which are required for making rational inferences from the data.

6.4 Discussion

This study explored methods to make real time particulate pollution data widely usable in web platform. Essentiality of seamlessly bringing together data from different sources and effective utilization and distribution of the same were addressed in the study. Data interoperability, an important enabler, was carried out by SOS. The study used SOS for implementing interoperability measures for monitoring and modeling data. Following the ‘Model as sensor’ concept [257], it used SOS to describe the model output. However, this approach has limitations in terms of data storage and describing background information. The environmental model outputs especially in the case of WRF-CHEM are available in gridded format covering a spatial area pre-defined during simulation. Large data array produced in simulation are provided as binary format such as *Netcdf* form. This format has limitation in

¹²File named *SosAdvantagesExEval.R* in Nishadh et al. [131]

web based visualization and time series data dissemination. The current study addressed this limitation by converting the binary format into serial format. Only the point-of-interest grids are selected from the simulation output grids and imported into SOS, neglecting non-chosen grids; thus restricting the user to pre-selected point-of-interest time series data from model grid. Moreover, the SOS has limitations in describing the metadata information regarding the selected model parametrization scheme and chosen grid . Developing a web service based platform for model output can be a solution for such limitations. That platform can be designed with facility for user requested query on time series model data in SOS specification. The extended API facility provided by *Netcdf* [265, 266] and programming language to access *Netcdf* such as *nco* [212] can be used in dynamic web application background. The studies on environmental models as web services [256, 267, 268] or OGC defined sensor web enablement specification such as web processing service [269] are providing important facilitating option in this aspect. Automated means [270] of generating metadata can be followed and provided along with the user query response. This can be used for providing the model specific metadata on the time series data provided in SOS specification.

Other than air pollution monitoring and modeling data, volunteer-provided data is an important source [271] for addressing gaps in air pollution data. Current developments in computing and smart phone technology [272] offer advanced tools for volunteer community to collect and share air pollution data [125, 133]. Low cost air quality sensors [273, 274] and image based data as proxy information on pollution level enables common citizens to share such data with geographical information [275]. The recent initiatives on volunteer based data collection on air pollution is concentrating on sensor developments and focusing on data collection in real-time [276]. However, there are issues to be addressed with such data, such as data quality, before to use. In the current study, SOS implementation with extended steps in registering the sensors and data import largely hinders wide usability of SOS system in volunteer information initiatives. Internet of Things (IOT) based platform which are greatly simplifying real time data streaming from common sensors and consumer electronic devices are used for such initiatives [277]. Free IOT platforms are providing simple HTTP based REST-API to import the data without any metadata data and SOS specification. The developments in

Open Geospatial Consortium (OGC) specification [278] are attempting to find solutions for such shortcomings. OGC Sensor things API Standard is one that can be used for involving the volunteer information with SOS [279].

Studies on air pollution have shown the need for wider distribution of related information and air quality indices in social-media [280–286]. REST-API provided by SOS has important technological capability for wide distribution of air pollution data. It helps in reuse and visualizing the data in different dynamic web platforms such as web pages of press and social media. In India, recent reports of National Air Quality Index highlight the need for disseminating it in web-based platforms. The API form of SOS output would be helpful in real time AQI calculation and displaying the outputs in various web-mapping applications. Various studies have concluded the need of increasing the quantity and quality of air pollution data for improving the public information on air pollution [287, 285]. The data interoperability provided by SOS can improve the spread, quality and quantity of information by collating the data from various monitoring and modeling routines as documented in the current study. It can also help in including volunteer data with further development in interoperability measures. It might be noted that the objective of air quality information has to be widened from a tool merely for generating awareness to a one that is engaging and capable of instigating behavioural change among the public [284].

Engaging a larger segment of people in web based tools to explore air pollution data, analyse and draw rational conclusions from it using appropriate easy to use statistical and modeling tools can be promoted by the interoperability of data demonstrated in the current study. Application programming Interface based on-line statistical analysis and data dissemination facility shown in the current study can be used for such web-based data exploration and analyses. The open source statistical program R used in the current study is enabled with such web-based tools [288] such as on-line data and statistical analysis enriched document generation tools [289], online regression exploration tools, and animated and dynamic graphs [290]. These types of tools are increasingly coming up as important interventions for engaging the users. That gives the highly wanted openness and capability to explore the data and prompt the public user to know about

air pollution causes and effects, and be fittingly pro-active. Furthermore, this openness in the data fosters much needed trust and involvement of public in understanding and combating air pollution and related issues. This characteristic of data provision from preservative of social and behavioural science research are increasingly observed as an important enabler for behavioural change in the public [291]. The advantages of interoperability measures discussed in the current study can be a start in developing public engaging services for air quality information in the country.

6.5 Conclusion

1. Methods and advantages of interoperability measures on real time particulate pollution data was explored and it is found to be important enabler to make the data widely usable in web platform
2. Open standard, Sensor Observation Service (SOS) and its *Python* based implementation software *IstSOS*, used in the current study shows its effectiveness in achieving various interoperability levels
3. SOS implementation is well suited for describing real time particulate pollution monitoring data but it has limitations in terms of describing and disseminating the large model output grids
4. The web application and R based platform-independent statistical analysis shows the advantages of interoperability measures for web based data dissemination

Chapter 7

Conclusion

The data intensive approach can provide necessary tools for addressing the particulate pollution problem in urban centres. The current study focused on developing basic infrastructure for data intensive approach in a second tier urban centre of India. The study addresses the development of particulate pollution monitoring system using low cost commodity sensor and assess its effectiveness in the study area. The study attempted a real-time particulate pollution modeling system using WRF-CHEM model and address its computational requirements. The particulate emission inventory for the Coimbatore region was developed for satisfying the data requirement for WRF-CHEM modeling. The advantages of interoperability measures applied for real time monitoring and modeling data was demonstrated by a web application and on-line statistical analysis work flow.

Particulate pollution and its resultant health effects are considered one of the large health problems across the world. Particulate pollutants (or suspended Particulate Matter - PM) are particles with small aerodynamic diameter. The size distribution of the PM is one crucial aspects of monitoring their environmental and health implications. Hence, PM is monitored primarily with objective of finding the size distribution and concentration. It is commonly monitored in terms of standard measures in two size bins of $PM_{2.5}$ and PM_{10} , by manual air filtration, optical methods or by continuous and integrated instruments determining mass concentration. However, high investment and running cost for automatic real-time instruments largely hinders

their wider deployment. Consequently, particle counters are being widely tried for their explicit low cost, ease of use and very short response time in measuring particle concentration in the ambient air. In this context, the current study aims evaluating the effectiveness of a customized particulate matter monitor for Indian situation. It is based on a low cost commodity monitor, fabricated with locally available open-source hardware, tools and expertise. Dylos™ model number DC1100 is a patented indoor air quality monitor commercialized by Dylos Corporation (USA). To customize Dylos™ into a real time ambient particulate monitor, it was equipped with data collection, storage and real-time data communication setup. For field deployment, the whole setup (monitor with data storage and communication) was housed in an all weather steel box, specially designed for the purpose. To evaluate the functionality of the developed / customized real time PM monitor, four units were deployed in various parts of Coimbatore city. To assess the validity of data from the units, collocation samplings were carried out concurrent with industrially calibrated portable particulate monitor MetOne™ Aerocet-531S. To assess the relationship / collinearity across the readings from collocated monitors, plots and adjusted R^2 values of linear regression between pair's of hourly average samples were used. Four locations in Coimbatore urban region were selected for ambient testing and data validation of the developed monitor. Cost-wise, PM_{10} monitor stands for around 70% of the total expenditure of the monitor customization. The customized monitor development, for the loosely coupled power source and single board computer, called for intermittent manual interventions. That suggests design overhaul replacing with embedded boards, and stable connectivity with battery power source. *Python* based scripts were used for real-time data communication. Ambient testing reflected general trend and stations' pollution profile, and the monitor reading satisfactorily exhibiting the pollution trend. The system effectiveness of real time particulate monitor was also assessed in terms of system persistence and data communication. It is observed that the monitor was effective in continuous operation for data collection, storage and real time communication and long-duration monitor operation. *Collocation* sampling was used for data validation and it is found that the developed monitor is correlating well with the industrially calibrated monitor. However, the variations in agreement

at different periods indicate the requirement of continuous and more robust calibration of the monitor.

Emission inventory (EI) is breakup of the quantity of air pollutants emitted from anthropogenic or natural sources; it quantifies various pollutants generated from wide-ranging emission causing activities within a stipulated geographical area and time span. EI development is an important and primary task in air pollution management. It is the key input data for air quality forecasting and in planning management interventions. Air quality models especially chemical transport and numerical weather prediction-based models require spatially allocated and gridded EI that provides initial condition values (seed values) for mathematical computation of spatio-temporal air pollution forecasts. Based on the availability and specificity of data, two approaches, top-down or bottom-up, are adopted for EI preparation. Top-down approach takes overall fuel consumption as surrogate for emission. Then the area level emissions are derived using related proxies for the emission activity and respective spatial segregates. On the other hand, bottom-up approach considers every pollutant emitting sectors in an area based on their respective spatial extent. Bottom up approach was used for local urban level EI preparation. Most of the past works and EIs in the country are limited in their utility, for using proprietary software or do not give out detailed methodology. That hinders their wider replication to generate EI for other urban locations and their use in air quality models. In view of the above, the current study attempted developing PM ($PM_{2.5}$ and PM_{10}) emission inventory for Coimbatore region through a bottom-up approach using free and open source programming tools. An area of 3000 km^2 surrounding Coimbatore Corporation and spread over the nearby villages were included while preparing current EI with a resolution of $1 \times 1\text{ km}$. Four major emission sources viz., sectors such as residential, industrial, transport and roadside windblown dust were considered for the inventory. Relevant data from governmental and non-governmental organization were used for emission source identification. Emission factor from the past studies were used considering the surrogate urban Indian condition. Spatial allocation of emission inventory was carried out using free and open source geospatial tools of the *Python* programming language and its libraries such as *Numpy*, *Pandas*, *Fiona*, *Geopandas*, *Shapely* and *Rtree* index. The study addressed localized data requirement and geospatial programming needs

for urban level EI development. Data in dispersed and non-reusable multi-data format was a challenging step in collating the data in a usable structure. *Python* based programmes were used for various geometry operation such as *interaction with-in*, measuring the proportionate of geometrical objects such as line, polygon etc. It is found that *Python* based programmes are self-contained in carrying out various operations and its extensive programs are promising to develop the regional data-backed stand-alone application. The EI shows that in Coimbatore region, Industrial sector is largest source of particulate pollutant followed by transport and residential sector. The spatial distribution of the emission shows that southern part of the geographic domain is host to largest industrial emitters. Furthermore, the present exercise of developing EI for Coimbatore region brings out the prevailing data gaps. The availability of data, its formats, validity and indifference from officialdom are constraints in EI development.

Significant health and environmental effect of particulate pollution can be managed by pollutant concentration forecasts. That would help in issuing early warnings / alerts and taking precautionary measures against pollutants level hikes. For this, Real Time - Air Quality Forecast systems (RT-AQF) are established in some major urban centres across the world. Numerical models are an important form of RT-AQF system. The models helps in predicting future state of physical and chemical constituents such as aerosols or other atmospheric species, their distributions and levels (the state of air quality in a particular time frame) in the context of prevalent atmospheric dynamics. Numerical models are complex and time taking. However, recent advances in scientific understanding and advent of high-power computational (HPC) capacity has considerably alleviated the difficulty of running numerical models. Consequently, these models are also gaining superior prediction accuracies and better reflection of the complexity of atmospheric processes in simulation. Weather Research and Forecast model with Chemistry (WRF-CHEM) is a widely used model in RT-AQF system. Being a typical numerical model, WRF-CHEM involves complex and numerous work flows and requires large computational power. This largely limits its wider usage and operational air quality modeling for second tier urban centres with less technical and financial resources. In this context, the current study was intended to addresses technical requirement for automatic real time execution of WRF-CHEM. To overcome the constraint of low access to HPC environment, we explored using

the advantages of on-line cloud computer services. WRF-CHEM model performance over Coimbatore region was assessed using 1 km resolution simulation and the output of the exercise was evaluated. The major components of the WRF-CHEM model are WRF Pre-Processing system (WPS), real data initialization and dynamical core solver advanced research WRF (ARW). The computational requirement for real time execution of WRF-CHEM simulation covering Coimbatore region was evaluated comparing desktop computers and cloud computer clusters. It was found that the cloud computer cluster Amazon™web Services Elastic Computer Cloud (AWS-EC2) takes the shortest time for one complete simulation. For evaluating the performance of WRF-CHEM simulation over Coimbatore region, a separate three-level nested domain with lowest resolution of 1 km was used. Particulate pollutants $PM_{2.5}$ and PM_{10} in Coimbatore urban and surrounding region were simulated and evaluated based on six days of WRF-CHEM modeling. In terms of WRF-CHEM performance for Coimbatore region, good agreement was found between observation and simulation for meteorological variables such as temperature, relative humidity; but it agreement was below the required model performance and criteria goals for $PM_{2.5}$ and PM_{10} . The accuracy assessment of EI shows the Coimbatore emission inventory compiled in current study is better while comparing with global level emission inventory. The Coimbatore emission inventory reflected the spatial variability and concentrations well in the range similar to the observation. The study warrants forecast improvements in terms of emission inventory and model parameterizations used in the simulation and observation data for model evaluation.

Interoperability has crucial role in data discovery, open access, utilization and reuse. It is a necessary attribute of data, especially in multidisciplinary fields such as air pollution studies and management. The process of information integration to generate actionable knowledge and intervention measures requires data interoperability, especially for the current reliance upon heterogeneous information sources from wide types and arrays of sensor resources, atmospheric models and voluntary information in air pollution management. Non-interoperability of the information sources severely impedes the process of information integration to generate actionable knowledge and intervention measures currently. Environmental regulatory bodies such as national or state level Pollution Control Boards (PCBs) monitor real time air pollution

in major urban centres of India. However, the data collected and disseminated under these programs do not comply with interoperability measures and hence, the data would not be open access, machine-readable with any metadata and Application Programming Interface(API) to enable the capability to process data seamlessly. Such limitations hinder integration of data with air quality modeling system and other volunteer information systems to address data gaps prevailing presently in air pollution management. Data interoperability is enabled through compliance with open standards. Sensor Observation Service (SOS) is an important interoperability open standard for sensor observations and its web based dissemination on real-time basis. The current study used *Python* programming language based implementation of SOS standard namely IstSOS. The characteristics of SOS for monitoring and modeling data could be described in terms of four interoperability levels namely semantic, system, structure and syntax achieved by importing the data into IstSOS program. Semantic interoperability describes meaning and background information on sensor resource and its observations. IstSOS provide a set of facilities to describe the metadata of imported data into it, used for describing the semantic details of the imported monitoring and modeling data. System interoperability addresses variability in sensor resources such as type of sensors used in the network, and its mode of collection of data and communication. Structure interoperability deals with common encodings for real time data and its metadata. Information models in SOS provided the common encoding for describing the imported data and its metadata to enable structure interoperability. Software platform independent automatic machine readability is meant by syntax interoperability. This is achieved by provisioning the data in automatically recognizable form through REST-API. IstSOS provides a set of REST-API request methods to query and import the data. Advantages of interoperability measures are demonstrated by a dynamic web application and on-line statistical analysis routine. The software platform-independent and interoperable advantages of machine readable REST-API provided by SOS is well represented in those applications and that demonstrate the advantages of SOS.

Future direction

The study addressed the generation of real time particulate pollution data based on low cost monitor and air quality modeling suite. The following points describe the future study sought for follow-up of the current study.

1. Improve the particulate pollution monitor with low cost dust sensor. Calibrate the developed monitor with regulatory compliant particulate pollution monitor
2. Address the data limitations in Coimbatore emission inventory preparation. Using video-based image analysis techniques derive the vehicular kilometre travel, proportionate number of different types of vehicles passing over various road categories in Coimbatore
3. Utilizing the scale and open source tools for *Python* based emission inventory programming routine, develop software or web application for emission inventory calculation for India. Compile the required data for such application
4. Further study the evaluation of WRF-CHEM model results for Coimbatore region. Carry out sensitivity analysis of various parameterization schemes for particulate pollutant forecast in the region
5. Carry out study on developing online web application for WRF CHEM simulation for small urban centres of India
6. Address the issue with interoperability measures for model output data. Carry out a study on effectiveness of interoperability measures for dissemination of particulate pollution information in various social networking platforms

References

- [1] Daniel Vallero. *Fundamentals of air pollution*. Academic press, 2014.
- [2] Renyi Zhang, Gehui Wang, Song Guo, Misti L Zamora, Qi Ying, Yun Lin, Weigang Wang, Min Hu, and Yuan Wang. Formation of urban fine particulate matter. *Chemical reviews*, 115(10):3803–3855, 2015.
- [3] Sarath K. Guttikunda, Rahul Goel, and Pallavi Pant. Nature of air pollution, emission sources, and management in the Indian cities. *Atmospheric Environment*, 95:501–510, October 2014.
- [4] Jaidit Brar, Shishir Gupta, Anu Madgavkar, Barnik C Maitra, and Sunali Rohra. India’s economic geography in 2025: states, clusters and cities. Technical report, McKinsey Global Institute, McKinsey & Company, India, 2014.
- [5] S Fuzzi, U Baltensperger, K Carslaw, S Decesari, H Denier Van Der Gon, MC Facchini, D Fowler, I Koren, B Langford, and U Lohmann. Particulate matter, air quality and climate: lessons learned and future needs. *Atmospheric Chemistry and Physics*, 15(14): 8217–8299, 2015.
- [6] John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2016.
- [7] Sagnik Dey, Larry Di Girolamo, Aaron van Donkelaar, SN Tripathi, Tarun Gupta, and Manju Mohan. Variability of outdoor fine particulate (PM 2.5) concentration in the Indian subcontinent: a remote sensing approach. *Remote Sensing of Environment*, 127: 153–161, 2012.
- [8] Kalpana Balakrishnan, Santu Ghosh, Bhaswati Ganguli, Sankar Sambandam, Nigel Bruce, Douglas F Barnes, and Kirk R Smith. State and national household concentrations of PM 2.5 from solid cookfuel use: results from measurements and modeling in India for estimation of the global burden of disease. *Environmental Health*, 12(1):1, 2013.
- [9] Amit Awasthi, Nirankar Singh, Susheel Mittal, Prabhat K Gupta, and Ravinder Agarwal. Effects of agriculture crop residue burning on children and young on PFTs in North West India. *Science of the total environment*, 408(20):4440–4445, 2010.
- [10] Zheming Tong and K Max Zhang. The near-source impacts of diesel backup generators in urban environments. *Atmospheric Environment*, 109:262–271, 2015.

- [11] ZB Wang, M Hu, XY Pei, RY Zhang, P Paasonen, J Zheng, DL Yue, ZJ Wu, M Boy, and A Wiedensohler. Connection of organics to atmospheric new particle formation and growth at an urban site of Beijing. *Atmospheric Environment*, 103:7–17, 2015.
- [12] M Kanakidou, JH Seinfeld, SN Pandis, I Barnes, FJ Dentener, MC Facchini, R Van Dingenen, B Ervens, ANCJSE Nenes, and CJ Nielsen. Organic aerosol and global climate modelling: a review. *Atmospheric Chemistry and Physics*, 5(4):1053–1123, 2005.
- [13] Mattias Hallquist, JC Wenger, Urs Baltensperger, Y Rudich, David Simpson, M Claeys, J Dommen, NM Donahue, C George, and AH Goldstein. The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmospheric Chemistry and Physics*, 9(14):5155–5236, 2009.
- [14] Tanjima Pervin, Ulf-G Gerdtham, and Carl Hampus Lyttkens. Societal costs of air pollution-related health hazards: A review of methods and results. *Cost Effectiveness and Resource Allocation*, 6(1):1, 2008.
- [15] Anna Alberini and Alan Krupnick. Cost-of-illness and willingness-to-pay estimates of the benefits of improved air quality: evidence from Taiwan. *Land Economics*, pages 37–53, 2000.
- [16] Sachin D Ghude, DM Chate, C Jena, G Beig, R Kumar, MC Barth, GG Pfister, S Fadnavis, and Prakash Pithani. Premature mortality in India due to PM_{2.5} and ozone exposure. *Geophysical Research Letters*, 43(9):4650–4658, 2016.
- [17] C Arden Pope III and Douglas W Dockery. Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association*, 56(6):709–742, 2006.
- [18] Alexander Millman, Deliang Tang, and Frederica P Perera. Air pollution threatens the health of children in China. *Pediatrics*, 122(3):620–628, 2008.
- [19] G Brooke Anderson, Jenna R Krall, Roger D Peng, and Michelle L Bell. Is the relation between ozone and mortality confounded by chemical components of particulate matter? Analysis of 7 components in 57 US communities. *American journal of epidemiology*, page 188, 2012.
- [20] Robert D Brook, Sanjay Rajagopalan, C Arden Pope, Jeffrey R Brook, Aruni Bhatnagar, Ana V Diez-Roux, Fernando Holguin, Yuling Hong, Russell V Luepker, and Murray A Mittleman. Particulate matter air pollution and cardiovascular disease an update to the scientific statement from the American Heart Association. *Circulation*, 121(21):2331–2378, 2010.
- [21] Robert D Brook, Jeffrey R Brook, Bruce Urch, Renaud Vincent, Sanjay Rajagopalan, and Frances Silverman. Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation*, 105(13):1534–1536, 2002.
- [22] C Arden Pope, Joseph B Muhlestein, Heidi T May, Dale G Renlund, Jeffrey L Anderson, and Benjamin D Horne. Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution. *Circulation*, 114(23):2443–2448, 2006.

- [23] Jesus A Araujo, Berenice Barajas, Michael Kleinman, Xuping Wang, Brian J Bennett, Ke Wei Gong, Mohamad Navab, Jack Harkema, Constantinos Sioutas, and Aldons J Lusis. Ambient particulate pollutants in the ultrafine range promote early atherosclerosis and systemic oxidative stress. *Circulation research*, 102(5):589–596, 2008.
- [24] Frederica Perera, Karl Hemminki, Wieslaw Jedrychowski, Robin Whyatt, Ulka Campbell, Yanzhi Hsu, Regina Santella, Richard Albertini, and James P O'Neill. In utero DNA damage from environmental pollution is associated with somatic gene mutation in newborns. *Cancer Epidemiology Biomarkers & Prevention*, 11(10):1134–1137, 2002.
- [25] Laura Gaspari, Seong-Sil Chang, Regina M Santella, Seymour Garte, Paola Pedotti, and Emanuela Taioli. Polycyclic aromatic hydrocarbon-DNA adducts in human sperm as a marker of DNA damage and infertility. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 535(2):155–160, 2003.
- [26] IPCC Climate Change. The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. K., Tignor, M., Allen, SK, Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, PM, Eds, page 1535, 2013.
- [27] National Research Council (US). Committee on Research Priorities for Airborne Particulate Matter. *Research Priorities for Airborne Particulate Matter: Continuing Research Progress. IV*. National Academies Press, 2004.
- [28] Y Wang, Q Wan, W Meng, F Liao, H Tan, and R Zhang. Long-term impacts of aerosols on precipitation and lightning over the Pearl River Delta megacity area in China. *Atmospheric Chemistry and Physics*, 11(23):12421–12436, 2011.
- [29] Spyros N Pandis, Ksakousti Skyllakou, Kalliopi Florou, Evangelia Kostenidou, Christos Kaltsonoudis, Erion Hasa, and Albert A Presto. Urban particulate matter pollution: a tale of five cities. *Faraday Discussions*, 189:277–290, 2016.
- [30] Dan Costa. The confluence of air quality, health and climate change: a challenge to air scientist and everyone else. Paper presented at the meeting of Frontiers in Air Quality Science: Celebrating 21 years of the Environmental Research Group, London, 2014.
- [31] Olivier Chanel, Susann Henschel, Patrick G Goodman, Antonis Analitis, Richard W Atkinson, Alain Le Tertre, Ariana Zeka, and Sylvia Medina. Economic valuation of the mortality benefits of a regulation on SO₂ in 20 European cities. *The European Journal of Public Health*, 24(4):631–637, 2014.
- [32] Pallavi Pant, Sarath K Guttikunda, and Richard E Peltier. Exposure to particulate matter in India: A synthesis of findings and future directions. *Environmental research*, 147: 480–496, 2016.
- [33] Zhen Cheng, Lina Luo, Shuxiao Wang, Yungang Wang, Sumit Sharma, Hikari Shimadera, Xiaoliang Wang, Michael Bressi, Regina Maura de Miranda, and Jingkun Jiang. Status and characteristics of ambient PM 2.5 pollution in global megacities. *Environment international*, 89:212–221, 2016.

- [34] World Health Organization. Global urban ambient air pollution, 2016. data retrieved from WHO Global Urban Ambient Air Pollution Database, http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/.
- [35] Peter Brimblecombe. *The big smoke: a history of air pollution in London since medieval times*. Routledge, 2011.
- [36] Michael Greenstone and Rema Hanna. Environmental regulations, air and water pollution, and infant mortality in India. *The American Economic Review*, 104(10):3038–3072, 2014.
- [37] Abhijit V Banerjee, Esther Duflo, and Rachel Glennerster. Putting a Band-Aid on a corpse: Incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6(2-3):487–500, 2008.
- [38] Esther Duflo, Michael Greenstone, Rohini Pande, and Nicholas Ryan. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. Technical report, National Bureau of Economic Research, 2013.
- [39] Abhijit Banerjee, Sendhil Mullainathan, and Rema Hanna. Corruption. In Jan Fagerberg, David C. Mowery, and Richard R. Nelson, editors, *The handbook of organizational economics*, chapter 27, pages 1109–1047. Princeton University Press, Princeton, 2013.
- [40] Nancy B Grimm, Stanley H Faeth, Nancy E Golubiewski, Charles L Redman, Jianguo Wu, Xuemei Bai, and John M Briggs. Global change and the ecology of cities. *science*, 319(5864):756–760, 2008.
- [41] Nancy B Grimm, Charles L Redman, Christopher G Boone, Daniel L Childers, Sharon L Harlan, and BL Turner II. Viewing the urban socio-ecological system through a sustainability lens: lessons and prospects from the Central Arizona–Phoenix LTER Programme. In *Long Term Socio-Ecological Research*, pages 217–246. Springer, 2013.
- [42] Julia Schmale, Erika von Schneidemesser, Ilan Chabay, Achim Maas, and Mark G Lawrence. Building Interfaces That Work: A Multi-stakeholder Approach to Air Pollution and Climate Change Mitigation. In *Communicating Climate-Change and Natural Hazard Risk and Cultivating Resilience*, pages 65–76. Springer, 2016.
- [43] Jianguo Liu, Harold Mooney, Vanessa Hull, Steven J Davis, Joanne Gaskell, Thomas Hertel, Jane Lubchenco, Karen C Seto, Peter Gleick, and Claire Kremen. Systems integration for global sustainability. *Science*, 347(6225):1258832, 2015.
- [44] Julia Schmale and Ilan Chabay. Coupled Human-Atmosphere-System Thinking. In *EGU General Assembly Conference Abstracts*, volume 16, page 4879, 2014.
- [45] Ana Nuno, Nils Bunnefeld, and Eleanor J Milner-Gulland. Managing social-ecological systems under uncertainty: Implementation in the real world. *Ecology and Society*, 19(2):52, 2014.
- [46] Brian Walker^{1a}, Stephen Carpenter, John Anderies^{1b}, Nick Abel^{1b}, Graeme Cumming, Marco Janssen, Louis Lebel, Jon Norberg, Garry D Peterson, and Rusty Pritchard. Resilience management in social-ecological systems: a working hypothesis for a participatory approach. *Conservation ecology*, 6(1):14, 2002.

- [47] Shreekant GUPTA. Environmental Policy and Governance in a Federal Framework: Perspectives from India. In *Environmental Policies in Asia: Perspectives from Seven Asian Countries*, pages 15–42. World Scientific, 2014.
- [48] Susan G Hadden. Statutes and standards for pollution control in India. *Economic and political weekly*, pages 709–720, 1987.
- [49] Madhav Gadgil. Towards a cooperative commonwealth of knowledge. Paper presented at the workshop on Open Source Development of Geo-spatial Technology Community level planning, Hyderabad, 2014.
- [50] Christine J Kirchhoff, Maria Carmen Lemos, and Suraje Dessai. Actionable knowledge for environmental decision making: broadening the usability of climate science. *Annual review of environment and resources*, 38(1):393, 2013.
- [51] David W Cash, William C Clark, Frank Alcock, Nancy M Dickson, Noelle Eckley, David H Guston, Jill Jäger, and Ronald B Mitchell. Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, 100(14):8086–8091, 2003.
- [52] Karl Aberer, Saket Sathe, Dipanjan Chakraborty, Alcherio Martinoli, Guillermo Barrenetxea, Boi Faltings, and Lothar Thiele. OpenSense: open community driven sensing of environment. In *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming*, pages 39–42. ACM, 2010.
- [53] Lei Zhang, Arthur PJ Mol, Guizhen He, and Yonglong Lu. An implementation assessment of China’s environmental information disclosure decree. *Journal of Environmental Sciences*, 22(10):1649–1656, 2010.
- [54] Cedric Pugh. *Sustainability the Environment and Urbanisation*. Routledge, 2014.
- [55] Devendra K Punia and KBC Saxena. Managing inter-organisational workflows in eGovernment services. In *Proceedings of the 6th international conference on Electronic commerce*, pages 500–505. ACM, 2004.
- [56] Timothy D Hogan. *Data for Effective Policy and Decision-making in Indiana: Assessing Its Availability, Accessibility, and Analysis*. Central Indiana Corporate Partnership, 2005.
- [57] Sven Schade and Max Craglia. A Future Sensor Web for the Environment in Europe. In *EnviroInfo*, pages 529–539, 2010.
- [58] National Research Council (US). Division on Earth, Life Studies. Committee on Key Challenge Areas for Convergence, and Board on Life Sciences Health. *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. National Research Council (US), 2014.
- [59] Neil Davies, Dawn Field, David Gavaghan, Sally J Holbrook, Serge Planes, Matthias Troyer, Michael Bonsall, Joachim Claudet, George Roderick, and Russell J Schmitt. Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium. *GigaScience*, 5(1):1, 2016.

- [60] Rob Kitchin. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.
- [61] Stephanie E Hampton, Carly A Strasser, Joshua J Tewksbury, Wendy K Gram, Amber E Budden, Archer L Batcheller, Clifford S Duke, and John H Porter. Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3):156–162, 2013.
- [62] Chris A Mattmann. Computing: A vision for data science. *Nature*, 493(7433):473–475, 2013.
- [63] CL Philip Chen and Chun-Yang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275:314–347, 2014.
- [64] Tony Hey, Stewart Tansley, and Kristin M Tolle. Jim Gray on eScience: A transformed scientific method. In *The fourth paradigm: data-intensive scientific discovery*, pages xvii–xxxi. Microsoft Research Redmond, WA, 2009.
- [65] Gordon Bell, Tony Hey, and Alex Szalay. Beyond the data deluge. *Science*, 323(5919):1297–1298, 2009.
- [66] Jeff Dozier and William B Gail. The emerging science of environmental applications. In *The fourth paradigm: data-intensive scientific discovery*, pages 13–20. Microsoft Research Redmond, WA, 2009.
- [67] Alexander Kotsev, Sven Schade, Massimo Craglia, Michel Gerboles, Laurent Spinelle, and Marco Signorini. Next Generation Air Quality Platform: Openness and Interoperability for the Internet of Things. *Sensors*, 16(3):403, 2016.
- [68] Vladimir Vujovic and Milan Maksimovic. Raspberry Pi as a Wireless Sensor node: Performances and constraints. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*, pages 1013–1018. IEEE, 2014.
- [69] Khaiwal Ravindra, Maninder Kaur Sidhu, Suman Mor, Siby John, and Saumyadipta Pyne. Air Pollution in India: Bridging the Gap between Science and Policy. *Journal of Hazardous, Toxic, and Radioactive Waste*, page A4015003, 2015.
- [70] JA Dearing, B Acma, S Bub, FM Chambers, Xuezhong Chen, J Cooper, Darren Crook, XH Dong, M Dotterweich, and ME Edwards. Social-ecological systems in the Anthropocene: the need for integrating social and biophysical records at regional scales. *Anthropocene Review*, pages 1–27, 2015.
- [71] Rob Kitchin. The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1):1–14, 2014.
- [72] White House. Open Government Directive. <http://www.whitehouse.gov/sites/default/files/microsites/ogi-directive.pdf>, 2009. (Visited on 08/22/2015).
- [73] Michael Batty. *The new science of cities*. MIT Press, 2013.

- [74] Arne Bröring, Johannes Echterhoff, Simon Jirka, Ingo Simonis, Thomas Everding, Christoph Stasch, Steve Liang, and Rob Lemmens. New generation sensor web enablement. *Sensors*, 11(3):2652–2699, 2011.
- [75] Directorate of Census Operations, Tamil Nadu. District census handbook-coimbatore, 2011. data retrieved from District Census Handbook, http://www.censusindia.gov.in/2011census/dchb/3331_PART_B_DCHB_COIMBATORE.pdf.
- [76] Census of India. Population and decadal growth rate by residence-Tamil Nadu. Technical report, Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India, 2011.
- [77] Department of Economics and Statistics. Season and Crop Report- Tamil Nadu. Technical report, Government of Tamil Nadu, 2015.
- [78] Niveta Jain, Arti Bhatia, and Himanshu Pathak. Emission of air pollutants from crop residue burning in India. *Aerosol and Air Quality Research*, 14(1):422–430, 2014.
- [79] Coimbatore District Administration. Coimbatore district profile 2015. Technical report, Government of Tamil Nadu, 2015.
- [80] Tamil Nadu Pollution Control Board(TNPCB). Final action plan report for critically polluted area Kurichi industrial cluster, Coimbatore. Technical report, Government of Tamil Nadu, 2010.
- [81] State Transport Authority. Growth of vehicles in tamil nadu, 2016. data retrieved from State Transport Authority, Government of Tamil Nadu, <http://www.tn.gov.in/sta/tables.htm>.
- [82] R. Mohanraj and P.a. Azeez. Urban development and particulate air pollution in Coimbatore city, India. *International Journal of Environmental Studies*, 62(1):69–78, February 2005.
- [83] R Mohanraj, S Dhanakumar, and G Solaraj. Polycyclic aromatic hydrocarbons bound to PM 2.5 in urban Coimbatore, India with emphasis on source apportionment. *The Scientific World Journal*, 2012:980843, January 2012.
- [84] ENVIS Centre. Air pollution database in tamil nadu, 2014. data retrieved from Department of Environment, Government of Tamil Nadu, http://cpcb.nic.in/divisionsofheadoffice/ess/Air_Pollution_Database_for_Tamil_Nadu_-_2014.pdf.
- [85] Rangaswamy Mohanraj and Parapurath Abdul Azeez. Polycyclic aromatic hydrocarbons in PM 10 of urban and suburban Coimbatore, India. *Fresenius Environmental Bulletin*, 13(4):332–335, 2004.
- [86] R Mohanraj, P a Azeez, and T Priscilla. Heavy metals in airborne particulate matter of urban Coimbatore. *Archives of environmental contamination and toxicology*, 47(2): 162–7, August 2004.
- [87] Central Pollution Control Board. CPCB Bulletin. Technical report, Ministry of Environment, Forest and Climate Change (MoEFCC), Government of India, 2016.

- [88] Sangeeta Nandi and Shama Gamkhar. Urban challenges in India: A review of recent policy measures. *Habitat International*, 39:55–61, 2013.
- [89] Nirvikar Singh. Fiscal federalism and decentralization in India. *Available at SSRN 1282267*, 2007.
- [90] Shirish Sankhe, Ireena Vittal, Richard Dobbs, Ajit Mohan, Ankur Gulati, Jonathan Ablett, Shishir Gupta, Alex Kim, Sudipto Paul, and Aditya Sanghvi. India's Urban Awakening: Building Inclusive Cities, Sustaining Economic Growth. Technical report, McKinsey Global Institute, McKinsey & Company, India, 2012.
- [91] Joseph James Whitworth. *The Role of Planning in Contemporary Urban India: Consequences and Lessons from the Hyderabad Metropolitan Rail: Telangana, India*. PhD thesis, University of Otago, 2015.
- [92] National Institute of Urban Affairs. Report on urban infrastructure and services. Technical report, Ministry of Urban Development, Government of India, 2011.
- [93] DM Holstius, A Pillarisetti, KR Smith, and E Seto. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmospheric Measurement Techniques*, 7(4):1121–1131, 2014.
- [94] Michael Brauer, Markus Amann, Rick T Burnett, Aaron Cohen, Frank Dentener, Majid Ezzati, Sarah B Henderson, Michal Krzyzanowski, Randall V Martin, and Rita Van Dingenen. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environmental science & technology*, 46(2):652–660, 2012.
- [95] Stephen S Lim, Theo Vos, Abraham D Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A AlMazroa, Markus Amann, H Ross Anderson, and Kathryn G Andrews. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, 380(9859):2224–2260, 2012.
- [96] A Tittarelli, A Borgini, M Bertoldi, E De Saeger, A Ruprecht, R Stefanoni, G Tagliabue, P Contiero, and P Crosignani. Estimation of particle mass concentration in ambient air using a particle counter. *Atmospheric Environment*, 42(36):8543–8548, 2008.
- [97] Jonathan M Samet, Francesca Dominici, Frank C Currier, Ivan Coursac, and Scott L Zeger. Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *New England journal of medicine*, 343(24):1742–1749, 2000.
- [98] Dena H Jaffe, Mendel E Singer, and Alfred A Rimm. Air pollution and emergency department visits for asthma among Ohio Medicaid recipients, 1991–1996. *Environmental Research*, 91(1):21–28, 2003.
- [99] Douglas W Dockery, C Arden Pope, Xiping Xu, John D Spengler, James H Ware, Martha E Fay, Benjamin G Ferris Jr, and Frank E Speizer. An association between air pollution and mortality in six US cities. *New England journal of medicine*, 329(24):1753–1759, 1993.

- [100] C Arden Pope III, Richard T Burnett, Michael J Thun, Eugenia E Calle, Daniel Krewski, Kazuhiko Ito, and George D Thurston. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, 287(9):1132–1141, 2002.
- [101] W James Gauderman, Edward Avol, Frank Gilliland, Hita Vora, Duncan Thomas, Kiro Berhane, Rob McConnell, Nino Kuenzli, Fred Lurmann, and Edward Rappaport. The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine*, 351(11):1057–1067, 2004.
- [102] Chandra Venkataraman. Major Sources Contributing to Current and Future Fine Particulate Pollution in India. Paper presented at the AAAS 2016 Annual Meeting: Global Science Engagement, 2016.
- [103] Ning Li, Constantinos Sioutas, Arthur Cho, Debra Schmitz, Chandan Misra, Joan Sempf, Meiyang Wang, Terry Oberley, John Froines, and Andre Nel. Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Environmental health perspectives*, 111(4):455, 2003.
- [104] Federico Karagulian, Claudio A Belis, Carlos Francisco, C Dora, Annette M Prüss-Üstün, Sophie Bonjour, Heather Adair-Rohani, and Markus Amann. Contributions to cities' ambient particulate matter (PM_{2.5}): A systematic review of local source contributions at global level. *Atmospheric Environment*, 120:475–483, 2015.
- [105] World Health Organization (WHO). Burden of disease from ambient air pollution for 2012. *World Health Organization*, 2014.
- [106] World Health Organization. Ambient air pollution, 2014. data retrieved from WHO's Ambient Air Pollution Database, http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_results_2014.pdf.
- [107] Judith C. Chow, Prakash Doraiswamy, John G. Watson, L.-W. Antony Chen, Steven Sai Hang Ho, and David A. Sodeman. Advances in Integrated and Continuous Measurements for Particle Mass and Chemical Composition. *Journal of the Air & Waste Management Association*, 58(2):141–163, February 2008.
- [108] Central Pollution Control Board. Air quality monitoring, EI and source apportionment study for Indian cities National Summary Report. Technical report, Ministry of Environment, Forest and Climate Change (MoEFCC), Government of India, 2010.
- [109] Thomas E McKone, P Barry Ryan, and Halûk Özkaynak. Exposure information in environmental health research: Current opportunities and future directions for particulate matter, ozone, and toxic air pollutants. *Journal of Exposure Science and Environmental Epidemiology*, 19(1):30–44, 2009.
- [110] Simone Simões Amaral, João Andrade de Carvalho, Maria Angélica Martins Costa, and Cleverson Pinheiro. An Overview of Particulate Matter Measurement Instruments. *Atmosphere*, 6(9):1327–1345, 2015.
- [111] M Obaidullah, S Bram, VK Verma, and J De Ruyck. A review on particle emissions from small scale biomass combustion. *International Journal of Renewable Energy Research (IJRER)*, 2(1):147–159, 2012.

- [112] WE Wilson, Judith C Chow, Candis Claiborn, Wei Fusheng, Johann Engelbrecht, and John G Watson. Monitoring of particulate matter outdoors. *Chemosphere*, 49(9): 1009–1043, 2002.
- [113] Barouch Giechaskiel, Matti Maricq, Leonidas Ntziachristos, Christos Dardiotis, Xiaoliang Wang, Harald Axmann, Alexander Bergmann, and Wolfgang Schindler. Review of motor vehicle particulate emissions sampling and measurement: From smoke and filter mass to particle number. *Journal of Aerosol Science*, 67:48–86, 2014.
- [114] World Health Organization. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Technical report, Geneva: World Health Organization, 2006.
- [115] Ki-Hyun Kim, Ehsanul Kabir, and Shamin Kabir. A review on the human health impact of airborne particulate matter. *Environment international*, 74:136–143, 2015.
- [116] Jill Engel-Cox, Nguyen Thi Kim Oanh, Aaron van Donkelaar, Randall V. Martin, and Erica Zell. Toward the next generation of air quality monitoring: Particulate Matter. *Atmospheric Environment*, 80:584–590, December 2013.
- [117] Pramod Kulkarni and Paul A Baron. An Approach to Performing Aerosol Measurements. In *Aerosol Measurement: Principles, Techniques, and Applications, Third Edition*, pages 55–65. John Wiley & Sons, Inc., 2011.
- [118] H-Erich Wichmann, Claudia Spix, Thomas Tuch, G Wölke, Annette Peters, Joachim Heinrich, Wolfgang G Kreyling, and Joachim Heyder. Daily mortality and fine and ultrafine particles in Erfurt, Germany part I: role of particle number and particle mass. Technical Report 98, Health Effects Institute, 2000.
- [119] B Gomišček, H Hauck, S Stopper, and O Preining. Spatial and temporal variations of PM 1, PM 2.5, PM 10 and particle number concentration during the AUPHEP—project. *Atmospheric Environment*, 38(24):3917–3934, 2004.
- [120] J Ruuskanen, Th Tuch, H Ten Brink, A Peters, A Khlystov, A Mirme, GPA Kos, B Brunekreef, HE Wichmann, and G Buzorius. Concentrations of ultrafine, fine and PM 2.5 particles in three European cities. *Atmospheric Environment*, 35(21):3729–3738, 2001.
- [121] Joshua M Pearce. Building research equipment with free, open-source hardware. *Science*, 337(6100):1303–1304, 2012.
- [122] Milena Jovašević-Stojanović, Alena Bartonova, Dušan Topalović, Ivan Lazović, Boris Pokrić, and Zoran Ristovski. On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. *Environmental Pollution*, 206:696–704, 2015.
- [123] Jan Theunis, Matthias Stevens, and Dick Botteldooren. Sensing the Environment. In *Participatory Sensing, Opinions and Collective Awareness*, pages 21–46. Springer, 2017.
- [124] Yang Wang, Jiayu Li, He Jing, Qiang Zhang, Jingkun Jiang, and Pratim Biswas. Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement. *Aerosol Science and Technology*, 49(11):1063–1077, 2015.

- [125] Emily G Snyder, Timothy H Watkins, Paul A Solomon, Eben D Thoma, Ronald W Williams, Gayle SW Hagler, David Shelow, David A Hindin, Vasu J Kilaru, and Peter W Preuss. The changing paradigm of air pollution monitoring. *Environmental science & technology*, 47(20):11369–11377, 2013.
- [126] Amanda L Northcross, Rufus J Edwards, Michael A Johnson, Zhong-Min Wang, Kun-ning Zhu, Tracy Allen, and Kirk R Smith. A low-cost particle counter as a realtime fine-particle mass monitor. *Environmental Science: Processes & Impacts*, 15(2):433–439, 2013.
- [127] Meiling Gao, Junji Cao, and Edmund Seto. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM_{2.5} in Xi’an, China. *Environmental Pollution*, 199:56–65, 2015.
- [128] Matthias Budde, Lin Zhang, and Michael Beigl. Distributed, Low-cost Particulate Matter Sensing: Scenarios, Challenges, Approaches. In *Proc. 1st Int. Conf. Atmospheric Dust*, 2014.
- [129] ThingSpeak: Internet Of Things. <https://thingspeak.com/>, 2016. (Visited on 02/16/2016).
- [130] Justin Arling, Kyle O’Connor, and Michael Mercieca. Air Quality Sensor Network for Philadelphia, 2010.
- [131] K A Nishadh, P A Azeez, and R Mohanraj. Particulate Air pollution Data for Coimbatore-Tools for real time monitoring and modeling, August 2016. URL <https://doi.org/10.5281/zenodo.60499>.
- [132] Ji Yi Lee, Hye Jung Shin, Soo Ya Bae, Yong Pyo Kim, and Chang-Hee Kang. Seasonal variation of particle size distributions of PAHs at Seoul, Korea. *Air Quality, Atmosphere & Health*, 1(1):57–68, 2008.
- [133] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, page 15. ACM, 2013.
- [134] N Ramanathan, M Lukac, T Ahmed, A Kar, PS Praveen, T Honles, I Leong, IH Rehman, JJ Schauer, and V Ramanathan. A cellphone based system for large-scale monitoring of black carbon. *Atmospheric Environment*, 45(26):4481–4487, 2011.
- [135] YR Kolupula, MA Aceves-Fernandez, GR Jones, AG Deakin, and JW Spencer. Airborne particle monitoring with urban closed-circuit television camera networks and a chromatic technique. *Measurement Science and Technology*, 21(11):115204, 2010.
- [136] MA Aceves-Fernandez, GR Jones, YR Kolupula, T Reichelt, JW Spencer, C Pedraza-Ortega, and E Gorrostieta. Capability of a Portable Chromatic Unit for Monitoring Airborne Particles over Wide Urban Areas. *Journal of Sensors*, 2009, 2009.
- [137] Ron Williams, Vasu Kilaru, Emily Snyder, Amanda Kaufman, Timothy Dye, Andrew Rutter, Ashley Russell, and Hilary Hafner. Air Sensor Guidebook. Technical report, United States Environmental Protection Agency, 2014.

- [138] Bean T Chen, Robert A Fletcher, and YUNG-SUNG Cheng. Calibration of Aerosol Instruments. *Aerosol Measurement: Principles, Techniques, and Applications*, 2011.
- [139] David Hasenfratz, Olga Saukh, and Lothar Thiele. On-the-fly calibration of low-cost gas sensors. In *Wireless Sensor Networks*, pages 228–244. Springer, 2012.
- [140] Claire Granier, Bertrand Bessagnet, Tami Bond, Ariela D’Angiola, Hugo Denier van der Gon, Gregory J. Frost, Angelika Heil, Johannes W. Kaiser, Stefan Kinne, Zbigniew Klimont, et al. Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic Change*, 109(1):163–190, 2011.
- [141] Rainer Friedrich and Stefan Reis. *Emissions of air pollutants: measurements, calculations and uncertainties*. Springer Science & Business Media, 2013.
- [142] Nicolas Moussiopoulos. *Air quality in cities*. Springer Science & Business Media, 2003.
- [143] Gregory J Frost, Paulette Middleton, Leonor Tarrasón, Claire Granier, Alex Guenther, Beatriz Cardenas, Hugo Denier van der Gon, Greet Janssens-Maenhout, Johannes W Kaiser, and Terry Keating. New Directions: GEIA’s 2020 vision for better air emissions information. *Atmospheric environment*, 81, 2013.
- [144] Gregory J Frost, Stefan R Falke, Claire Granier, Terry Keating, Jean-François Lamarque, Megan L Melamed, Paulette Middleton, Gabrielle Pétron, and Steven J Smith. New Directions: Toward a community emissions approach. *Atmospheric environment*, 51: 333–334, 2012.
- [145] ACAP. What is an Emission Inventory? <http://www.acap.asia/publication/pdf/emissioneng.pdf>, 2016. (Visited on 08/14/2016).
- [146] SR Freitas, KM Longo, MF Alonso, M Pirre, V Marecal, G Grell, R Stockler, RF Mello, and M Sánchez Gácita. PREP-CHEM-SRC–1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models. *Geoscientific Model Development*, 4(2):419–433, 2011.
- [147] Alice B Gilliland, Christian Hogrefe, Robert W Pinder, James M Godowitch, Kristen L Foley, and ST Rao. Dynamic evaluation of regional air quality models: assessing changes in O₃ stemming from changes in emissions and meteorology. *Atmospheric Environment*, 42(20):5110–5123, 2008.
- [148] Saroj Kumar Sahu, Gufran Beig, and Neha Parkhi. High Resolution Emission Inventory of NO_x and CO for Mega City Delhi, India. *Aerosol and Air Quality Research*, 15: 1137–1144, 2015.
- [149] P Thunis, B Degraeuwe, K Cuvelier, M Guevara, L Tarrason, and A Clappier. A novel approach to screen and compare emission inventories. *Air Quality, Atmosphere & Health*, 9(4):325–333, 2016.
- [150] Armistead Russell and Robin Dennis. NARSTO critical review of photochemical models and modeling. *Atmospheric environment*, 34(12):2283–2324, 2000.

- [151] S François, E Grondin, S Fayet, and J-L Ponche. The establishment of the atmospheric emission inventories of the ESCOMPTE program. *Atmospheric Research*, 74(1):5–35, 2005.
- [152] Daniel Q Tong, Pius Lee, and Rick D Saylor. New Directions: The need to develop process-based emission forecasting models. *Atmospheric Environment*, 47:560–561, 2012.
- [153] Neal Fann, Charles M Fulcher, and Bryan J Hubbell. The influence of location, source, and emission type in estimates of the human health benefits of reducing a ton of air pollution. *Air Quality, Atmosphere & Health*, 2(3):169–176, 2009.
- [154] Maria de Fatima Andrade, Regina Maura de Miranda, Adalgiza Fornaro, Americo Kerr, Beatriz Oyama, Paulo Afonso de Andre, and Paulo Saldiva. Vehicle emissions and PM_{2.5} mass concentrations in six Brazilian cities. *Air Quality, Atmosphere & Health*, 5(1): 79–88, 2012.
- [155] Greet Janssens-Maenhout, Frank Dentener, John Van Aardenne, Suvi Monni, Valerio Pagliari, Lorenzo Orlandini, Zbigniew Klimont, Jun-ichi Kurokawa, Hajime Akimoto, and Toshimasa Ohara. EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories. *European Commission Publications Office, Ispra (Italy). JRC68434, EUR report No EUR*, 25:299–2012, 2012.
- [156] Anonymous. Emissions Inventory. http://www.clarkcountynv.gov/airquality/planning/Pages/Planning_EmissionsInventory.aspx, 2016. (Visited on 02/15/2016).
- [157] Manju Mohan, Shweta Bhati, Preeti Gunwani, and Pallavi Marappu. Emission inventory of air pollutants and trend analysis based on various regulatory measures over megacity delhi. In G.L.Badilla, Benjamin Valdez, and Michael Schorr, editors, *Air Quality - New Perspective*. INTECH Open Access Publisher, Croatia, 2004.
- [158] Ye Huang, Huizhong Shen, Han Chen, Rong Wang, Yanyan Zhang, Shu Su, Yuanchen Chen, Nan Lin, Shaojie Zhuo, and Qirui Zhong. Quantification of Global Primary Emissions of PM_{2.5}, PM₁₀, and TSP from Combustion and Industrial Process Sources. *Environmental science & technology*, 48(23):13834–13843, 2014.
- [159] Saroj Kumar Sahu, Gufran Beig, and Neha S Parkhi. Emissions inventory of anthropogenic PM 2.5 and PM 10 in Delhi during commonwealth games 2010. *Atmospheric Environment*, 45(34):6180–6190, 2011.
- [160] Greet Janssens-Maenhout, Valerio Pagliari, Diego Guizzardi, and Marilena Muntean. Global emission inventories in the Emission Database for Global Atmospheric Research (EDGAR)—Manual (I). *Gridding: EDGAR emissions distribution on global gridmaps, Publications Office of the European Union, Luxembourg*, 2013.
- [161] M Li, Q Zhang, J Kurokawa, JH Woo, KB He, Z Lu, T Ohara, Y Song, DG Sreets, and GR Carmichael. MIX: a mosaic Asian anthropogenic emission inventory for the MICS-Asia and the HTAP projects. *Atmospheric Chemistry and Physics Discussion*, 2015.

- [162] J Kurokawa, T Ohara, T Morikawa, S Hanayama, G Janssens-Maenhout, T Fukui, K Kawashima, and H Akimoto. Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2. *Atmospheric Chemistry and Physics*, 13(21):11019–11058, 2013.
- [163] Dharendra Singh, Sheo Prasad Shukla, Mukesh Sharma, Sailesh N Behera, Devendra Mohan, Narendra Bahadur Singh, and Govind Pandey. GIS-based on-road vehicular emission inventory for Lucknow, India. *Journal of Hazardous, Toxic, and Radioactive Waste*, page A4014006, 2014.
- [164] Census of India : Office of the Registrar General & Census Commissioner, India. <http://censusindia.gov.in/>, 2012. (Visited on 03/19/2015).
- [165] National Sample Survey Office(NSSO). House hold consumption of various goods and services in india, 2011-2012, nss 68th round, 2014. data retrieved from Ministry of Statistics and Programme Implementation Government of India, http://mospi.nic.in/mospi_new/upload/Report_no558_rou68_30june14.pdf.
- [166] Department of Economics and Statistics, Government of Tamil Nadu. Statistical hand book of tamil nadu, 2014. data retrieved from District Statistical Hand Book, <http://www.tn.gov.in/deptst/>.
- [167] P Goyal and Neeru Jaiswal. Vehicular Emissions Intervention in Delhi, India. http://www.indianjournals.com/glogift2k6/glogift2k6-1-1/theme_1/Article%202.htm, 2004. (Visited on 03/20/2015).
- [168] Coimbatore City Portal-A perfect solution for City Bus search. <http://www.way2cbe.com/index.php>, 2015. (Visited on 03/24/2015).
- [169] Ministry of New and Renewable Energy. Development of Coimbatore Solar City Final Master Plan. Technical report, Government of India, 2012.
- [170] SK Sahu, G Beig, and C Sharma. Decadal growth of black carbon emissions in India. *Geophysical Research Letters*, 35(2), 2008.
- [171] Zbigniew Klimont, Janusz Cofala, Imrich Bertok, Markus Amann, Chris Heyes, and F Gyarmas. Modelling particulate emissions in Europe: a framework to estimate reduction potential and control costs. *International Institute for Applied Systems Analysis*, 2002.
- [172] Tami C Bond, David G Streets, Kristen F Yarber, Sibyl M Nelson, Jung-Hun Woo, and Zbigniew Klimont. A technology-based global inventory of black and organic carbon emissions from combustion. *Journal of Geophysical Research: Atmospheres* (1984–2012), 109(D14), 2004.
- [173] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.
- [174] Wes McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56, 2010.

- [175] Gillies Sean, Buffat Rene, Wurster Kevin, Perry Matthew, Arnott Joshua, Jordahl Kelsey, and Cochran Micah. Fiona: OGR's neat, nimble, no-nonsense API. <https://github.com/Toblerity/Fiona>, 2016.
- [176] K Jordahl. GeoPandas: Python tools for geographic data. <https://github.com/geopandas/geopandas>, 2014.
- [177] Gillies Sean, Tonnhofer Oliver, Toews Mike, Arnott Joshua, Wasserman Jacob, Bierbaum Aron, and Schönberger Johannes. Shapely: Manipulation and analysis of geometric objects. <https://github.com/Toblerity/Shapely>, 2016.
- [178] Gillies Sean, Butler Howard, Pedersen Brent, and Matthias. Rtree: spatial index for Python GIS. <https://github.com/Toblerity/Shapely>, 2016.
- [179] Ministry of Electronics & Information Technology. Policy on Open Application Programming Interfaces for Government of India. Technical report, Government of India, 2015.
- [180] Bibiana McHugh. Pioneering open data standards: The GTFS story. *Edited by Brett Goldstein with Lauren Dyson*, pages 125–135, 2013.
- [181] Ajay Singh Nagpure and BR Gurjar. Development and evaluation of vehicular air pollution inventory model. *Atmospheric environment*, 59:160–169, 2012.
- [182] Carl Vondrick, Deva Ramanan, and Donald Patterson. Efficiently scaling up video annotation with crowdsourced marketplaces. In *European Conference on Computer Vision*, pages 610–623. Springer, 2010.
- [183] Fernando Pérez and Brian E Granger. IPython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29, 2007.
- [184] Helen Shen. Interactive notebooks: Sharing the code. *Nature*, 515(7525):151–152, 2014.
- [185] Dimitrios Gkatzoflias, Giorgos Mellios, and Zissis Samaras. Development of a web GIS application for emissions inventory spatial allocation based on open source software tools. *Computers & Geosciences*, 52:21–33, 2013.
- [186] Georg A Grell, Steven E Peckham, Rainer Schmitz, Stuart A McKeen, Gregory Frost, William C Skamarock, and Brian Eder. Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment*, 39(37):6957–6975, 2005.
- [187] David J Stensrud. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.
- [188] James R Holton and Gregory J Hakim. *An introduction to dynamic meteorology*, volume 88. Academic press, 2012.
- [189] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.
- [190] Y Zhang. Online-coupled meteorology and chemistry models: history, current status, and outlook. *Atmospheric Chemistry and Physics*, 8(11):2895–2932, 2008.

- [191] Yang Zhang, Marc Bocquet, Vivien Mallet, Christian Seigneur, and Alexander Baklanov. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*, 60:632–655, 2012.
- [192] Rajesh Kumar, Manish Naja, GG Pfister, MC Barth, C Wiedinmyer, and GP Brasseur. Simulations over South Asia using the Weather Research and Forecasting model with Chemistry (WRF-Chem): chemistry evaluation and initial results. *Geoscientific Model Development*, 5(3):619–648, 2012.
- [193] Medhavi Gupta and Manju Mohan. Assessment of contribution to PM 10 concentrations from long range transport of pollutants using WRF/Chem over a subtropical urban airshed. *Atmospheric Pollution Research*, 4(4):405–410, 2013.
- [194] Reka Srinivas, Abhilash S Panicker, Neha S Parkhi, Sunil K Peshin, and Gufran Beig. Sensitivity of online coupled model to extreme pollution event over a mega city Delhi. *Atmospheric Pollution Research*, 2015.
- [195] M Michael, A Yadav, SN Tripathi, VP Kanawade, A Gaur, P Sadavarte, and C Venkataraman. Simulation of trace gases and aerosols over the Indian domain: evaluation of the WRF-Chem model. *Geoscientific Model Development Discussions*, 7(1):431–482, 2014.
- [196] Tabish U Ansari, Aswathy E Valsan, N Ojha, R Ravikrishna, Balaji Narasimhan, and Sachin S Gunthe. Model simulations of fungal spore distribution over the Indian region. *Atmospheric Environment*, 122:552–560, 2015.
- [197] SAFAR - India. <http://safar.tropmet.res.in/>, 2016. (Visited on 02/27/2016).
- [198] RA Rozumalski. A Nearly Complete Guide to the WRF EMS V3. *NOAA/NWS National SOO Science and Training Resource Coordinator, Boulder Colorado*, 2010.
- [199] AM Guerrero-Higueras, E García-Ortega, JL Sánchez, J Lorenzana, and V Matellán. Schedule WRF model executions in parallel computing environments using Python. In *Third Symposium on Advances in Modeling and Analysis Using Python*, 2013.
- [200] RL John. johnrobertlawson/lazyWRF · GitHub. <https://github.com/johnrobertlawson/lazyWRF>, 2015. (Visited on 03/24/2015).
- [201] Andre R Erler. WRF-Tools: Tools and scripts to set up and run WRF climate simulations. <https://github.com/aerler/WRF-Tools>, 2015. (Visited on 02/27/2016).
- [202] Viera Šipková and Ladislav Hluchý. Software support for the execution of WRF (Weather Research and Forecasting) simulations on HPC infrastructures. <http://ups.savba.sk/~sipka/wrf-KCI.pdf>, 2015. (Visited on 02/27/2016).
- [203] Manju Mohan and Shweta Bhati. Analysis of WRF model performance over subtropical region of Delhi, India. *Advances in Meteorology*, 2011, 2011.
- [204] Gaurav Govardhan, Ravi S Nanjundiah, SK Satheesh, K Krishnamoorthy, and VR Kotamarthi. Performance of WRF-Chem over Indian region: Comparison with measurements. *Journal of Earth System Science*, 124(4):875–896, 2015.

- [205] Alexey A Voinov, Cecelia DeLuca, Raleigh R Hood, Scott Peckham, Christopher R Sherwood, and James PM Syvitski. A community approach to earth systems modeling. *Eos, Transactions American Geophysical Union*, 91(13):117–118, 2010.
- [206] J Michalakes, S Chen, J Dudhia, L Hart, J Klemp, J Middlecoff, and W Skamarock. Development of a next generation regional weather research and forecast model. In *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the use of high performance computing in meteorology*, volume 1, pages 269–276. World Scientific, 2001.
- [207] J Michalakes, J Dudhia, D Gill, T Henderson, J Klemp, W Skamarock, and W Wang. The weather research and forecast model: software architecture and performance. In *Proceedings of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology*, pages 156–168, 2005.
- [208] Thomas Tomkins Warner. *Numerical weather and climate prediction*. Cambridge University Press, 2010.
- [209] David C Carslaw and Karl Ropkins. Openair—an R package for air quality data analysis. *Environmental Modelling & Software*, 27:52–61, 2012.
- [210] David R Legates and Gregory J McCabe. A refined index of model performance: a rejoinder. *International Journal of Climatology*, 33(4):1053–1056, 2013.
- [211] David R Legates and Gregory J McCabe. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water resources research*, 35(1):233–241, 1999.
- [212] Charles S Zender. Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). *Environmental Modelling & Software*, 23(10):1338–1342, 2008.
- [213] Online Tutorial-WRF. Compiling WRF. http://www2.mmm.ucar.edu/wrf/OnLineTutorial/compilation_tutorial.php, 2012. (Visited on 07/29/2016).
- [214] G Janssens-Maenhout, M Crippa, D Guizzardi, F Dentener, M Muntean, G Pouliot, T Keating, Q Zhang, J Kurokawa, and R Wankmüller. HTAP_v2: a mosaic of regional and global emission gridmaps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmospheric Chemistry & Physics Discussions*, 15(8), 2015.
- [215] Christopher W Harrop, Ligia Bernardet, Mark Govett, Jeff S Smith, and Stephen Weygandt. A workflow management system for automating weather and climate simulations. In *CIRES’Annual, Institute-wide Symposium*. Citeseer, 2008.
- [216] Jeff S Smith, MW Govett, and PT McCaslin. WRF Portal and Domain Wizard: Graphical tools for WRF. In *23rd Conf. on IIPS*, 2007.
- [217] What’s wrong with Perl. <http://www.garshol.priv.no/download/text/perl.html>, 2009. (Visited on 07/27/2016).
- [218] Michel F Sanner. Python: a programming language for software integration and development. *J Mol Graph Model*, 17(1):57–61, 1999.

-
- [219] Gregory V Wilson. Where's the real bottleneck in scientific computing? *American Scientist*, 94(1):5, 2006.
- [220] Travis E Oliphant. Python for scientific computing. *Computing in Science & Engineering*, 9(3):10–20, 2007.
- [221] Fernando Perez, Brian E Granger, and John D Hunter. Python: an ecosystem for scientific computing. *Computing in Science & Engineering*, 13(2):13–21, 2011.
- [222] Johnny Wei-Bing Lin. Why Python is the next wave in earth sciences computing. *Bulletin of the American Meteorological Society*, 93(12):1823–1824, 2012.
- [223] Jon Sáenz, Juan Zubillaga, and Jesús Fernández. Geophysical data analysis using Python. *Computers & geosciences*, 28(4):457–465, 2002.
- [224] Martinho Marta-Almeida, Manuel Ruiz-Villarreal, Pablo Otero, Marcos Cobas, Alvaro Peliz, Rita Nolasco, Mauro Cirano, and Janini Pereira. A Python engine for automating regional and coastal ocean forecasts. *Environmental Modelling & Software*, 26(5):680–682, 2011.
- [225] Steven K Morley, Daniel T Welling, Josef Koller, Brian A Larsen, Michael G Henderson, and Jonathan Niehof. SpacePy—a python-based library of tools for the space sciences. In *Proceedings of the 9th Python in science conference (SciPy 2010)*, pages 39–45. Citeseer, 2010.
- [226] Peter JA Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, and Bartek Wilczynski. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.
- [227] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.
- [228] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: A CPU and GPU math compiler in Python. In *Proc. 9th Python in Science Conf*, pages 1–7, 2010.
- [229] Johnny Wei-Bing Lin. *A Hands-On Introduction to Using Python in the Atmospheric and Oceanic Sciences*. Lulu. com, 2012.
- [230] Sarath Guttikunda, Giuseppe Calori, Fanny Velay-Lasry, and Richard Ngo. Air quality forecasting system for cities: modeling architecture for Delhi. Technical report, Simple Interactive Models for Better Air Quality SIM-air Working Paper Series: 36, 2011.
- [231] CV Srinivas, R Venkatesan, NV Muralidharan, Someshwar Das, Hari Dass, and P Esvara Kumar. Operational mesoscale atmospheric dispersion prediction using a parallel computing cluster. *Journal of earth system science*, 115(3):315–332, 2006.

- [232] Chandan Maheshwari. The e-infrastructure at c-dac for climate & weather research. Paper presented at the Conference on the Role of e-Infrastructures for Climate Change Research, Trieste - Italy, 2011.
- [233] Constantinos Evangelinos and C Hill. Cloud computing for parallel scientific HPC applications: Feasibility of running coupled atmosphere-ocean climate models on Amazon's EC2. *ratio*, 2(2.40):2–34, 2008.
- [234] Lei Wang, Jianfeng Zhan, Weisong Shi, and Yi Liang. In cloud, can scientific communities benefit from the economies of scale? *IEEE Transactions on Parallel and Distributed Systems*, 23(2):296–303, 2012.
- [235] Lamia Youseff, Maria Butrico, and Dilma Da Silva. Toward a unified ontology of cloud computing. In *2008 Grid Computing Environments Workshop*, pages 1–10. IEEE, 2008.
- [236] Alexandru Iosup, Simon Ostermann, M Nezh Yigitbasi, Radu Prodan, Thomas Fahringer, and Dick Epema. Performance analysis of cloud computing services for many-tasks scientific computing. *IEEE Transactions on Parallel and Distributed systems*, 22(6):931–945, 2011.
- [237] Christian Vecchiola, Suraj Pandey, and Rajkumar Buyya. High-performance cloud computing: A view of scientific applications. In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, pages 4–16. IEEE, 2009.
- [238] Gideon Juve, Ewa Deelman, Karan Vahi, Gaurang Mehta, Bruce Berriman, Benjamin P Berman, and Phil Maechling. Scientific workflow applications on Amazon EC2. In *2009 5th IEEE International Conference on E-Science Workshops*, pages 59–66. IEEE, 2009.
- [239] Jaliya Ekanayake and Geoffrey Fox. High performance parallel computing with clouds and cloud technologies. In *International Conference on Cloud Computing*, pages 20–38. Springer, 2009.
- [240] Qiming He, Shujia Zhou, Ben Kobler, Dan Duffy, and Tom McGlynn. Case study for running HPC applications in public clouds. In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, pages 395–401. ACM, 2010.
- [241] Peter J Walton, Morgan B Yarker, Michel d S Mesquita, and Friederike EL Otto. Helping to Make Sense of Regional Climate Modeling: Professional Development for Scientists and Decision Makers Anytime, Anywhere. *Bulletin of the American Meteorological Society*, 97(2015), 2015.
- [242] CW Tessum, JD Hill, and JD Marshall. Twelve-month, 12 km resolution North American WRF-Chem v3. 4 air quality simulation: Performance evaluation. *Geoscientific Model Development Discussions (Online)*, 7(6), 2014.
- [243] James W Boylan and Armistead G Russell. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmospheric environment*, 40(26):4946–4959, 2006.
- [244] World Health Organization. Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia. *World Health Organization Regional Office for Europe, Copenhagen*, 2013.

- [245] Michal Krzyzanowski, John Vandenberg, and Dave Stieb. Perspectives on air quality policy issues in Europe and North America. *Journal of Toxicology and Environmental Health, Part A*, 68(13-14):1057–1061, 2005.
- [246] Sarath K. Guttikunda, Rahul Goel, and Pallavi Pant. Nature of air pollution, emission sources, and management in the Indian cities. *Atmospheric Environment*, 95:501–510, October 2014.
- [247] Yang Zhang, Marc Bocquet, Vivien Mallet, Christian Seigneur, and Alexander Baklanov. Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. *Atmospheric Environment*, 60:656–676, 2012.
- [248] George Percivall. The application of open standards to enhance the interoperability of geoscience information. *International Journal of Digital Earth*, 3(S1):14–30, 2010.
- [249] ISO. ISO/IEC 2382:2015 - Information technology – Vocabulary. Technical report, International Organization for Standardization, 2013.
- [250] Aris M Ouksel and Amit Sheth. Semantic interoperability in global information systems. *ACM Sigmod Record*, 28(1):5–12, 1999.
- [251] Sarath Guttikunda and Puja Jawahar. Urban Air Pollution Analysis in India. Technical report, UrbanEmissions.Info, New Delhi, India, 2011.
- [252] Paul Edwards, Matthew S Mayernik, Archer Batcheller, Geoffrey Bowker, and Christine Borgman. Science friction: Data, metadata, and collaboration. *Social Studies of Science*, page 0306312711413314, 2011.
- [253] Steve C Hankin, Jon D Blower, Thierry Carval, Kenneth S Casey, Craig Donlon, Olivier Lauret, Thomas Loubrieu, Ashwanth Srinivasan, Joaquim Trinanes, and Oystein Godoy. NetCDF-CF-OPeNDAP: Standards for ocean data interoperability and object lessons for community data standards processes. In *Oceanobs 2009, Venice Convention Centre, 21-25 September 2009, Venice*, 2010.
- [254] Mike Botts, George Percivall, Carl Reed, and John Davidson. OGC® sensor web enablement: Overview and high level architecture. In *GeoSensor networks*, pages 175–190. Springer, 2006.
- [255] OGC. Open Geospatial Consortium, OpenGIS Sensor Observation Service. OpenGIS® Implementation Standard. Technical report, Open Geospatial Consortium, 2007.
- [256] Stefano Nativi, Paolo Mazzetti, and Gary N Geller. Environmental model access and interoperability: The GEO Model Web initiative. *Environmental Modelling & Software*, 39:214–228, 2013.
- [257] Kym Watson and Hylke van der Schaaf. Describing models on the web using OGC standards. Paper presented at the 20th International Congress on Modelling and Simulation, Adelaide, 2013.
- [258] M Cannata, M Antonovic, M Molinari, and M Pozzoni. istSOS, a new sensor observation management system: software architecture and a real-case application for flood protection. *Geomatics, Natural Hazards and Risk*, 6(8):635–650, 2015.

- [259] Arne Bröring, Eike Hinderk Jürrens, Simon Jirka, and Christoph Stasch. Development of sensor web applications with open source software. In *First open source GIS UK conference (OSGIS 2009)*, volume 22. Citeseer, 2009.
- [260] G van Rossum and JFL Drake. The Python Language Reference Manual. Technical report, Network Theory Ltd., 2006.
- [261] David Flanagan. *JavaScript: the definitive guide*. O'Reilly Media, Inc., 2006.
- [262] John Worsley and Joshua D Drake. *Practical PostgreSQL*. O'Reilly Media, Inc., 2002.
- [263] Mpfeil. QualitySCHU- a web application. <https://github.com/mpfeil/qualitySCHU>, 2012. (Visited on 03/21/2015).
- [264] Daniel Nüst, Christoph Stasch, and Edzer Pebesma. Connecting R to the sensor web. In *Advancing Geoinformation Science for a Changing World*, pages 227–246. Springer, 2011.
- [265] Yu Su and Gagan Agrawal. Supporting user-defined subsetting and aggregation over parallel netcdf datasets. In *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on*, pages 212–219. IEEE, 2012.
- [266] Kui Gao, Wei-keng Liao, Alok Choudhary, Robert Ross, and Robert Latham. Combining I/O operations for multiple array variables in parallel netCDF. In *2009 IEEE International Conference on Cluster Computing and Workshops*, pages 1–10. IEEE, 2009.
- [267] Benjamin D Best, Patrick N Halpin, Ei Fujioka, Andrew J Read, Song S Qian, Lucie J Hazen, and Robert S Schick. Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics*, 2(3):210–223, 2007.
- [268] Jeffrey D Walker and Steven C Chapra. A client-side web application for interactive environmental simulation modeling. *Environmental Modelling & Software*, 55:49–60, 2014.
- [269] Anthony M Castronova, Jonathan L Goodall, and Mostafa M Elag. Models as web services using the open geospatial consortium (ogc) web processing service (wps) standard. *Environmental Modelling & Software*, 41:72–83, 2013.
- [270] Jacqueline Mize and R Ted Habermann. Automating metadata for dynamic datasets. In *OCEANS 2010 MTS/IEEE SEATTLE*, pages 1–6. IEEE, 2010.
- [271] S Schade and M Craglia. A future sensor web for the environment in Europe. In *24th International Conference on Informatics for Environmental Protection-Enviroinfo2010*, pages 212–219. Aachen, Germany, 2010.
- [272] David Hasenfratz, Olga Saukh, Silvan Sturzenegger, and Lothar Thiele. Participatory air pollution monitoring using smartphones. *Mobile Sensing*, pages 1–5, 2012.

- [273] MI Mead, OAM Popoola, GB Stewart, P Landshoff, M Calleja, M Hayes, JJ Baldovi, MW McLeod, TF Hodgson, and J Dicks. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203, 2013.
- [274] Jill Engel-Cox, Nguyen Thi Kim Oanh, Aaron van Donkelaar, Randall V Martin, and Erica Zell. Toward the next generation of air quality monitoring: Particulate Matter. *Atmospheric Environment*, 80:584–590, 2013.
- [275] Muki Haklay. Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge*, pages 105–122. Springer, 2013.
- [276] Prashant Kumar, Lidia Morawska, Claudio Martani, George Biskos, Marina Neophytou, Silvana Di Sabatino, Margaret Bell, Leslie Norford, and Rex Britter. The rise of low-cost sensing for managing air pollution in cities. *Environment international*, 75:199–205, 2015.
- [277] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi. Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32, 2014.
- [278] CY Huang and CH Wu. Design and Implement AN Interoperable Internet of Things Application Based on AN Extended Ogc Sensorthings Api Standard. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 263–266, 2016.
- [279] J Pradilla, R González, M Esteve, and C Palau. Sensor Observation Service (SOS)/Constrained Application Protocol (CoAP) proxy design. In *2016 18th Mediterranean Electrotechnical Conference (MELECON)*, pages 1–5. IEEE, 2016.
- [280] David Shooter and Peter Brimblecombe. Air quality indexing. *International journal of Environment and Pollution*, 36(1-3):305–323, 2008.
- [281] Daniel Wartenberg. Some considerations for the communication of results of air pollution health effects tracking. *Air Quality, Atmosphere & Health*, 2(4):207–221, 2009.
- [282] Antonella Plaia and Mariantonietta Ruggieri. Air quality indices: a review. *Reviews in Environmental Science and Bio/Technology*, 10(2):165–179, 2011.
- [283] Renjie Chen, Xi Wang, Xia Meng, Jing Hua, Zhijun Zhou, Bingheng Chen, and Haidong Kan. Communicating air pollution-related health risks to the public: an application of the Air Quality Health Index in Shanghai, China. *Environment international*, 51: 168–173, 2013.
- [284] Sef van den Elshout, Karine Léger, and Fabio Nussio. Comparing urban air quality in Europe in real time: A review of existing air quality indices and the proposal of a common alternative. *Environment International*, 34(5):720–726, 2008.
- [285] R Beaumont, RS Hamilton, N Machin, J Perks, and ID Williams. Social awareness of air quality information. *Science of the total environment*, 235(1):319–329, 1999.

-
- [286] Branden B Johnson. Communicating air quality information: Experimental evaluation of alternative formats. *Risk Analysis*, 23(1):91–103, 2003.
 - [287] Judith Bush, Suzanne Moffatt, and Christine E Dunn. Keeping the public informed? Public negotiation of air quality information. *Public Understanding of Science*, 10(2): 213–229, 2001.
 - [288] Jeff Banfield. Rweb: Web-based statistical analysis. *Journal of Statistical Software*, 4 (i01), 1999.
 - [289] Chris Beeley. *Web application development with R using Shiny*. Packt Publishing Ltd, 2016.
 - [290] Yihui Xie. Animation: An R package for creating animations and demonstrating statistical methods. *Journal of Statistical Software*, 53(1):1–27, 2013.
 - [291] Christian Oltra and Roser Sala. Communicating the risks of urban air pollution to the public. A study of urban air pollution information services. *Revista Internacional de Contaminación Ambiental*, 31(4):361–375, 2015.

Appendix A

Table A.1 Linear regression Adjusted R^2 for collocation sampling for the periods 14 January 2015 to 19 January 2015, APM-Aerocet Particulate Matter monitor, CRPM-Customized Real time Particulate Matter monitor, $\sim$ indicate ‘modelled as a function of’

	Model Formula	Adjusted R^2	Significance
Mass reading raw values	$APM_{2.5} \sim CRPM_{2.5}$	0.5583	<0.01
	$APM_{10} \sim CRPM_{10}$	0.2826	<0.01
Mass reading hourly values	$APM_{2.5} \sim CRPM_{2.5}$	0.752	<0.01
	$APM_{10} \sim CRPM_{10}$	0.7831	<0.01
Count reading raw values	$APM_{0.03} \sim CRPM_{2.5}$	0.5453	<0.01
	$APM_{0.05} \sim CRPM_{2.5}$	0.5377	<0.01
	$APM_1 \sim CRPM_{2.5}$	0.4059	<0.01
	$APM_5 \sim CRPM_{2.5}$	0.02065	<0.01
	$APM_{10} \sim CRPM_{2.5}$	0.004437	0.01416
	$APM_{0.03} \sim CRPM_{10}$	0.2181	<0.01
	$APM_{0.05} \sim CRPM_{10}$	0.2347	<0.01
	$APM_1 \sim CRPM_{10}$	0.2862	<0.01
	$APM_5 \sim CRPM_{10}$	0.09908	<0.01
	$APM_{10} \sim CRPM_{10}$	0.04476	<0.01
Count reading hourly values	$APM_{0.03} \sim CRPM_{2.5}$	0.6753	<0.01
	$APM_{0.05} \sim CRPM_{2.5}$	0.6947	<0.01
	$APM_1 \sim CRPM_{2.5}$	0.6406	<0.01
	$APM_5 \sim CRPM_{2.5}$	0.1206	<0.01
	$APM_{10} \sim CRPM_{2.5}$	0.03746	0.04192
	$APM_{0.03} \sim CRPM_{10}$	0.2939	<0.01
	$APM_{0.05} \sim CRPM_{10}$	0.3471	<0.01
	$APM_1 \sim CRPM_{10}$	0.5834	<0.01
	$APM_5 \sim CRPM_{10}$	0.5114	<0.01
	$APM_{10} \sim CRPM_{10}$	0.3087	<0.01