

Hybrid Approaches for Aerosol Source Apportionment

A THESIS

*submitted in partial fulfilment of the requirements
for the award of the dual degree of*

Bachelor of Science - Master of Science

in

Earth and Environmental Sciences

by

Akarsh Ralhan

(16014)



**DEPARTMENT OF EARTH AND ENVIRONMENTAL
SCIENCES**
**INDIAN INSTITUTE OF SCIENCE EDUCATION AND
RESEARCH BHOPAL**
BHOPAL – 462 066

May 2021



भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान भोपाल
Indian Institute of Science Education and Research Bhopal
(Estb. By MHRD, Govt. of India)

CERTIFICATE

This is to certify that **Mr. Akarsh Ralhan**, BS-MS (Dual Degree) student in the Department of Earth and Environmental Sciences, has completed bonafide work on the thesis entitled '**Hybrid Approaches for Aerosol Source Apportionment**' under my supervision and guidance.

May 2021
IISER Bhopal


Dr. Ramya Sunder Raman

Committee Members

Signature

Date

Dr. Sanjeev Kumar Jha



24/05/2021

Dr. Vinee Srivastava



24/05/2021

ACADEMIC INTEGRITY AND COPYRIGHT DISCLAIMER

I hereby declare that this project report is my own work and due acknowledgement has been made wherever the work described is based on the findings of other investigators. This report has not been accepted for the award of any other degree or diploma at IISER Bhopal or any other educational institution. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I certify that all copyrighted material incorporated into this document is in compliance with the Indian Copyright (Amendment) Act (2012) and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and safeguard IISER Bhopal from any claims that may arise from any copyright violation.



May 2021
IISER Bhopal

Akarsh Ralhan

ACKNOWLEDGEMENTS

First, my sincere gratitude to my advisor Dr. Ramya Sunder Raman, for her continuous support from the beginning, which has carved my research interests. It was a privilege to have you as a mentor, great teacher, friend and support system throughout. I have learnt a lot from your infectious enthusiasm, consistent encouragement to achieve perfection and presentation skills. These insights were crucial for completing this work and an invaluable asset for a long journey ahead.

I would like to thank Mr Kushal and Prof. Venkatraman from IIT Bombay for the EI data and constant support. I also want to thank open-source platforms and the entire community of Stack Overflow, without whom it would not be possible to complete this work. Special thanks to all my fellow mates for unravelling the mysteries of ESS with me and all the teachers for the exceptional academic training and wonderful grades.

I am indebted to Krishna and Prem for always being there to solve any tiny problem. Thank you for listening to my rants, crying and pampering me like a baby. This thesis is the result of all the brainstorming sessions, long discussions and debugging. I can never thank Akanksha enough for introducing me to this world, believing in me and being an ever radiating source of guidance (MOM) throughout. I am grateful to all the ACers Samresh, Shivangi, Ankur, Sandeep, Kajal, Anusmita, Deeksha, Diksha and Priyabrata to make Lab a home. Thank you for teaching me to operate all the “toys”, food, canteen trips, keeping my little secrets, online shopping, investment discussions and most awaited TT sessions. They have made life enjoyable and created a positive environment for work in this depressing and challenging time of COVID-19.

The last five years at IISER were bearable and like a perfect dream only because of Pyne, Bhondu, Vips, Butuks, Aditi, YJ, Cheepda, Kallol, PapaRon, Tanay, Radha and Shreyansh: Thank you for uncountable memories and for shaping my evolution in these years. Cheers to endless chill sessions, trips, fights, cooking, DJ nights, awkward moments and love. Thank you, Creepita for increasing the entropy of my life and for all the unpredictable, whimsical and impulsive things we did.

I am forever grateful to my parents and my baby sister. Thank you for your unwavering support, care, love and trust in me. I am blessed to have them as my family. Any accomplishment would not have been possible without them.

Akarsh Ralhan

ABSTRACT

Identifying sources, their characteristics and geographical locations impacting pollutants at a receptor site are essential for any pollution mitigation and management strategy. This thesis presents the development and application of air parcel trajectory ensemble tools and hybrid receptor models for PM_{2.5} source identification and apportionment. This study determined the location of major sources affecting PM_{2.5} over 1) Van Vihar National Park (VVNP), Bhopal and 2) 11 sites in the pan-India national network - COALESCE.

We use the Potential Source Contribution Function (PSCF) and Concentration Weighted Trajectory (CWT) to identify the location of regional sources. The ability of these techniques to identify primary PM_{2.5} source locations were assessed by the determining spatial correlation with a national PM_{2.5} Emissions Inventory (EI) database - SMOG. The results showed that the regions identified by PSCF, CWT and the ensemble of these techniques were highly correlated with EI data (r -values > 1.6). Moreover, the ensemble of the two methods best agreed with the EI for most sources over VVNP. The trajectory ensemble results were also used to determine the geographical location of major regional sources and the sectoral contribution of each sector of the EI to PM_{2.5} over each of the 11 COALESCE sites. Thus, an important outcome of this thesis is a quantitative estimate of primary PM_{2.5} sectoral shares across locations in India. This output can help guide immediate PM_{2.5} source reductions and mitigation actions on a national scale.

In this thesis, a Hybrid receptor model was applied to VVNP PM_{2.5} and its chemical species. The output of two Receptor Models (Positive Matrix Factorisation - PMF5.0 and Chemical Mass Balance -CMB) and one Chemical Transport Model, CTM product (MERRA-2), were hybridised. This hybrid model overcame some limitations of the individual models, including a significant decrease in (unrealistic) zero modelled source contribution and was capable of apportioning mass on days when output was not available from one or more of the individual models. The hybridisation framework developed in this thesis is very

flexible and can easily be applied to a combination of any number of RMs and CTMs and can further be used to constrain RMs to obtain more robust region-specific source profiles. Both of these aspects of hybrid models will strengthen PM_{2.5} source apportionment based policy actions.

LIST OF SYMBOLS OR ABBREVIATIONS

Δ	Change
$^\circ$	Degree
σ	Standard Error
RM	Receptor Model
PMF	Positive Matrix Factorization
CTM	Chemical Transport Model
CMB	Chemical Mass Balance
SA	Source Apportionment
SCI	Spatial Correlation Index
PM	Particulate Matter
PSCF	Potential Source Contribution Function
CWT	Concentrated Weighted Trajectory
RTWC	Residence Time Weighted Trajectory
TPSCF	Total Potential Source Contribution
VVNP	Van Vihar National Park
BS	Bootstrap (Analysis)
DISP	Displacement (Analysis)
ME	Multilinear Engine
EI	Emissions Inventory
BC	Black Carbon
OC	Organic Carbon
BF	Brute Force
TS	Tagged Species
BT	Back Trajectory
MNB	Mean Normalised Bias
AOD	Aerosol Optical Depth

LIST OF FIGURES

Figure 2.1 Types of RM (from (Belis et al. 2013))	15
Figure 2.2 Conjugate gradient method to understand PMF solution search (from USEPA PMF 5.0 User Manual, 2014)	16
Figure 3.1 Location of 11 regional representative sites across India (adapted from Lekinwala et al. 2020)	29
Figure 3.2 Method to assign MEERA-2 OC and EC to three PMF factors	32
Figure 3.3 Flowchart illustrating study design	33
Figure 4.1 PSCF (left) and CWT (right) maps of source locations for various factors of VVNP PM _{2.5} mass to visualise MNB results	37
Figure 4.2 Comparisons between trajectory ensemble and EI locations of Black Carbon (BC) from combustion aerosol sources. The large map (leftmost) is the combined EI map of BC emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost)	38
Figure 4.3 Comparisons between trajectory ensemble and EI locations of SO ₂ from secondary sulphate sources. The large map (leftmost) is the combined EI map of SO ₂ emissions from ELGN and INDS sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlaps with EI (rightmost) .	40
Figure 4.4 Daily time-series plots (a,b) and seasonal box and whiskers plot (d) showing PM _{2.5} source contributions (mass concentration) from each model	44
Figure 4.5 PM _{2.5} Emission Inventory map - 2019 (all sectors)	45
Figure 4.6 PSCF (left) and overlap with EI (right) for sites in the Eastern cluster	46
Figure 4.7 PSCF (left) and overlap with EI (right) for sites in the Northern Cluster	48
Figure 4.8 PSCF (left) and overlap with EI (right) for sites in the Southern and Central cluster	49
Figure 4.9 Sector-wise contribution to primary PM _{2.5} mass over 11 site COALESCE network across India for 2019 (Ensemble)	51
Figure 5.1 Schematic of hybrid CTM-RM constrained receptor model framework	54

LIST OF TABLES

Table 2.1 SMOG-India Emissions Inventory sectors and associated categories	25
Table 3.1 Major factors (sources) and potential source locations identified by Kumar and Sunder Raman, 2020 over VVNP, Bhopal	28
Table 3.2 Source profiles used for each source category to run CMB 8.2 model	31
Table 4.1 MNB values PSCF vs CWT geographical source locations of VVNP PM _{2.5} mass	36
Table 4.2 Number of zero-impact days for each source calculated by three SA models and the hybrid model	42
Table 4.3 Results (r-statistic values) of Spatial Correlation hypothesis test for each site of the COALESCE network	50

CONTENTS

Certificate	1
Academic Integrity and Copyright Disclaimer	2
Acknowledgements	3
Abstract	4
List of Symbols and Abbreviations	5
List of Figures	6
List of Tables	7
Introduction	11
1.1 Motivation	12
1.2 Objectives	13
Context and Techniques	14
2.1 Source Apportionment	14
2.1.1 Positive Matrix Factorization, PMF 5.0	15
2.1.2 Chemical Mass Balance, CMB 8.2	17
2.1.3 MERRA-2 Chemical Species	19
2.1.4 Hybrid Model	20
2.2. HYSPLIT Air Parcel Trajectories	21
2.3 Trajectory Ensemble Techniques	22
2.3.1 Potential Source Contribution Function, PSCF	23
2.3.2 Concentration Weighted Trajectories, CWT	24
2.4 Emissions Inventory, EI	24
2.5 Spatial Correlation Index, SCI	26
Study Design and Methodology	28
3.1 Input Data Sets	28

3.1.1 Van Vihar National Park (VVNP) Mass and Chemical Species	28
3.1.2 COALESCE Network PM2.5 Measurements	29
3.1.3 Receptor Model Source Profiles	30
3.1.4 Chemical Transport Model, CTM Source Apportionment	32
3.2 Study Design	33
Results and Discussion	35
4.1 Van Vihar National Park (VVNP) Bhopal	35
4.1.1 Comparison Of Trajectory Ensemble Derived Source Locations for VVNP Aerosol	35
4.1.2 Reconciling Trajectory Ensemble Source Locations With Emissions Inventory	37
4.1.3 Hybrid RM-CTM source apportionment	41
4.2 Pan-India COALESCE Network (11 Sites)	44
Summary and Conclusions	52
Appendices	55
Appendix I - Trajectory Ensemble Results - VVNP	55
Appendix II- Hybrid Model	65
Appendix III - Trajectory Ensemble Results - COALESCE Network	70
References	74

1. Introduction

Identification of sources responsible for the emission of various pollutants is the primary objective of air quality management, which seeks to mitigate air pollution and its associated health and climate effects. Receptor modelling tools like USEPA PMF (Positive Matrix Factorization), UNMIX, CMB (Chemical Mass Balance) and other factor analysis based approaches typically solve the chemical mass balance equation to identify and estimate the contribution of sources (Paatero and Tapper 1993; Belis et al. 2020; Hopke 2016; Keeler 1987). However, these models do not provide information about the geographical location of the sources and/or preferred transport pathways to the receptor site. To address this limitation, air-parcel back trajectory models, in conjunction with the receptor model source contribution output, are typically employed.

These hybrid models calculate the position of the air parcel sampled at the receptor, backwards in time for a specified duration, to identify the geographical location of the sources of pollutants (Hopke 2016). Ensembles of trajectories can improve the statistical reliability of estimating the likely source locations and/or preferred transport pathways. Popular ensemble approaches include the Potential Source Contribution Function (PSCF), Concentration Weighted Trajectory (CWT), Residence Time Weighted Concentration (RTWC), Semi-Quantitative Trajectory Bias Analysis (SQTBA) and others (Hopke 2016; Masiol et al. 2019; Zhou, Hopke, and Liu 2004).

Another method is to construct emission inventories (EI) that map and estimate the geographical distribution of sources associated with different pollutant species and quantify the strength of emissions (Henry et al. 1984; Gordon 1988). These inventories are typically input to Chemical Transport Models (CTM) to estimate pollutant concentrations at various locations (Belis et al. 2020; Venkataraman et al. 2018). It stands to reason that the geographical locations identified by air parcel trajectory ensemble methods and those identified on the EI, for primary pollutants, should reconcile if both of these approaches

provide expected outputs within their uncertainty bounds. Spatial Correlation Indices (SCI) can be used to assess the degree of correspondence between the receptor model/trajectory ensemble results and the known EI maps (Hopke, et al., 2005). These indices provide a quantitative basis to further strengthen our understanding of source location and their contribution.

Receptor modelling tools like PMF and CMB use the concentration of pollutants received at the receptor site to determine their potential sources and quantitatively apportion them to these sources. These methods apply different algorithms to solve the general chemical mass balance equation for the measured chemical concentrations. CMB employs an Effective Variance algorithm (EV-CMB) that uses predefined source profiles (Watson et al. 1990). In contrast, PMF employs a Multilinear Engine (ME-2), that uses an iterative conjugate gradient algorithm to approach a global minimum without using source profiles (Paatero and Tapper 1993). However, apportionment by EV-CMB or PMF alone suffers from limitations in resolving collinear sources and their contributions. The other possible solution is to track the pollution released from the source to the receptor using Chemical Transport Models (CTM) like the Weather Research Forecasting coupled with Chemistry (WRF-Chem) model (Grell et al. 2005). Such models provide time-varying concentrations of several sources, and coupling Receptor Models (CMB and PMF) outputs with the CTM output iteratively is believed to generate more realistic source profiles and contributions than either one of these individual approaches (Lee et al. 2009).

1.1 Motivation

PM_{2.5} has received very high scientific and media attention in recent times. It is in the spotlight because of the associated adverse effects on health and high uncertainty associated with its climate effects (Feng et al. 2016; Tai et al. 2010). Identification of sources responsible for the emission of various pollutants is the primary objective of air quality management and mitigation strategies. HYSPLIT and other air-parcel trajectory ensemble models are widely used as a tool to identify the source locations (Han et al. 2007; Hsu et al. 2003). The motivation for

this study is to determine the application and assess the effectiveness of these models in identifying potential source locations of PM_{2.5} measured over various locations. Also, it is necessary to identify the type of sources affecting the receptor site along with the potential source location. The aerosol research community uses both top-down (RMs), and bottom-up (CTMs) approaches to identify the characteristics of these sources. We would like to assess the power of the hybrid model developed by combining these two approaches and its ability to overcome the limitations associated with each of the individual approaches.

1.2 Objectives

Objective 1: Assessment of the effectiveness of trajectory ensemble modelling techniques in identifying potential primary aerosol source locations by comparing them with emission inventories. Specific objectives include:

- 1.1. Intercomparison of two trajectory ensemble receptor techniques (PSCF and CWT)
- 1.2. Assessment of Spatial Correlation (SC) between emissions inventory maps and trajectory receptor model maps for each of the PM_{2.5} sources resolved over Van Vihar National Park, VVNP, in a previous study (Kumar and Sunder Raman, 2020)
- 1.3. To determine the SC between emissions inventory maps for PM_{2.5} and trajectory ensemble maps for the entire COALESCE Network (Venkataraman et al. 2020) - 11 locations across India.

Objective 2: To assess the potential of hybrid approaches in aerosol source apportionment of PM_{2.5} mass over VVNP, Bhopal

- 2.1 To demonstrate the framework of hybridising receptor (CMB-PMF) and CTM (MERRA-2) model outputs for source apportionment
- 2.2 To inter-compare the results of the hybrid approach with the receptor and CTM methods of source apportionment.

2. Context and Techniques

We have used different tools and methods to achieve the objectives of this thesis. In this chapter, a review of the fundamentals and relevance of each technique, along with modifications adopted to accomplish the thesis goals utilising available resources, are presented.

2.1 Source Apportionment

Source Apportionment (SA) is a method of assigning the total pollutant (typically, mass concentration) measured over a location to individual sources, often quantifying their contributions to the pollutant mass. Conventionally, there are two ways to accomplish this task: applying source-oriented models (chemical transport models, CTM) or receptor oriented models to the pollutant dataset. The first method tracks the pollutant from its point of release at the source to the receptor by superimposing meteorology, associated chemical reactions and transformations on the pollutant. These CTM models provide excellent spatial resolution but are computationally intensive. These models are also highly dependent on input data quality, especially the emissions inventory (EI), which itself makes several assumptions and approximations. These uncertainties are further propagated during the modelling effort.

Receptor oriented models (RM) use chemical concentrations as a fingerprint to quantify the contribution of sources to each sample. The effect of meteorology and physio-chemical transformation of pollutants during transport from source to the receptor, although not broken down into separate components, is captured in the receptor location's measurements. These models provide a good estimate of associated uncertainties in the source apportioned contributions using various approaches. The ability of RMs to apportion sources and provide insights on atmospheric transformations increases with increasing time resolution of the measurements. However, even for the most commonly applied time-integrated (typically 24 hours) measurements for regulatory purposes, the RM approach

requires extensive chemical analysis and fieldwork and is not applicable for non-conservative atmospheric pollutants (Belis et al. 2013).

All RMs are mathematical tools to solve a general mass balance equation which can be expressed as

$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (2.1)$$

where i is the number of samples; j is chemical species; e_{ij} is residual for each sample/species; p the number of factors (or known sources); the species profile f of each source; and mass g is the contribution by each factor to every individual sample (Norris et al. 2008).

Based on the amount of information required, RMs can be divided into a complete spectrum of models (Figure 2.1). Of all these models, we can broadly classify the widely used models into two domains: a) If f matrix (source profiles) is known but g is unknown - CMB b) No information on f or g matrices is available- PMF and UNMIX. Following is a brief description of the RMs, CTM, and other ancillary models and methods used to achieve the objectives of this thesis.

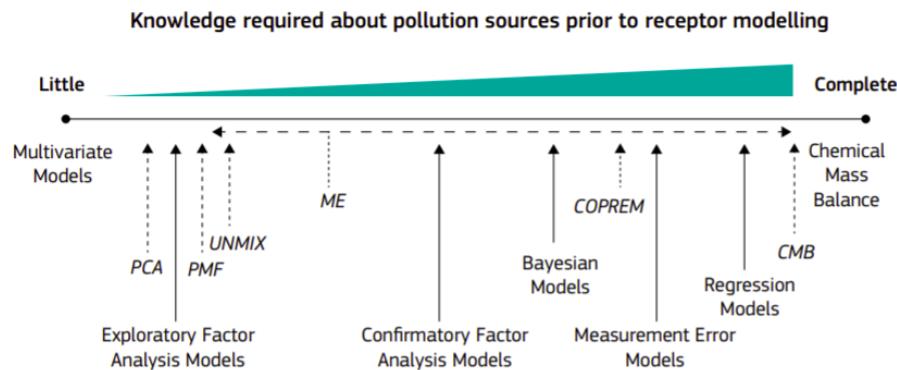


Figure 2.1 - Types of RM (from (Belis et al. 2013)

2.1.1 Positive Matrix Factorization, PMF 5.0

Positive Matrix Factorization (PMF) is a powerful SA technique that solves the general chemical mass balance equation using a weighted, constrained,

least-squares approach (Paatero and Tapper 1993; Paatero et al. 2005). The model assumes p distinct sources (termed factors) that influence a receptor site. The linear combination of the impacts from individual factors gives the observed concentrations of different chemical species at the receptor (Eq 2.1).

PMF minimises the objective function ‘ Q ’, i.e. the sum of the squares of the residuals, weighted inversely by the error estimates of measurements using a Multilinear Engine (ME) algorithm. The objective function Q is defined as

$$Q = \sum_{i=1}^n \sum_{j=1}^m \left[\frac{x_{ij} - \sum_{k=1}^p g_{ik} f_{kj}}{u_{ij}} \right]^2 \quad (2.2)$$

The ME-2 algorithm starts with defining the solution space (using input observations and specification). It begins traversing this solution space using the conjugate gradient approach (Brown et al. 2015), starting from a random point in the multidimensional space (Figure 2.2). The model’s search for the global minima (optimum solution -minimum Q) goes from coarser to a finer scale defined by the change in $Q(dQ)$ in consecutive steps. The model can reach a local maximum in the space due to randomness in the starting point; therefore, the model is run multiple times (minimum 100) with different starting points to obtain values close to the global minimum.

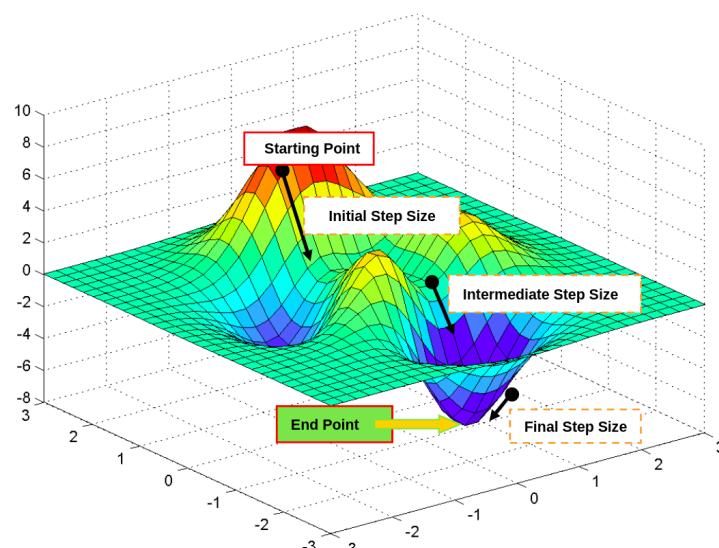


Figure 2.2. Conjugate Gradient Method to understand PMF solution search (from USEPA PMF 5.0 User Manual, 2014)

Variability in the PMF solutions can arise due to chemical transformations or changes in the process which indirectly affect the model by changing factor profiles for different PMF runs (Paatero et al. 2014). The uncertainty associated with this variability can be estimated using three methods: 1) Bootstrap (BS) analysis - to identify the impact of small subsets of observations. It considers the effects of random errors and rotational ambiguity (Infinite solution to CMB problem). 2) Displacement (DISP) analysis - sensitivity of the solution to minor changes. It only takes into account the rotational ambiguity. 3) BS-DISP analysis - a hybrid approach that accounts for the effects of both random errors and rotational ambiguity. It is a more robust technique than either of the two (Paatero et al. 2014; Brown et al. 2015).

PMF can handle missing and below detection limit data by adjusting the error estimates (Polissar et al. 1998; Kim 2004; Dutton et al. 2010). PMF weighs individual points and accounts for uncertainty in every sample (which accounts for the confidence in the measurement by propagating sampling and chemical analyses uncertainties). There is no need to input source profiles or select relevant species manually. The model gives a range of solutions for interpretation along with better isolation of sources that are minor contributors to mass. However, the model requires a large data set and generally isolates secondary factors from precursor emissions, and are not actual sources. These secondary factors must be interpreted as the source representing atmospheric formation/transformations of precursor species into secondary particles.

2.1.2 Chemical Mass Balance, CMB 8.2

The CMB model estimates emission source contributions to receptors by solving the mass balance equation (Eq 2.1) using an effective-variance least-squares approach (e.g., Watson et al. 1994; Watson et al. 1990; Watson et al. 1997; Coulter 2004). CMB requires the species fractional compositions of source categories (source profiles) and species concentration data at the receptor site as inputs. Ultimately, a weighted least-squares approach is used to find the source contributions (g).

The model starts with a set of linear equations to express the chemical species' concentration (measured at the receptor site) and expresses them as the sum of the product of source compositions and source contributions. The resulting set of equations has many possible solutions (overdetermined); the number of chemical species is greater than the number of source types. Uncertainty estimates (source profiles and concentration measurements) and physical constraints are imposed to obtain the optimum solution. These uncertainties are used as weights in an effective variance weighting scheme, i.e. species with higher uncertainties are given less weightage than species with lower relative uncertainties (Watson et al. 1990).

It is not always possible to get the exact profiles for the site; the sources can have different profiles, or composite profiles are needed. Various statistical parameters are used to determine the set of profiles (to get an optimum solution) and test the model performance (details presented in relevant section of results). The statistics used to assess model performance include the percentage mass explained by the model, r^2 and χ^2 (acceptable ranges and details are given in Table A1).

The major limitation of the model is associated with the strong dependence of model output on the input source profiles. The profiles are based on measurements followed by chemical analysis; this multi-step process can introduce errors. They are also unable to represent the ageing of the source material (which is how they will be present when measured at a receptor site). The profiles might have significant errors if employed for different geographical locations and times (than those at which the profile was obtained). CMB uses fixed source profiles for the entire data set, but in reality, source profiles substantially vary by region, time, and ambient conditions (e.g., the gasoline composition and engine efficiency changes with seasons). CMB uses primary source profiles; therefore, it is difficult to correctly assign secondary sources (e.g., sulfates, nitrates) to the receptor site.

2.1.3 MERRA-2 Chemical Species

In addition to quantifying the spatio-temporal distribution of aerosols from various sources, CTMs with a chemistry module can also explicitly quantify the distribution of particles formed from gaseous species emission from various sources on the PM at a receptor site, a limitation of RM based approaches. However, running a CTM is resource-intensive and out of the scope of this study. Therefore, in this study, to demonstrate the hybridization (RM and CTM) methodology, we used species wise $\text{PM}_{2.5}$ concentrations as obtained from a global reanalysis product, Modern-Era Retrospective analysis for Research and Application (MERRA-2), (CTM constrained with satellite observations). MERRA-2 aerosol data is available from 1978 at a spatial resolution of $0.5^\circ \times 0.625^\circ$ generated using GEOS-5 atmospheric model radiatively coupled with GOCART aerosol module to simulate source, sink and chemistry of 15 aerosol species (externally mixed) - hydrophobic and hydrophilic BC and OC, 5 size bins of both dust and sea salt each and, sulfate aerosols (Gelaro et al. 2017; Randles et al. 2017).

Emissions and transportation of both dust and sea salt are wind speed-dependent, while carbonaceous and sulfate aerosols are from fossil fuel combustion, biofuel consumption and biomass burning, along with other biogenic sources of OC. In contrast, secondary sulfate is obtained from the chemical oxidation of SO_2 and dimethyl sulphide (DMS) and is transported using winds.

The total columnar aerosol mass generated by the model is constrained using bias and non-bias-corrected satellite AOD data using Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging SpectroRadiometer (MISR), Advanced Very High-Resolution Radiometer (AVHRR), and AERONET stations (Buchard et al. 2017). When compared against ground data, MERRA-2 total $\text{PM}_{2.5}$ performed well ($r \sim 0.6$) over India (Navinya, Vinoj, and Pandey 2020) and ($r \sim 0.59$) over North China (Song et al. 2018). The chemical species obtained from this product were used in this thesis to demonstrate the hybridisation technique for aerosol SA discussed in the next section.

2.1.4 Hybrid Model

All commonly used SA models have limitations that affect their applicability to identifying certain sources. In general, RMs identify sources/factors accounting for 60-80% of inventoried emissions, the results are dependent on the methods used, and incomplete information about the sources introduces both errors and bias in the model output (Barregard et al. 2006; Ostro et al. 2007; Sarnat et al. 2008). SA results show excessive inter-day variability, unrealistic zero impact of significant sources on many days, and sometimes physically unrealistic large shifts in relative contributions of various sources on consecutive days. Each RM has its own limitations as well. CMB highly depends on the accuracy of source profiles and cannot consider the temporal variability of emissions while PMF can produce factors that include contributions from multiple sources that are difficult to separate (Lee et al. 2009).

These limitations of RMs are overcome by using detailed, physically based CTMs coupled with chemistry. However, the computing resource requirements inhibit their use over long periods and they are affected by uncertainties associated with emission rates and inventories, model parameterizations, the accuracy of chemical mechanisms, aerosol mixing state assumptions, and meteorological inputs. A key limitation is that due to the coarse-scale of meteorological inputs and temporally aggregated emissions, these models show very little day-to-day variability. Chemical mechanisms limit their ability to generate secondary aerosols, especially organic secondary particles, even within two-three orders of magnitude of observation-based estimates. However, these models offer broad, continuous spatial coverage of aerosol estimates from various sources.

Model evaluation is difficult for both CTMs and RMs. CTMs can be evaluated by their ability to simulate the spatio-temporal concentration trends of the observation. RMs are tested on simulated data sets, but the availability and accuracy of these data sets in emulating actual atmospheric concentrations and species inter-relationships are generally questionable (Brinkman et al. 2006).

One way to overcome these limitations of RMs and CTMs is to hybridize them (Lee et al. 2009; Maier et al. 2013; Ivey et al. 2017; Balachandran et al. 2012). The hybrid approach results have shown improvement in several parameters, including fewer days with zero contribution from known sources and reduction in day-to-day variability (compared to individual RMs).

To achieve better estimates of source contributions of PM_{2.5} measured over VVNP, we have taken the ensemble of two RMs results (CMB 8.2 and PMF 5) and one satellite AOD constrained reanalysis data set (MERRA-2) by obtaining a weighted source contribution given by:

$$\bar{S}_j(t_k) = \frac{\sum_{l=1}^L w_{jl}(t_k) \cdot S_{lj}(t_k)}{\sum_{l=1}^L w_{jl}(t_k)} \quad (2.3)$$

where $S_j(t_k)$ is the ensemble-calculated impact of source j (in $\mu\text{g}/\text{m}^3$) at time t_k , $S_{lj}(t_k)$ is the impact developed by model l (MERRA-2, CMB, PMF), and $w_{jl}(t_k)$ is the weight given as

$$w_{jl} = \frac{1}{\sigma_{S_{lj}}} \quad (2.4)$$

2.2. HYSPLIT Air Parcel Trajectories

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) (Draxler and Rolph 2010) determines the extent and direction of movement of the air parcel, and subsequently air pollutants, by calculating back trajectories. It is called hybrid because it uses different reference frames; advection and diffusion equations are solved in the Lagrangian frame of reference. The Eulerian frame is used to calculate concentrations in each grid.

HYSPLIT can also calculate dispersion, chemical transformation, and deposition of atmospheric pollutants in different modes. The model can run independently as puff, particle or the hybrid of the two, in which particle behaviour is treated in the vertical direction and Lagrangian puff in the horizontal; this modified approach is the most efficient and cost-effective.

For the study, HYSPLIT is used to obtain the back trajectory end-point files. Each trajectory is traced back for 120 hours (5 days), and a new trajectory is initiated every two hours from the vertical level of 500m above ground level. Daily GDAS meteorological data is used with 0.5-degree spatial resolution.

Back Trajectories are determined in two steps. First, the position vector P is guessed in the next time step using 3D velocity vector V obtained from the meteorological data set.

$$P'(t + \Delta t) = P(t) + V(P, t)\Delta t \quad (2.5)$$

The final position is obtained from the average velocities determined in current and first guess positions.

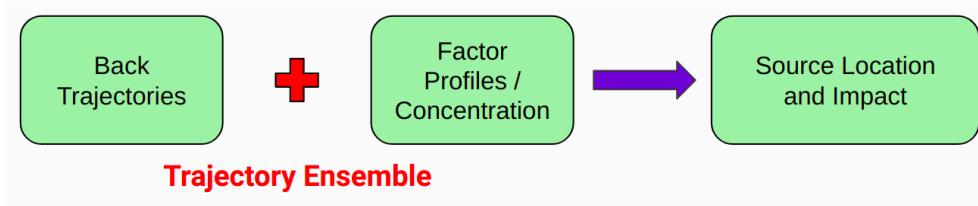
$$P(t + \Delta t) = P(t) + 0.5[V(P, t) + V(P', t + \Delta t)]\Delta t \quad (2.6)$$

2.3 Trajectory Ensemble Techniques

Air parcel trajectories can be calculated forward in time, starting from source to receptor. Another analogous technique is also widely used in which air-parcel transport is tracked backwards in time from the receptor to potential source locations, i.e., back trajectory (BT) analysis.

This analysis is based on the idea that the BT goes far enough in time with reasonable uncertainty in its position to capture the source's location for the pollutant of interest. This is so because the source must have been somewhere in the region over which the air parcel has passed to allow it to take up pollutants and transport them to the receptor site (if the pollutant source is outside the trajectory domain in a given time, then the region still indicates a preferred transport pathway for the pollutant). We assume that once the pollutants are incorporated into the air parcel, they are transported to the receptor site along with the parcel. The concentration data obtained at the receptor site through ambient sampling followed by chemical analysis is used to weigh these trajectories. If we set a concentration threshold, the trajectories corresponding to high concentration days have a high probability of having passed over the pollutant source. By

combining these two sets of information, we can comment on the impact and geographical location of the sources.



The different statistical trajectory ensemble methods used are discussed below

2.3.1 Potential Source Contribution Function, PSCF

PSCF is a popular technique (Zong et al. 2018; Pekney et al. 2006; Kumar and Sunder Raman 2020) developed by (Ashbaugh, et al. 1985) to identify the spatial distribution and location of potential sources of pollution at a regional scale. The total number of air parcel back trajectory endpoints over the ij^{th} cell are represented by n_{ij} . In the same grid cell, the number of endpoints corresponding to pollutant concentration higher than an arbitrary threshold is represented as m_{ij} . The PSCF value for the ij^{th} cell is defined as

The PSCF value for the ij^{th} cell is defined as

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \quad (2.7)$$

Thus, the PSCF is a conditional probability that describes the spatial distribution of source locations using BTs. Grid cells corresponding to high PSCF values are the potential source areas or the potential pollution pathway.

High PSCF values can be obtained in grid cells with the small n_{ij} (denominator). To reduce the associated uncertainty caused by small values of n_{ij} an arbitrary weight function $W(n_{ij})$ is multiplied with the PSCF value at each grid cell (Tellaetxe and Carslaw 2014; Hopke et al. 1995; Polissar et al. 1999).

There are two major limitations of this analysis: (1) Inability to estimate the geographical location of each source of the fine PM to the atmosphere. The species that can be successfully transported to the receptor site can only be identified. (2) It fails to identify possible sources outside the back trajectory domain (Kim et al. 2016; Wang et al. 2006).

2.3.2 Concentration Weighted Trajectories, CWT

CWT was developed by (Hsu et al. 2003) to overcome the limitations of PSCF and distinguish sources based on their relative strength by multiplying the measured total concentration with the corresponding trajectories. Similar to PSCF the spatial domain is divided into a 2D array of grids, and each grid cell gets a weighted concentration as follows

$$\text{CWT}_{i,j} = \frac{\sum_{l=1}^M C_l \tau_{i,j,l}}{\sum_{l=1}^M \tau_{i,j,l}} \quad (2.8)$$

C_l is the measured PM concentration tagged with trajectory l, M is the total number of trajectories, and $\tau_{i,j,l}$ is the number of trajectory endpoints in a grid corresponding to the C_l sample.

2.4 Emissions Inventory, EI

Emission Inventories (EI) are detailed accounting of the emissions released from multiple source categories in a particular geographical area and specific time frame (generally a year). The rate of activity for a particular source (such as fuel consumed, industrial production, diesel consumed in vehicles etc.) is multiplied with the emission factor (i.e. the amount of pollutant emitted per unit of activity) to estimate emissions. EI serves as an important tool to identify the sources of pollutants, understanding air quality/ atmospheric chemistry, and the impact on regional climate. This study used the Speciated Multi-pollutant Generator (SMOG) data set developed by a research group from IIT Bombay (Sadavarte and Venkataraman 2014; Pandey et al. 2014) for comparing the source locations identified using ensemble methods.

SMOG presents spatiotemporally resolved monthly emissions over India in the form of $0.25^\circ \times 0.25^\circ$ grids. The basket of pollutants considered in this analysis includes BC, OC, $\text{PM}_{2.5}$, SO_2 , NOx, NMVOCs, CO and Mineral matter. While the data set includes both historical emissions - “1996-2015” (Sadavarte and Venkataraman 2014; Pandey et al. 2014) and future projections - “2015-2030”

(Venkataraman et al. 2018; Tibrewal and Venkataraman 2020), we have used the datasets for the years 2012, 2013 and 2019 for this study.

The emissions are prepared using the bottom-up approach using fuel consumption at the sectoral level and technology-linked emission factors. Activity information was mostly collected from government resources. In the absence of availability of direct fuel consumption, sector-specific methodologies were developed to estimate fuel consumption. The historical emissions were extrapolated from the base year (2010) using the reported fuel consumption trends (Sadavarte and Venkataraman 2014; Pandey et al. 2014), while the future projections involved estimating future activities from the reported sectoral growth rate.

Table 2.1. SMOG-India Emissions Inventory sectors and associated categories

Sector Code	Sector Name	Source categories
REBM	Residential Biomass	Res. Cooking Biomass Res. Space Heating Res. Water Heating
AGRIB	Agricultural residue burning	Open burning
ELGN	Electricity generation	Thermal Power Plant
INDS	Industry	Heavy Industry Light Industry
BRIC	Brick Production	Brick Production
TRAN	Transport	On-road Gasoline/CNG On-Road Diesel Railways
DDSL	Dispersed Diesel	Agri. Tractors Agri. Pumps Diesel Gensets
OTHR	Others	Res. Cooking Kero/LPG Res. Lighting Informal Industry Trash burning Urban fugitive dust

Future activities were then split into various technologies under different penetration scenarios of clean technologies, and emissions are reported using the corresponding emission factors (Venkataraman et al. 2018; Tibrewal and Venkataraman 2020). The inventory reports emissions for 8 different categories after grouping several source categories given in Table 2.1.

2.5 Spatial Correlation Index, SCI

To spatially compare the pair of maps generated using various trajectory ensemble techniques, we have used a spatial correlation index introduced by (Hopke et al. 2005) and extensively used in other studies (Choi et al. 2008; Han et al. 2007; Seo et al. 2015). Each map (trajectory ensemble and EI) is considered to be a 2D matrix defined by the latitude and longitude (x, y). For a given location, i , (x_i, y_i) are the coordinates and p_i and q_i are the emissions and predicted probability value (PSCF or CWT), respectively. The matrices p and q are first normalised (to get values from 0 to 1) and then vectorised. To describe the spatial association between two maps, an auto-correlation statistic, Γ , is defined, which is given by:

$$\Gamma = \sum_{ij} D_{ij} C_{ij} \quad (2.9)$$

$$D_{ij} = 1 / \left[(x_i - x_j)^2 + (y_i - y_j)^2 + 1 \right] \quad (2.10)$$

$$C_{ij} = [(p_i - \bar{p})(q_j - \bar{q}) + (p_j - \bar{p})(q_i - \bar{q})]/2 \quad (2.11)$$

where \bar{p} / \bar{q} are the mean values of the complete vector. The product of either very high or very low values of f and g have the most significant impact on C . Weights D are based on the spatial proximity of two points compared, closer the points more the weightage.

Further, we determined the association using hypothesis testing. The null hypothesis (H_0): p and q are not spatially correlated. The first two moments of Γ are defined as

$$r = (\Gamma - E(\Gamma)) / (E(\Gamma^2) - (E(\Gamma))^2)^{1/2} \quad (2.12)$$

$$E_R(\Gamma) = S_0 T_0 / n(n - 1)$$

$$E_R(\Gamma^2) = S_1 T_1 / 2n^{(2)} + (S_2 - 2S_1)(T_2 - 2T_1) / 4n^{(3)} + (S_0^2 + S_1 - S_2)(T_0^2 + T_1 - T_2) / n^{(4)}$$

(2.13)

Where,

$$S_0 = \sum_{ij} D_{ij} \quad S_1 = \frac{1}{2} \sum_{ij} (D_{ij} + D_{ji})^2 \quad S_2 = \sum_i (D_{i,a} + D_{i,b})^2 \quad D_{i,a} = \sum_j D_{ij}$$

$$D_{i,b} = \sum_j D_{ji}$$

$$n^{(b)} = n(n-1)(n-2)\dots(n-b+1)$$

where T_k is the same as S_k but with D_{ij} replaced by C_{ij} . The variable r is transformed to an $N(0,1)$ distribution and the final value is compared with 1.6, i.e. 95th percentile of the one-tailed test; r values greater than 1.6 will result in rejection of H_0 , indicating that the two maps are statistically correlated.

3. Study Design and Methodology

3.1 Input Data Sets

3.1.1 Van Vihar National Park (VVNP) Mass and Chemical Species

Van Vihar National Park (23.23°N 77.37°E), situated in Bhopal, covering an area of over 5 km^2 , surrounded by a large lake ($\sim 31 \text{ km}^2$) on three sides and a hill on the fourth, was selected as the sampling site to assess the impacts of particulate pollution over the National Park. Site selections and sampling details are discussed elsewhere (Kumar and Sunder Raman, 2020 and references cited therein). In this thesis, we have used PMF5 results previously obtained (Kumar and Sunder Raman 2020), utilizing 240 twelve-hour integrated samples that were collected during 2012 and 2013. Briefly, the fine PM mass and chemical concentration of over 35 chemical species was apportioned using PMF5 into 7 major factors, which are equivalent to the pollutant sources (g matrix), and this matrix was used as an input to the trajectory ensemble assessment and hybrid RMs-CTM.

Table 3.1. Major factors (sources) and potential source locations identified by Kumar and Sunder Raman, 2020 over VVNP, Bhopal

Sources	Potential Source Regions
Biomass burning aerosol	Stubble burning (Madhya Pradesh, Chattisgarh) + crop residue burning (Madhya Pradesh, Chattisgarh, Rajasthan, Haryana)
Combustion aerosol	Thermal power plants (Madhya Pradesh, Chattisgarh, Delhi)
Secondary Sulphate	Industrial belt + thermal Power plants (Madhya Pradesh, Chattisgarh, Rajasthan, Haryana)
Secondary Nitrate	Precursor emissions for secondary nitrate

Re-suspended crustal dust	Arabian Sea, Middle East
Pyrolysis carbon-rich aerosols	Several non-mineral industries (Madhya Pradesh)
Sea salt	The Arabian Sea as a source region

3.1.2 COALESCE Network PM_{2.5} Measurements

COALESCE (Carbonaceous Aerosols Emission, Source apportionment and Climate Impacts) project aims to understand the atmospheric chemistry of carbonaceous aerosols, identify their primary sources and their impact on regional air quality and climate of India (Venkataraman et al. 2020). Under this multi-institutional project, a nationally coordinated network of 11 regional representative sites was strategically selected across India to capture regional PM_{2.5} (Lekinwala et al. 2020). 24 hour - integrated PM_{2.5} sampling was conducted every alternate day in all these sites during 2019 and 2020. We use the PM_{2.5} mass concentration for 2019 in this thesis.

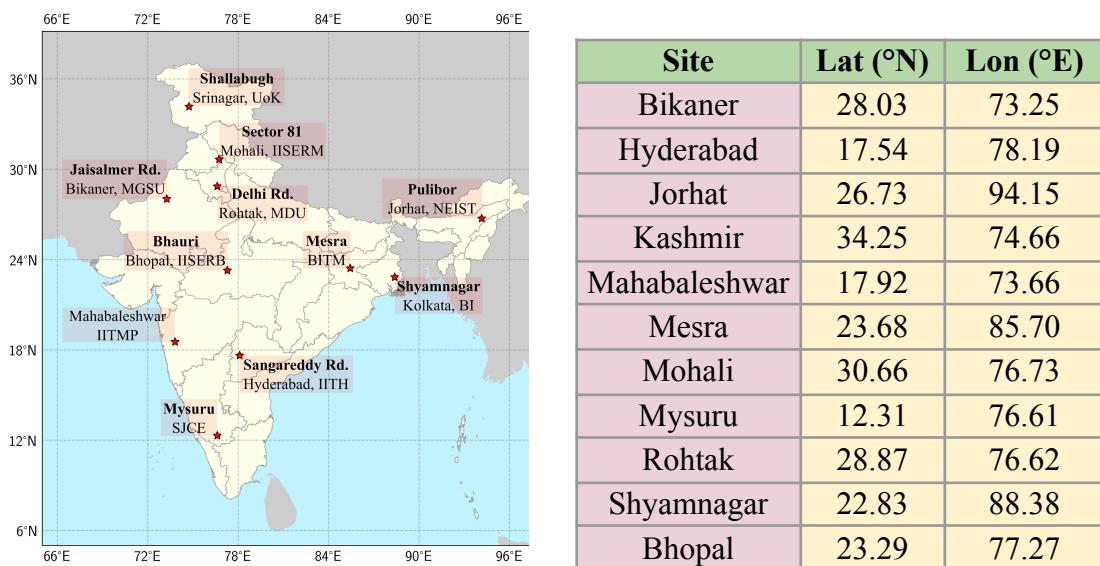


Figure 3.1. Location of 11 regional representative sites across India (adapted from (Lekinwala et al. 2020))

3.1.3 Receptor Model Source Profiles

Source profiles are the data set consisting of fractional mass abundances of measured chemical species (along with associated measurement uncertainty) relative to the PM mass from the emissions of a given source (Watson et al. 2001). They are typically generated for 1) developing emission inventories (Kuykendal et al. 1990; Cass and McRae, 1983; Chow et al. 2004), 2) inputs to CMB receptor model for SA (Watson et al. 1990, 2001) and 3) estimating emissions of hazardous and toxic pollutants (Chow et al. 2004).

The profiles vary across source categories, and the chemical signatures depend on a range of factors. For example, vehicle emission profiles are influenced by factors like fuel type, operating conditions, combustion technologies and season, whereas in the case of soil and dust profiles, they highly depend on local geology and origin (e.g., Bayram, 2008). Hence, it is recommended to develop local source profiles for use in CMB to obtain meaningful SA.

The Central Pollution Control Board (CPCB) in India provides $\text{PM}_{2.5}$ source profiles for some sources (CPCB, 2008). But these profiles are obtained from six large cities (Delhi, Mumbai, Chennai, Bangalore, Pune and Kanpur) from measurements made 8 hours a day for 90 days during 2007-2008, and fail to provide representative data for other semi-urban and rural locations in India. The profiles are outdated and updated versions are not yet released. Thus, SA studies from India using CMB RM mostly use the profiles from the USEPA SPECIATE data set. However, this introduces large uncertainties and limits the model ability to apportion mass accurately given that semi-urban and rural locations in the US do not really capture source profiles of comparable locations in India (Karar and Gupta, 2007; Matawle et al. 2014; Chakraborty and Gupta, 2010; Patil et al. 2013; Pant and Harrison, 2012; Banerjee et al. 2015).

In this study, attempts were made to obtain the best fit CMB using the source profiles from CPCB India at sites close to VVNP. The fits obtained were poor; therefore, we have used a combination of USEPA SPECIATE-2 data set Version 4.4 (<https://www.epa.gov/air-emissions-modeling/speciate-2>) and the

national profiles, wherever necessary, to obtain the best fit. PMF SA results from the VVNP study (Kumar and Sunder Raman 2020) provided important insights to select the source categories/profiles for running the model. Factor profiles obtained from PMF performed extremely well in explaining the measured mass when used as the input f matrix to the CMB model; therefore an attempt has been made to generate aggregate profiles, using available source profiles in literature across locations (Table 3.2), to mimic VVNP PMF profiles to whatever extent possible. Further, the profiles were selected to make composites after analysing the model performance statistics.

Table 3.2. Source profiles used for each source category to run CMB 8.2 model

Source Category	Data Source
Biomass burning aerosol	Bagasse combustion and Agricultural waste burning (Patil et al. 2013) + Bagasse combustion (CPCB)
Combustion aerosol	Powerplant Coal Combustion (USEPA SPECIATE 4.3) + Coal-based thermal power plant (Bano et al. 2018) + 2 Wheeler gasoline and 4 Wheeler diesel vehicles (Pervez et al. 2018)
Sea salt	PMF5 output
Secondary Sulphate	Secondary Sulphate (USEPA SPECIATE 4.3)
Secondary Nitrate	Ammonium Nitrate (USEPA SPECIATE 4.3)
Resuspended crustal dust	Crustal Dust and Resuspended Dust (USEPA SPECIATE 4.3) + Natural Soil Dust (Pervez et al. 2018; Patil et al. 2013) + Paved and Unpaved road dust (Pervez et al. 2018; Patil et al. 2013)
Pyrolysis carbon-rich aerosol	Residential solid fuel stoves (Matawle et al. 2014)

The profiles used for the final runs of the CMB are given in Table 3.2. For sea salt, the PMF factor profile was used as is because 1) the relative contribution

of this source to total PM mass was less, and 2) no representative profile was available for the Indian sub-continent.

3.1.4 Chemical Transport Model, CTM Source Apportionment

There are two widely used approaches for SA using CTMs: a) brute force (BF) and b) tagged species (TS). BF is relatively popular (Pillon et al. 2011; Lin and Wen 2015; Wang et al. 2015; Belis et al. 2020). To estimate each source's contribution, the model output of a run with reduced emission of the investigated source is subtracted from a run with all the sources (base case). MERRA-2 generated PM_{2.5} being a reanalysis product, can be used for CTM based SA, but the source code, computing time and complexities to run the model were beyond the scope of this thesis; Therefore, an indirect approach was used to obtain sectoral contributions and source profiles akin to RMs (Figure. 3.2)

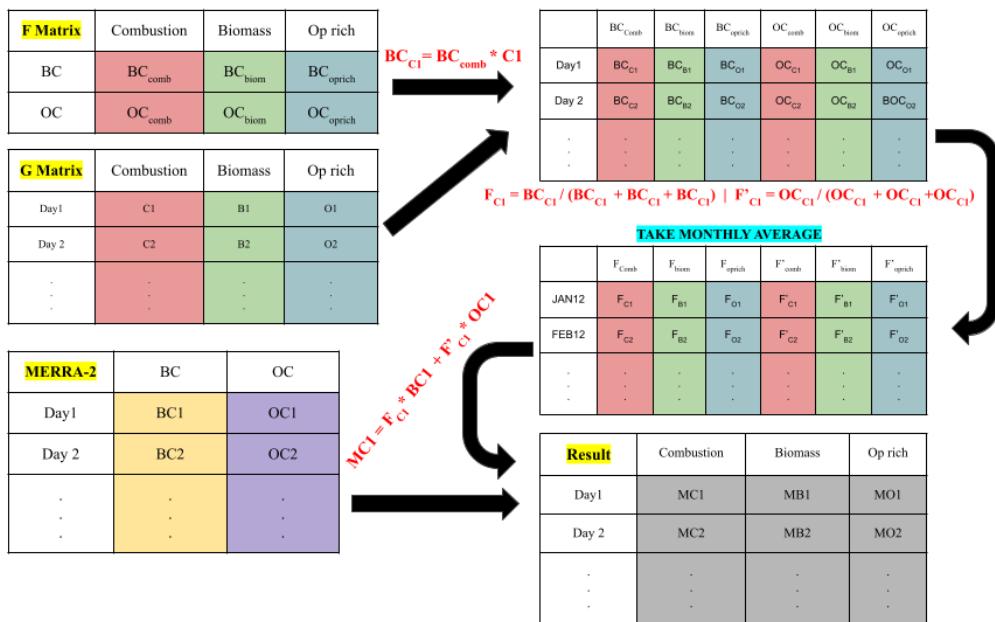


Figure 3.2. Method to assign MEERA-2 OC and EC to three PMF factors

MERRA-2 provides the PM_{2.5} mass concentration of OC, BC, sea salt, dust and sulphate. Analogues of RM like secondary sulphate, sea salt, and re-suspended crustal dust can be directly obtained, but there are no specific sectors to obtain the mass for combustion aerosol, biomass burning aerosol,

Pyrolysis carbon-rich aerosol or secondary nitrate equivalents of the RM outputs. Thus, to calculate CTM secondary nitrate, the daily nitrate ratio by sulphate from measured chemical concentrations was multiplied by MERRA-2 secondary sulphate (Ma et al. 2021). Since the major fraction of carbon was apportioned to combustion aerosol, biomass burning aerosol and pyrolysis carbon-rich aerosol factors by RMs, we have apportioned the total OC and BC from MERRA-2 into these three source categories. OC and BC fractions from source profiles (f matrix of PMF output) of each source factor were multiplied by the daily source contributions (g matrix of PMF output) of these three factors to obtain the daily OC and BC fractions contributed by each of these categories. The daily fraction of OC and BC contributed by each factor is determined and a monthly average is taken to get the monthly fractional contribution of OC and BC from each source. These fractions are then used to apportion daily OC and BC from MERRA-2 into three source categories and the OC and BC concentration are added to get the final factor contributions (Figure 3.2).

3.2 Study Design

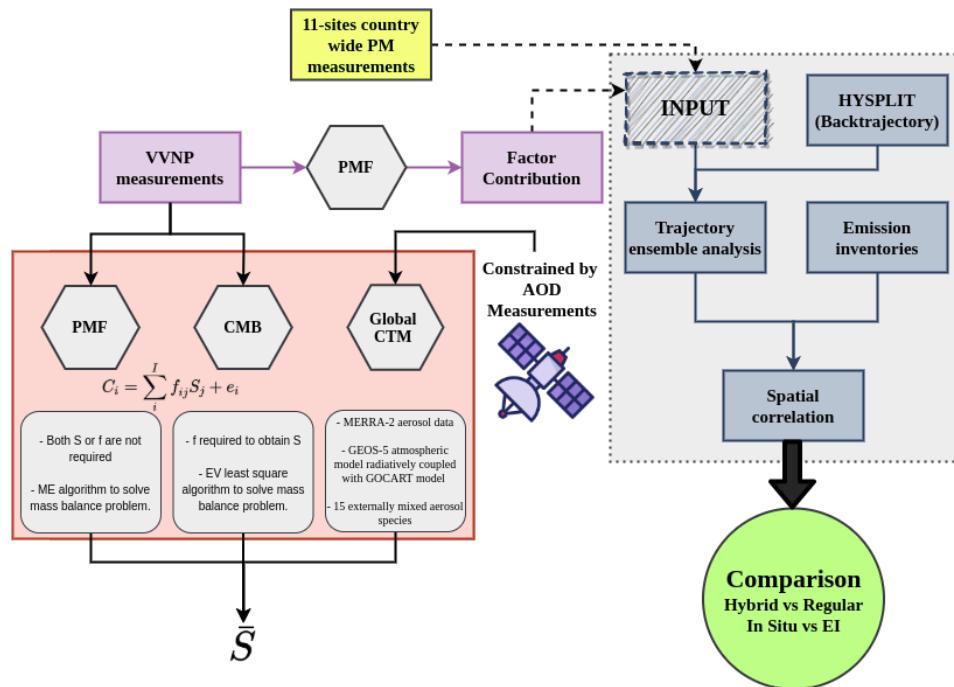


Figure 3.3. Flowchart illustrating study design

To perform trajectory ensemble analysis of 5-day back-trajectories (BT), starting at the Latitude and Longitude of receptor sites. A new BT is initiated every two hours utilizing the 0.5-degree GDAS meteorological dataset, with an arrival height of 500m AGL (e.g., Sunder Raman and Ramachandran 2011; Jaiprakash et al. 2017) resulting in ~1400 BT for VVNP and ~2200 for each site of the Network. Ultimately an R package - "Openair-Tools for the Analysis of Air Pollution Data" (Carslaw and Ropkins 2012) was used to generate PSCF and CWT maps. These Maps are then spatially correlated with EI to determine the major sources contributing to given site details discussed in appropriate sections of results.

4. Results and Discussion

In the previous chapters, the motivation and development of a framework and methods, tools/techniques, to achieve the objectives of this thesis were described. This chapter presents the important results and key findings of the study. The first part of this chapter is devoted to intercomparisons between trajectory ensemble methods derived from aerosol source locations and their reconciliation with EI, followed by a discussion on the Hybrid model developed for SA as applied to VVNP, Bhopal PM_{2.5} mass and chemical species. The next part of this chapter highlights the development and results of applying a methodology to identify and hybridize the trajectory analysis based potential source regions with EI to derive sectoral shares of primary PM_{2.5} contributions across India, at each of the pan-India multi-site COALESCE network locations.

4.1 Van Vihar National Park (VVNP) Bhopal

4.1.1 Comparison Of Trajectory Ensemble Derived Source Locations for VVNP Aerosol

To statistically compare the PSCF and CWT source location probability fields, we have used the Mean Normalised Bias (MNB) metric expressed as.

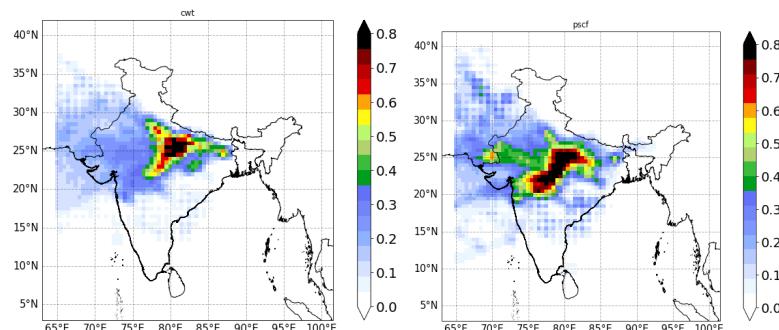
$$MNB(\%) = \frac{1}{N} \sum_{i=1}^N \left(\frac{PSCF_i - CWT_j}{CWT_i} \right) \times 100 \quad (4.1)$$

The MNB values obtained for each PMF apportioned VVNP PM_{2.5} factor source location maps by PSCF and CWT are given in Table 4.1. Positive bias indicates PSCF values are greater than CWT values; the larger the value, the higher the bias. For secondary nitrate (Figure 4.1.a), a positive MNB value results from the hotspot detected by PSCF in the region around MP and Maharashtra. Whereas for pyrolysis carbon-rich aerosol factor (Figure 4.1.b), both spatial domain and values obtained by both techniques are similar, resulting in a low bias. For secondary sulphate (Figure 4.1.c), the large green region in CWT (ability to

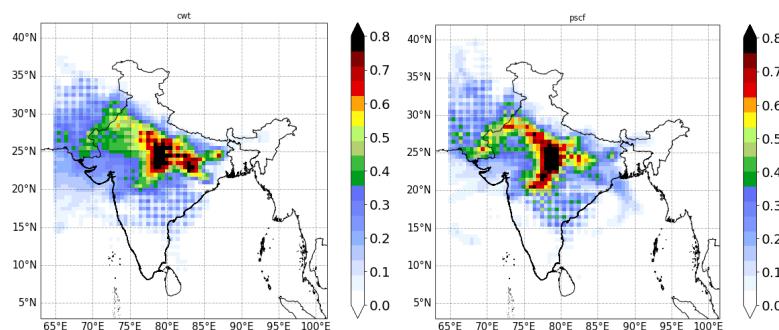
distinguish moderate and high source regions) governs the trend, resulting in negative values of MNB.

Table 4.1. MNB values PSCF vs CWT geographical source locations of VVNP PM_{2.5} mass

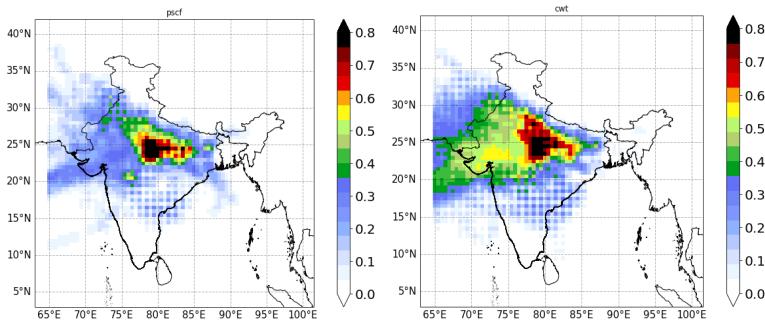
Factors	MNB (PSCF-CWT)
Biomass burning aerosol	-27.70
Combustion aerosol	-26.59
Pyrolysis carbon-rich aerosol	-1.57
Secondary Sulphate	-30.30
Secondary Nitrate	35.06



a) Secondary Nitrate



b) Pyrolysis carbon-rich aerosol



c) Secondary Sulphate

Figure 4.1. PSCF (left) and CWT (right) maps of source locations for various factors of VVNP PM_{2.5} mass to visualise MNB results.

The source regions identified and associated probability value assigned to a given grid by these two techniques are different, resulting in the range of magnitudes and sign of MNB. Both of the approaches have their strengths and limitations, as discussed in section 3. In the absence of information to assess which technique performs better in identifying ‘true’ source regions, an ensemble of the two to identify potential source regions was prepared. This ensemble was then compared with EI. The results of such comparisons are discussed in the following sections.

4.1.2 Reconciling Trajectory Ensemble Source Locations With Emissions Inventory

Spatial probability fields obtained from the trajectory ensemble receptor models (PSCF, CWT and their ensemble average) are spatially smooth while the EI data is sporadic, with few grid points having very high values (major sources of PM_{2.5}) adjacent to grid points with no or low values. Therefore, to obtain source location densities that will render a comparison of EI with outputs from the trajectory ensemble receptor models meaningful, the EI data were spatially smoothed using a 7 x 7 grid mean filter to obtain emission values at ~ 1° radius.

To capture the potential sources in the probability field regions obtained from the trajectory ensemble methods, an overlap map was generated using the

emission data grid points for which the PSCF/CWT values were greater than 0.4 for PSCF or 0.6 CWT, i.e. the top 80th percentile by each of these techniques, suggestive of the high likelihood of actual sources existing in these grid cells. Also, a dynamic colour scale, based on the maximum value of emission in the overlap region was chosen for better visualisation and interpretation of the overlap maps (Figure 4.2 and 4.3)

Appropriate sectoral outputs from EI were selected for comparisons, based on the source apportionment results over VVNP (Kumar and Sunder Raman, 2020). Although the analysis was performed for all sources, the overlap maps for select PM_{2.5} sources (combustion aerosol and secondary sulphate) are discussed in this section, as illustrative examples.

Combustion aerosol: Trajectory ensembles vs EI

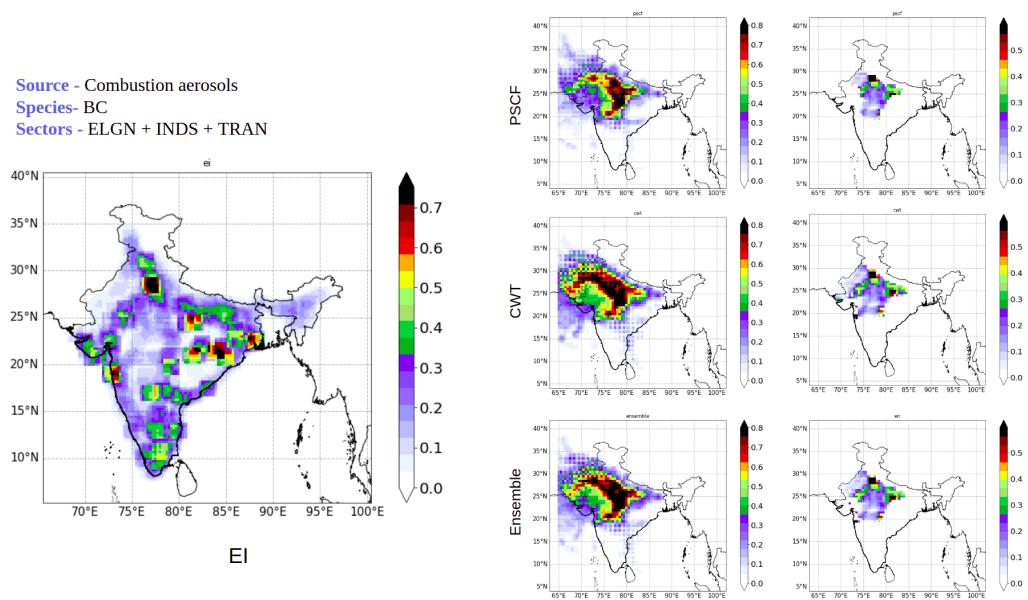


Figure 4.2. Comparisons between trajectory ensemble and EI locations of Black Carbon (BC) from combustion aerosol sources. The large map (leftmost) is the combined EI map of BC emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

The trajectory ensemble results of the combustion aerosol factor are compared with Black Carbon (BC) emissions from the combination of ELGN, INDS and TRAN sectors. BC is an exclusive combustion derived aerosol species and thus serves as the best proxy to assess the ability of trajectory ensemble receptor methods to identify sources in the EI. The results showed that potential source regions identified by PSCF/CWT coincide with known emission regions within the domain (Figure 4.2, see overlap maps).

While interpreting these results, one must be clear that PSCF and/or other trajectory techniques are not identical to emissions inventories. It is relevant to note that while we expect these methods to identify many of the sources on EIs, a low PSCF region does not necessarily indicate low emissions from the area. For instance, such a region might have high emissions from actual sources, but the air parcels travelling to the receptor site, may do not pass over those grids from that source to the measurement site, and/or that there is frequent precipitation may occur between the source and the receptor along the trajectory pathway, such that two locations that reduce $\text{PM}_{2.5}$ concentrations reduce faster than along such a pathway than other pathways to the receptor site.

Secondary sulphate: Trajectory ensembles vs EI

Primary pollutants are atmospheric particles emitted or injected directly into the atmosphere, whereas the chemical transformation of gaseous precursors and subsequent atmospheric transformations, including nucleation and condensational growth forms secondary aerosols. The SO_2 emissions from the ELGN and INDS sector are primary in nature, but the EI of this species is used as a proxy for a secondary pollutant (secondary sulphate) because SO_2 is the precursor gas of secondary sulphate aerosol and its chemical transformation time is of the order of hours to a couple of days (Rotstayn and Lohmann 2002), which is much faster than the transportation time (\sim five days) used in the air parcel back trajectory calculations. The results showed that both the PSCF and CWT methods successfully captured the primary emission sources (Figure 4.3, see overlap maps).

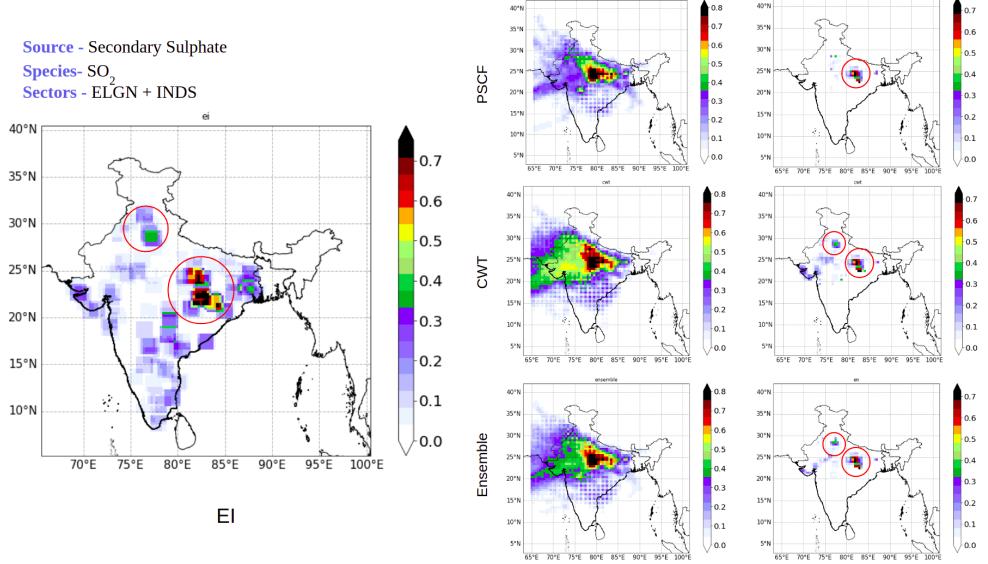


Figure 4.3. Comparisons between trajectory ensemble and EI locations of SO_2 from secondary sulphate sources. The large map (leftmost) is the combined EI map of SO_2 emissions from ELGN and INDS sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlaps with EI (rightmost).

Moreover, CWT distinguished between moderate and high concentration source regions, which allowed it to successfully capture a moderate source in the northern part of India (circled in red on the overlap map of the second row in Figure 4.3), which PSCF failed to identify. The ensemble (of PSCF and CWT) could capture both the moderate and high concentration source regions. Therefore, our results suggest that it is more meaningful to compare the ensemble of the PSCF and CWT maps with the EI rather than just any one of these techniques, especially for the secondary factors.

After comparing the VVNP factors with appropriate sectoral outputs from EI (as discussed above), we statistically compared the two maps using SCI (discussed in section - 2.5) to determine the effectiveness of trajectory ensemble tools in identifying potential source regions. The statistical hypothesis test result shows that “r - statistic” values greater than >1.6 (Appendix I Table A1.1) for all the performed comparisons, indicating a high spatial correlation. The ensemble of

PSCF and CWT has performed better than either of the two techniques in most of the cases (highest r-value - highlighted in Appendix I Table A1.1).

4.1.3 Hybrid RM-CTM source apportionment

PMF5 and MERRA-2 results were obtained as described in section 3.1.1 and 3.1.4 respectively, while CMB 8.2 was set up using profiles discussed in section 3.1.3 and optimum solution was achieved by assessing the ranges of various model test parameters given Appendix II Table A2.1. This section focuses on highlighting the procedure used to develop the hybrid model. Due to the unavailability of sectoral PM_{2.5} estimates and uncertainties associated with EI from the MERRA-2 products over VVNP, the hybridisation used in this thesis is not a “true” hybridisation of CTM and RM outputs. However, the method for hybridization that is presented in the following section still holds, and the best possible estimates (given the limitations associated with sectoral PM_{2.5} CTM results) of modified weighted ensemble concentrations were obtained using the equation below

$$\bar{S}_j = \frac{S_{j_{CTM}}}{3} + \frac{2}{3} \frac{w_{j_{PMF}} \cdot S_{j_{PMF}} + w_{j_{CMB}} \cdot S_{j_{CMB}}}{w_{j_{PMF}} + w_{j_{CMB}}}$$

In CMB, source contribution uncertainties are given as an output of the model and estimated using an effective variance approach (Watson et al. 1984):

$$\sigma_{s_j} = \left(\sum_{i=1}^n \frac{f_{ij}^2}{\sigma_{ci}^2 + \sum_{j=1}^m \sigma_{fij}^2 S_j^2} \right)^{-1/2}$$

where σ_{s_j} (in $\mu\text{g}/\text{m}^3$) is the uncertainty in source contribution S_j (in $\mu\text{g}/\text{m}^3$), σ_{ci} is the uncertainty in the measurement of ambient species i , and σ_{fij} is the uncertainty in the fraction of i th species, in the j th source profile. PMF does not provide uncertainties as standard output; therefore, we use the above equation to propagate factor profile uncertainties determined using bootstrapping (Lee et al. 2008; Lee et al. 2009) to obtain error bounds on factor contributions.

The SA models used in this study have their advantages and shortcomings, as previously discussed in section 2.1. For example - RMs show high day-to-day

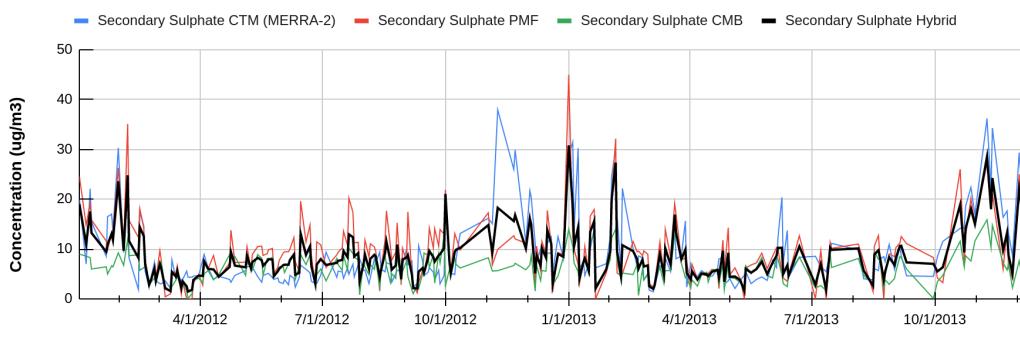
variability in source contribution to the total PM_{2.5} mass concentrations. These models often fail to capture the major sources of particulate pollution on particular days (Balachandran et al. 2012). On the other hand, CTMs show low day-to-day variability in source contribution over a location due to the coarse resolution of the meteorological models and monthly EI inputs. Daily variability is only driven by changing meteorology.

Zero impact days are characterised by the days when there is no PM contribution from a source to PM mass over a given site. The hybrid model developed in this thesis had the least number of zero impact days compared with PMF, CMB and the CTM used in this study. While PMF results showed a maximum number of zero impact days for almost all the sources/factors, as shown in Table 4.3., the CMB model was intermediate. Interestingly, the CTM (MERRA-2 product) output had zeroes only for a diffuse and ubiquitous source, i.e., pyrolysis carb-rich aerosol (Table 4.2), which was shown to be a proxy for residential fuel burning emissions.

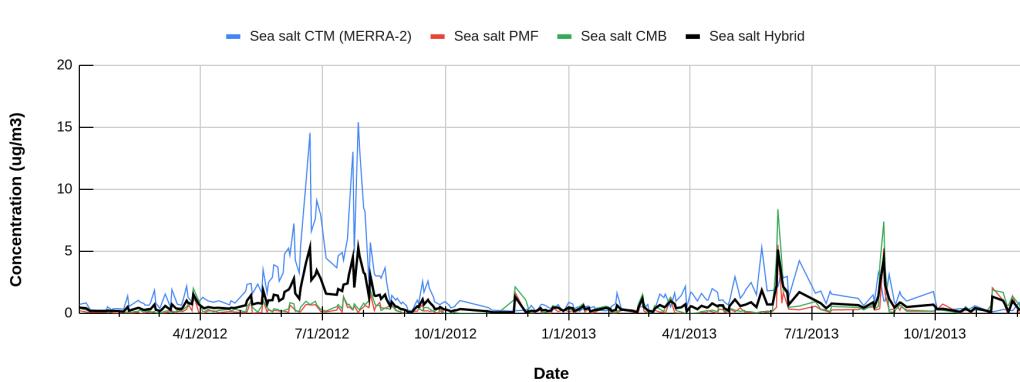
Table 4.2. Number of zero-impact days for each source calculated by three SA models and the hybrid model

Sources	MERRA-2	PMF	CMB	Hybrid
Combustion aerosol	0	24	1	0
Biomass burning aerosol	0	39	7	0
Pyrolysis carbon-rich aerosol	14	44	15	4
Sea salt	0	41	15	0
Secondary Sulphate	0	7	0	0
Secondary Nitrate	0	34	1	0
Re-suspended crustal dust	0	10	0	0

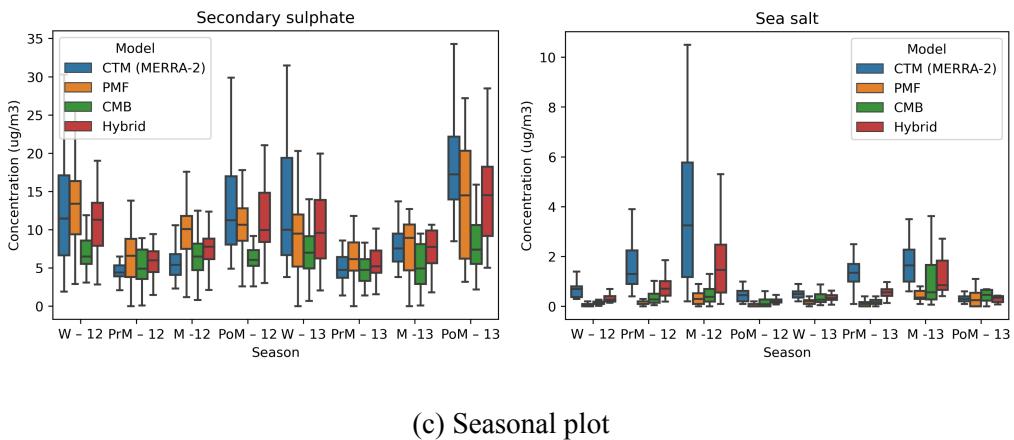
The results from this study demonstrated that the outputs from CMB, PMF and CTMs could suggest very different source impacts over the same receptor location, on a particular day. These differences are easily visible in both seasonal (box and whiskers plots, Figure 4.4.c) and daily (time-series plots, Figure 4.4.a,b) time scales, for select sources, other plots are present in Appendix III (Figure A2.1) For example, the time-series of the secondary sulphate factor showed a spike for the CTM, i.e, MERRA-2 ($37.9 \mu\text{g}/\text{m}^3$) output on 11/09/2012, which was ~5 times more than the concentrations obtained from the RMs ($9.9 \mu\text{g}/\text{m}^3$ - PMF and $5.7 \mu\text{g}/\text{m}^3$ - CMB) on the same day (Fig. 4.4. a). The SA results from individual models are widely used to comment on the relative impact of sources over different receptor sites. In the absence of any direct way to validate these individual model results, it is reasonable to assume that the hybridization of these models provides a more stable estimate than individual models, to determine the impact of a given source on the pollutants measured over a receptor site.



(a) Secondary Sulphate



(b) Sea salt



(c) Seasonal plot

Figure 4.4. Daily time-series plots (a,b) and seasonal box and whiskers plot (c) showing $\text{PM}_{2.5}$ source contributions (mass concentration) from each model.

Overall, it can be seen that the hybrid model developed in this study could resolve source impacts on days when individual models failed (Table 4.3.). The hybridization approach is also particularly valuable for the CMB model, which was unable to apportion mass due to collinearity between source profiles on over 14% of total sampling days (236 days). Further, the source profile derived from the hybrid model output can provide a locally relevant refined profile, for use in scenarios where only CMB is possible to help enhance the SA results obtained by this method.

4.2 Pan-India COALESCE Network (11 Sites)

In section 4.1, we demonstrated the application of the methodology developed in this thesis for aerosols over Van Vihar National Park, Bhopal, with multiple sources (7 factors from PMF). This section illustrates the application and effectiveness of the trajectory ensemble receptor model combinations with emissions inventory to identify the sectoral shares of primary $\text{PM}_{2.5}$ across India at each of the multi-site COALESCE network locations. In this analysis, the total fine PM mass is not apportioned to various sources. The in-situ surface concentration measurements are directly used in conjunction with air parcel back trajectories to obtain trajectory ensemble maps. We show the PSCF maps and

their overlaps with emissions inventory, to illustrate and discuss key results. The CWT maps and their overlaps with EI are placed in Appendix III (Figure A3.1).

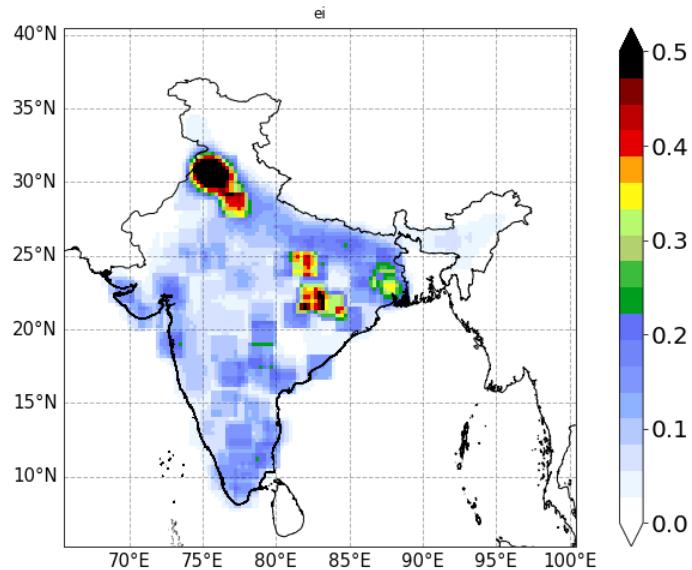
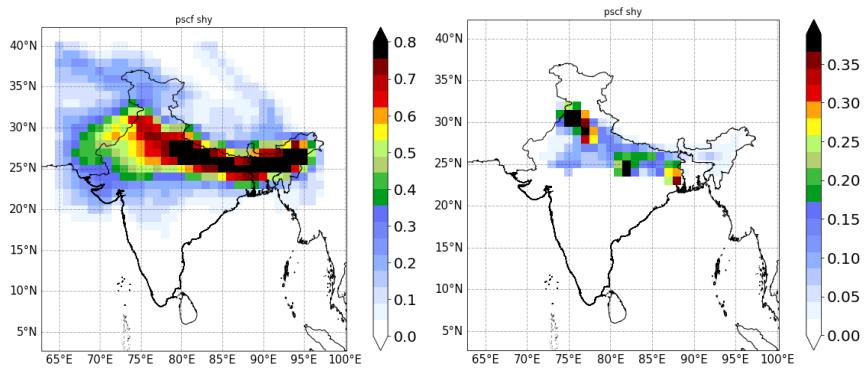
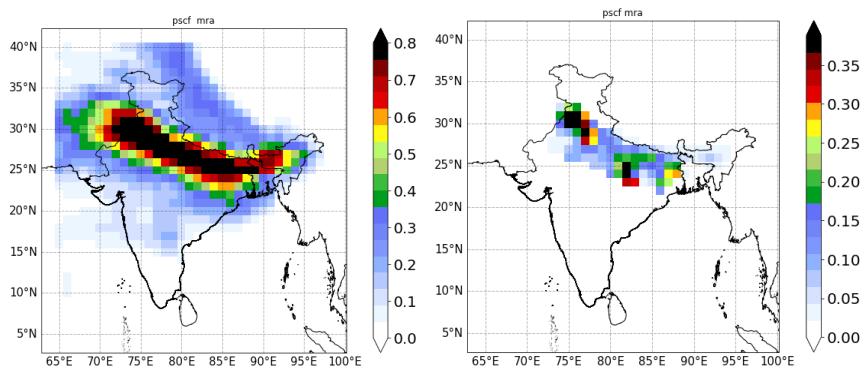


Figure 4.5. PM_{2.5} Emission Inventory map - 2019 (all sectors)

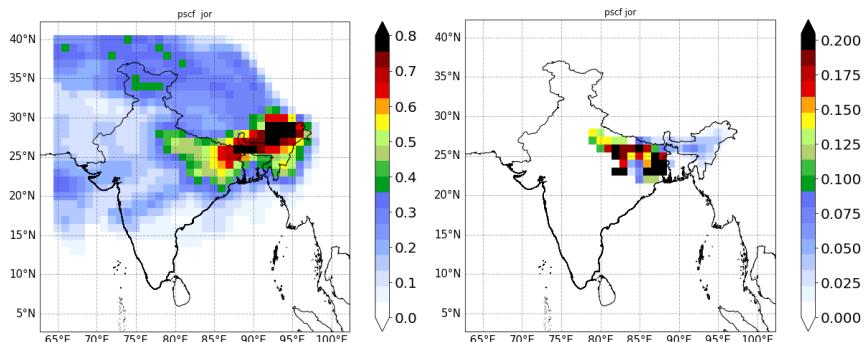
All the EI sectors were combined to obtain a total primary PM_{2.5} emissions map for all of India during 2019 (Figure 4.5). The network was divided into three regional clusters based on the geographical location and proximity of individual sites to one another. These clusters were: 1) Eastern - Shyamnagar, Mesra, Jorhat, 2) Northern - UoK, Rohtak, Bikaner, Mohali and 3) Southern and Central - Bhopal, Hyderabad, Mysuru, Mahabaleshwar, for ease of discussion. To visualise the major sources responsible for PM_{2.5} on an annual basis at any given site, the trajectory ensemble- EI overlap maps were obtained using the same criteria and methodology discussed for VVNP, Bhopal in section 4.2.



(a) Shyamnagar



(b) Mesra

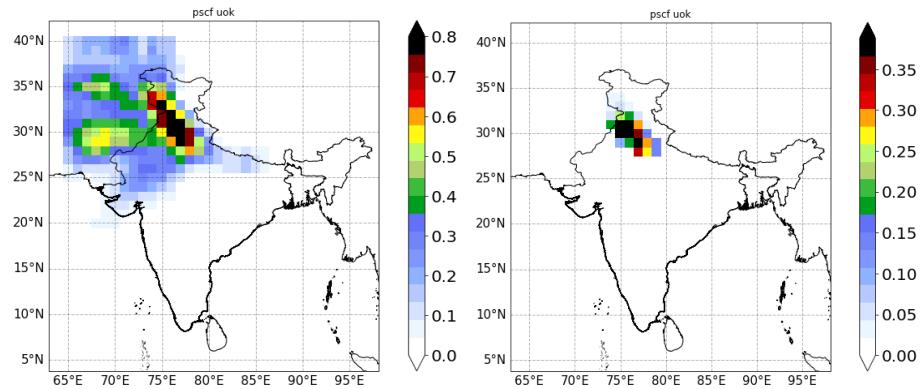


(c) Jorhat

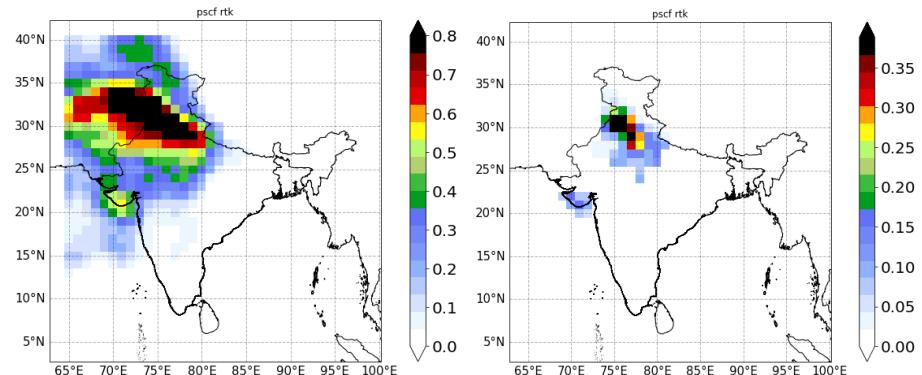
Figure 4.6. PSCF (left) and overlap with EI (right) for sites in the Eastern cluster

The results suggest that primary fine PM in the Indo Gangetic plain outflow influences fine PM mass concentrations measured over sites in the Eastern India cluster (Figure 4.6). Further, a comparison of the PSCF map generated for each site with the EI map revealed that PSCF successfully captured

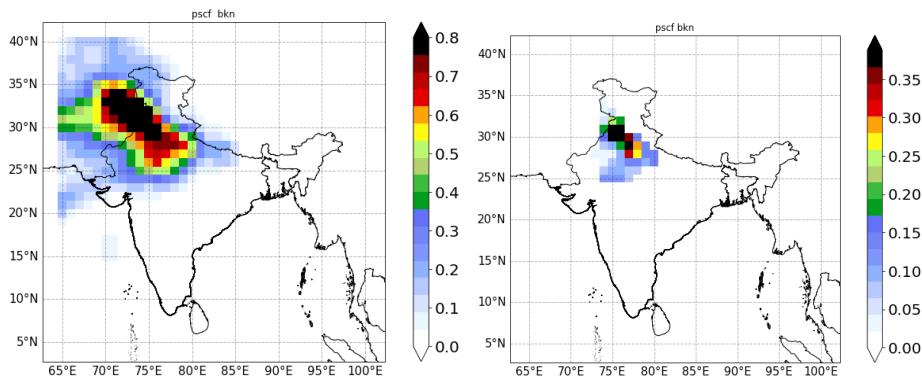
the location of major PM sources for all three locations. However, it must be borne in mind that PSCF analysis cannot distinguish between the preferred transport pathways and the actual source locations, especially if these locations lie outside the trajectory domains.



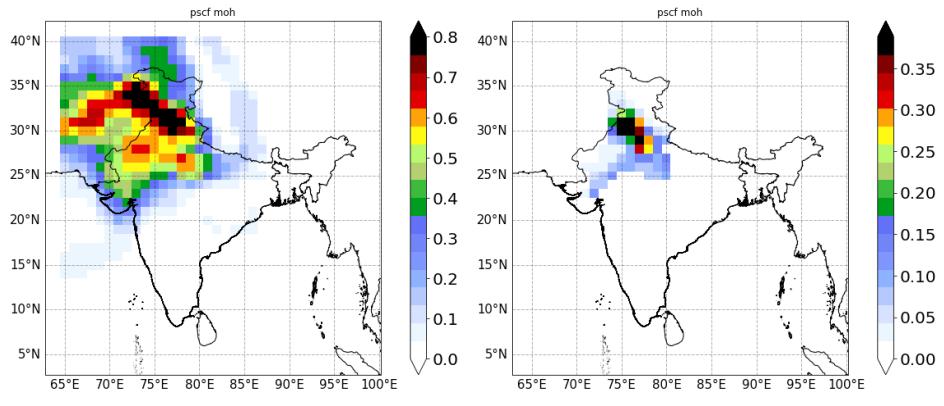
(a) UoK



(b) Rohtak



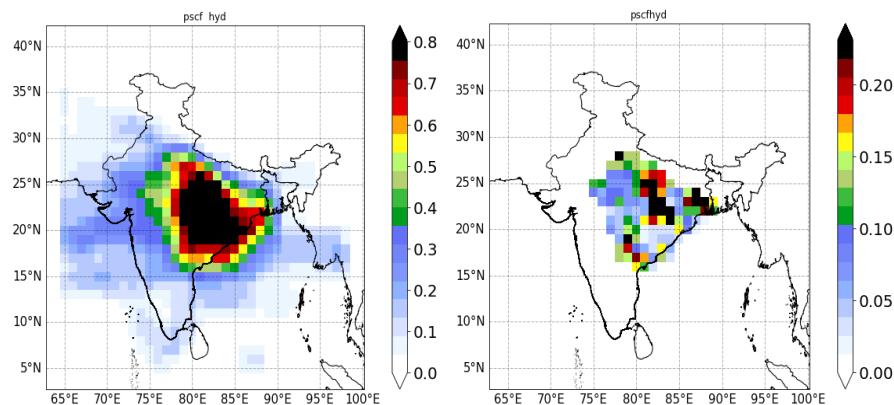
(c) Bikaner



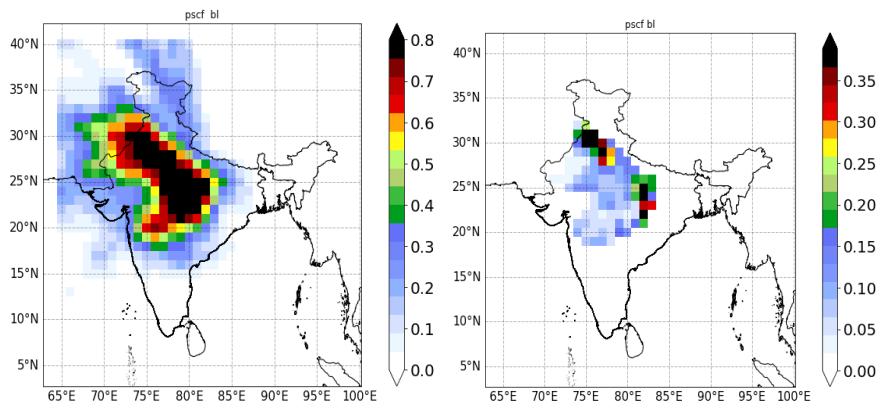
(d) Mohali

Figure 4.7. PSCF (left) and overlap with EI (right) for sites in the Northern cluster

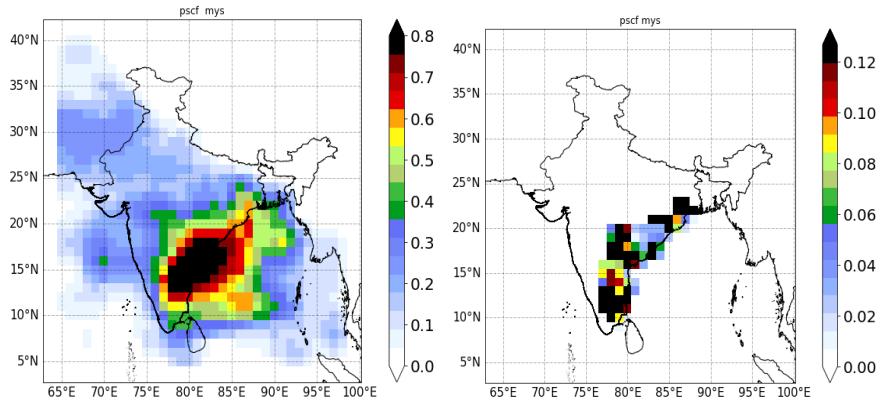
Primary PM_{2.5} emissions in the region around Punjab-Haryana influenced the concentration of this species for all the sites in the Northern India cluster (Figure 4.7) and two sites in the Eastern India cluster, i.e. Shyamnagar and Mesra (Figure 4.6). Also, the transboundary pollutants from sources in the outflow originating in Pakistan and/or Afghanistan may have been relevant mass contributors, but they were not in the scope of the national EI for India, and these potential overlaps with trajectory ensemble source locations could not be identified.



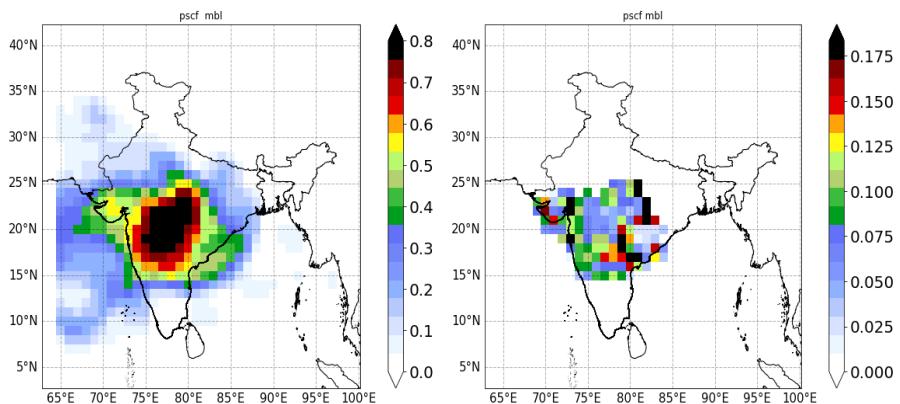
(a) Hyderabad



(b) Bhopal



(c) Mysuru



(d) Mohali

Figure 4.8. PSCF (left) and overlap with EI (right) for sites in the Southern and Central cluster

On a national EI scale (Figure 4.5), no major sources were found to influence the primary PM_{2.5} measured over locations in the Southern and Central India cluster, except for Bhopal, where Punjab-Haryana emissions hotspots influenced the fine PM. While the major sources on the national EI scale did not affect the concentration of PM much over southern India, the PM over these regions appeared to be driven by moderate to minor scale sources on the national emission scale, which were located in the vicinity of these sites (Figure 4.8).

Table 4.3. Results (r-statistic values) of Spatial Correlation hypothesis test for each site of the COALESCE network.

Site	PSCF	CWT
Bhopal	91.72	72.44
Rohtak	51.10	72.40
Mohali	58.13	67.02
Mesra	84.52	84.47
UoK	43.13	43.33
Mysuru	54.72	53.46
Shyamnagar	73.98	79.15
Bikaner	49.12	67.55
Hyderabad	97.13	89.74
Jorhat	40.67	63.14
Mahabaleshwar	80.77	84.21

After determining the potential source regions we statistically compared the two maps using SCI (discussed in section - 2.5) to determine the effectiveness of trajectory ensemble tools in identifying potential source regions. The statistical hypothesis test result shows that “ r - statistic” values greater than >> 1.6 (Table 4.3) for all the sites, indicating a high spatial correlation.

In the absence of previous studies over the study locations and/or evaluations with simulated datasets, there is no basis to comment on the superior performance of one trajectory ensemble method over another. The emissions over

grid cells in the overlap maps (CWT and PSCF ensemble average maps on an annual scale overlapped with the yearly averaged national PM_{2.5} EI map for 2019) were obtained to determine the sectoral contribution of primary PM_{2.5} over each site (Figure 4.9). The sectoral maps of PSCF and CWT are present in Appendix III (Figure A3.2).

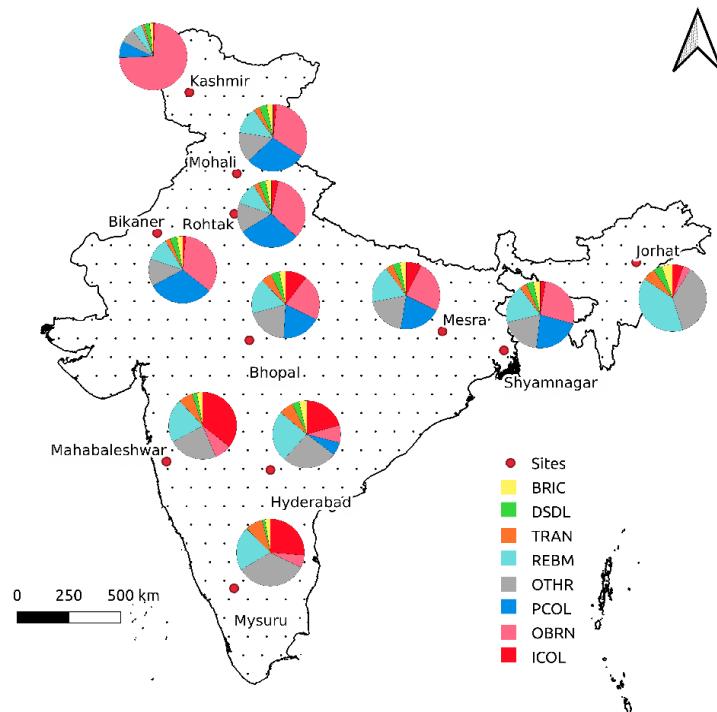


Figure 4.9. Sector-wise contribution to primary PM_{2.5} mass over 11 site COALESCE network across India for 2019 (Ensemble)

Open burning and thermal power plant coal combustion were the major contributing sectors over locations in northern India (Figure 4.9); whereas, industrial coal was a fairly large contributor to fine PM over southern India sites with negligible contribution from open burning and thermal power plants at these locations (Figure 4.9). Residential biomass emissions were rather ubiquitous and influenced all locations across the network to comparable extents (Figure 4.9). The insights provided by this apportionment, together with monthly and/or seasonal sectoral share maps will provide key inputs for immediate primary PM_{2.5} source mitigation actions, at the national scale.

5. Summary and Conclusions

The purpose of the study was to assess the potential of and demonstrate the applications of air parcel trajectory ensemble tools to identify potential aerosol source locations. Additionally, a hybrid CTM-RMs for PM_{2.5} source identification and apportionment was developed and applied to ambient fine PM measurements. This study used two ambient PM data sets: 1) VVNP PM_{2.5} mass and chemical species along with PMF results and 2) PM_{2.5} mass concentrations from the pan-India 11 sites national network - COALESCE.

In this thesis, the potential location of regional sources was identified by PSCF and CWT air parcel trajectory models and their ensemble. The ability of these tools to identify primary PM_{2.5} source locations was assessed by finding the association between PSCF/CWT/Ensemble identified source regions and the known locations of these sources from PM_{2.5} India Emissions Inventory (EI) - SMOG, by performing a spatial correlation hypothesis test. The test results showed a high correlation (r -values > 1.6) between EI source locations and the regions identified by PSCF, CWT and their ensemble for both VVNP aerosol and aerosol at all sites in the COALESCE network. In the case of VVNP, the ensemble of the two methods showed the best agreement (highest r -values) with the EI, which indicates the effectiveness of the ensemble in identifying primary PM_{2.5} source locations compared to either PSCF or CWT alone.

To determine potential regional sources affecting each of the 11 COALESCE network sites, these sites were divided into three major clusters and an approach similar to that for VVNP was used. After validating trajectory ensemble results with EI using the spatial correlation test, we determined the sectoral primary PM_{2.5} contribution over each site based on trajectory pathways for all COALESCE sites. Thus, the novelty of this study was its hybrid EI-trajectory ensemble reconciliation to obtain a quantitative estimate of primary PM_{2.5} sectoral shares across locations in India. These results can help guide PM_{2.5} source reductions and mitigation actions on a national scale.

In the future, other trajectory ensemble tools like SQTBA, RTWC can also be included in the approach developed in this thesis, the ensemble of which will be more robust in helping obtain location-specific primary PM_{2.5} sectoral contributions. Based on the trends of EI sectoral shares and similarity between the source locations of different sites, an attempt can be made to identify the reason for these similarities. Further, a multi-site trajectory ensemble analysis can be done to identify the common source locations.

To demonstrate the RM-CTM hybridisation methodology and assess its potential, in SA, we developed a hybrid model using two RMs (PMF5.0 and CMB8.2), and one CTM product (MERRA-2) applied it to VVNP PM_{2.5} and its chemical species. The hybrid model overcame some limitations of the individual models. The hybrid model results showed a significant decrease in (unrealistic) zero source contribution days obtained from RMs.. This approach was most useful to apportion PM mass when output was not available from one or more individual models on multiple days. The hybridisation framework developed in this thesis is very flexible and can include any combination of RMs and CTMs based on resource availability. The study is also the first of its kind to hybridise a satellite AOD constrained CTM product (MERRA-2) with two other receptors models with reasonable time-resolution (every other day measurements) for a relatively long span of time (236 days) over locations in India. Typically, such studies over India or elsewhere are performed on just seasonal or monthly scales.

Further, other receptor models like UNMIX, modified versions of CMB that include molecular markers and ensemble CTM outputs can also be incorporated into the hybridisation framework (Figure 5.1).

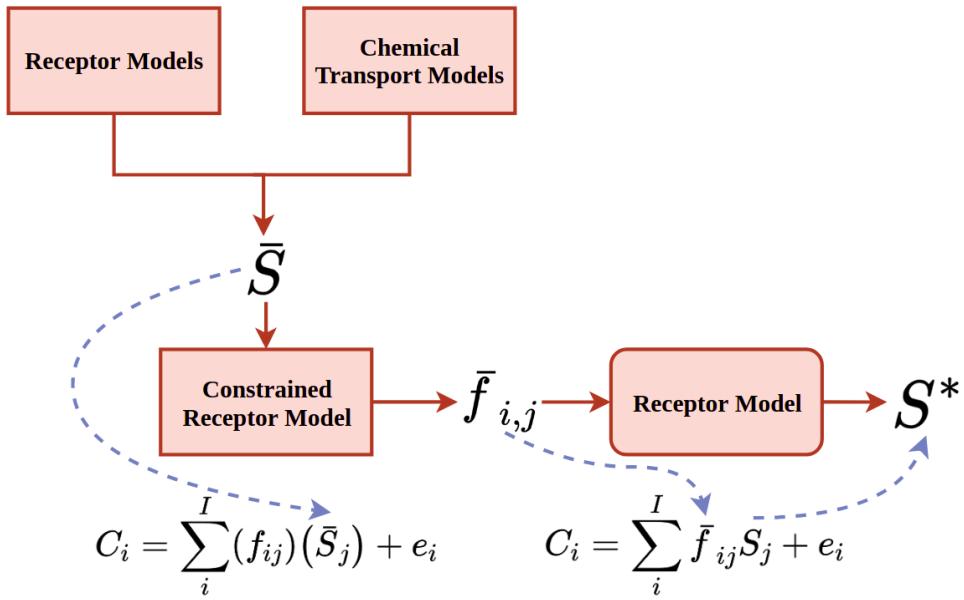


Figure 5.1. Schematic of hybrid CTM-RM constrained receptor model framework

An immediate benefit of such models is the fact that more accurate and representative source profiles can be obtained from the hybrid model. The methodology used for hybridization in this study can be further extended to develop ensemble trained receptor models by determining more robust location-specific source profiles for source apportionment (Figure 5.1), to strengthen policy actions and mitigation measures based on the results of such studies.

Appendices

Appendix I - Trajectory Ensemble Results - VVNP

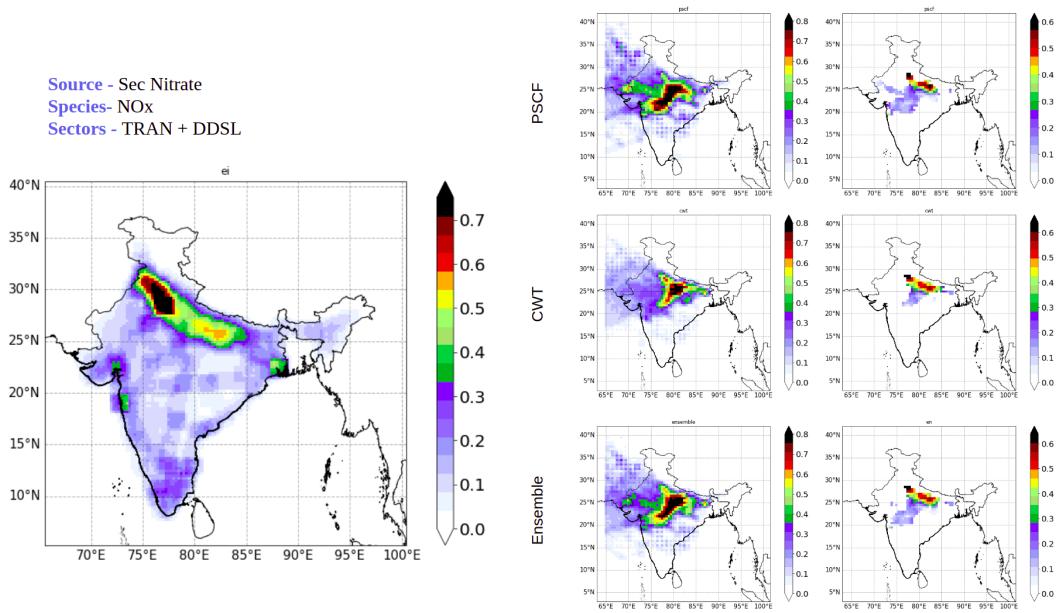


Figure A1.1. Comparisons between trajectory ensemble and EI locations of NOx from secondary nitrate sources. The large map (leftmost) is the combined EI map of NOx emissions from DDSL and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

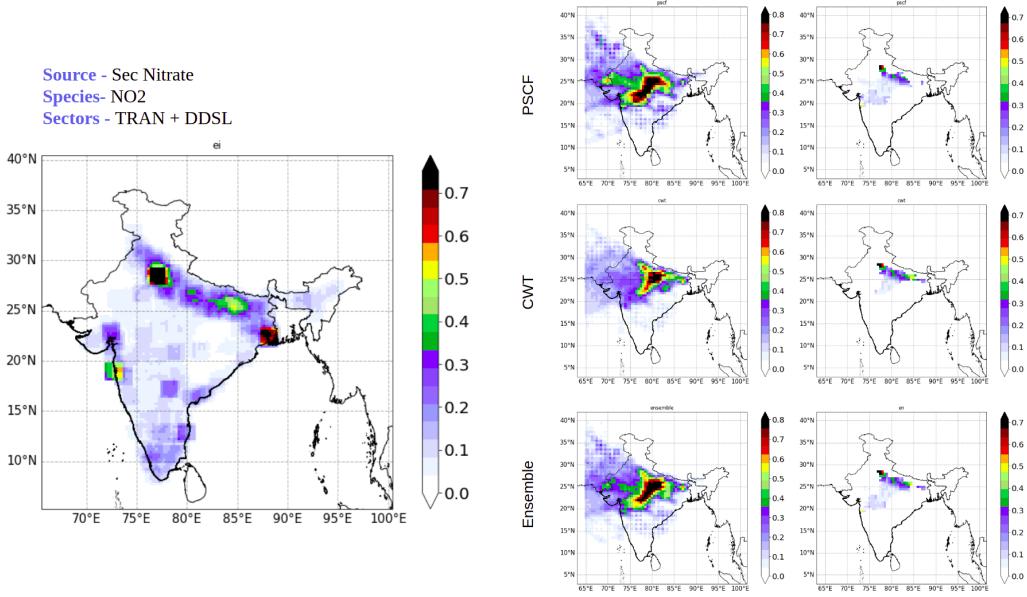


Figure A1.2. Comparisons between trajectory ensemble and EI locations of NO₂ from secondary nitrate sources. The large map (leftmost) is the combined EI map of NO₂ emissions from DDSL and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

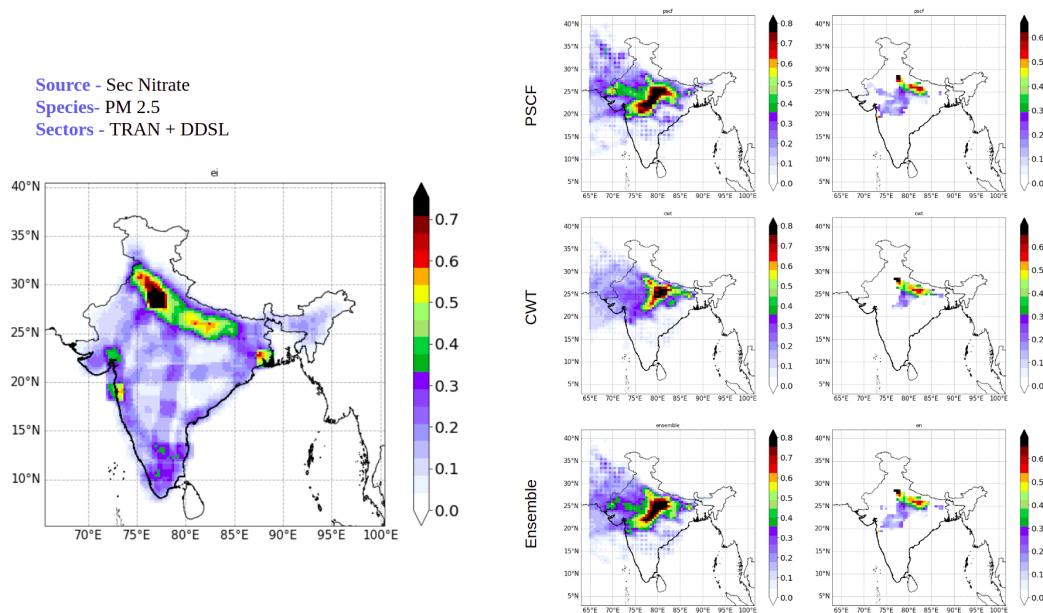


Figure A1.3. Comparisons between trajectory ensemble and EI locations of PM_{2.5} from secondary nitrate sources. The large map (leftmost) is the combined EI map of PM_{2.5} emissions from DDSL and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

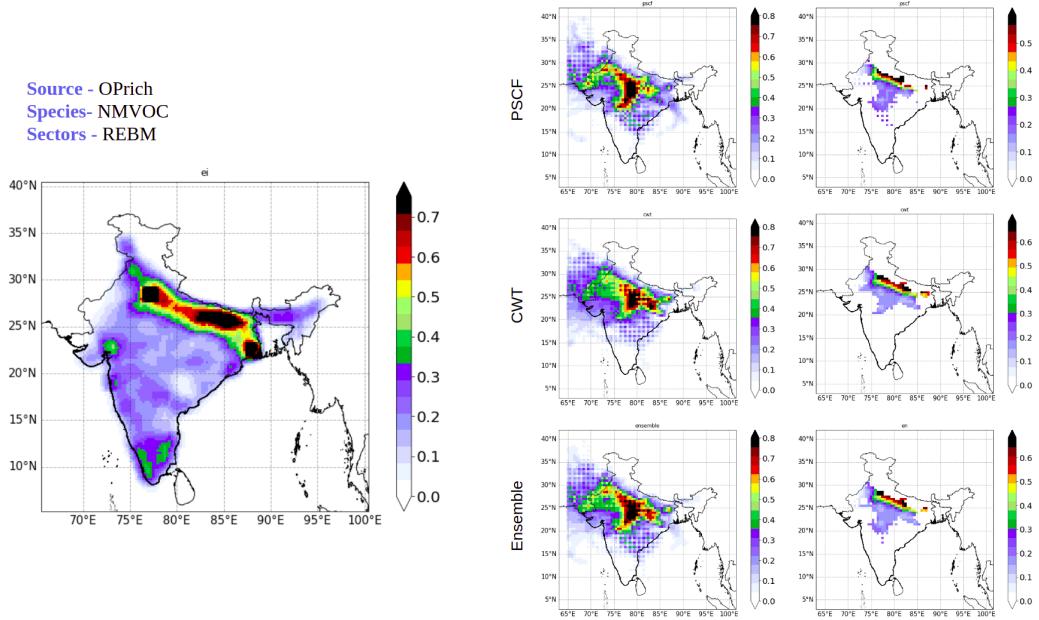


Figure A1.4. Comparisons between trajectory ensemble and EI locations of NMVOC from pyrolysed carbon-rich aerosol sources. The large map (leftmost) is the combined EI map of NMVOC emissions from REBM sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and their respective overlap maps with the EI (rightmost).

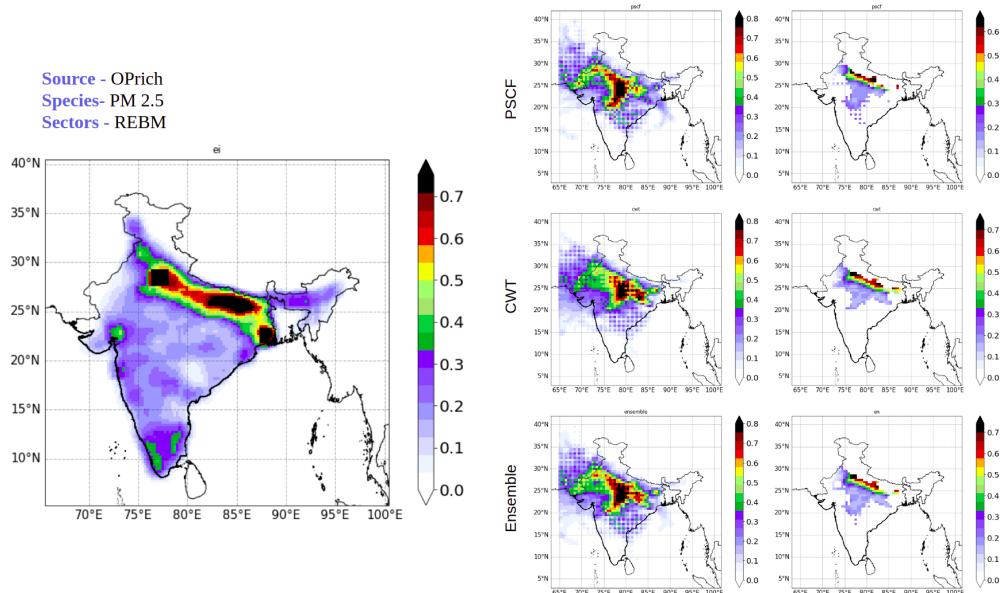


Figure A1.5. Comparisons between trajectory ensemble and EI locations of $\text{PM}_{2.5}$ from pyrolysed carbon-rich aerosol sources. The large map (leftmost) is the combined EI map of $\text{PM}_{2.5}$ emissions from REBM sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

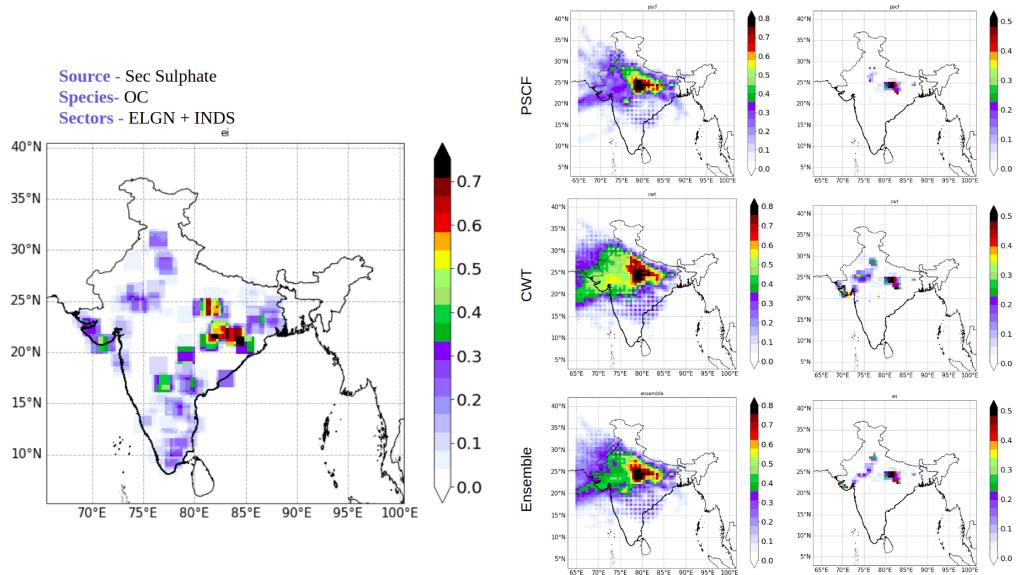


Figure A1.6. Comparisons between trajectory ensemble and EI locations of Organic Carbon (OC) from secondary sulphate sources. The large map (leftmost) is the combined EI map of OC emissions from ELGN and INDS sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

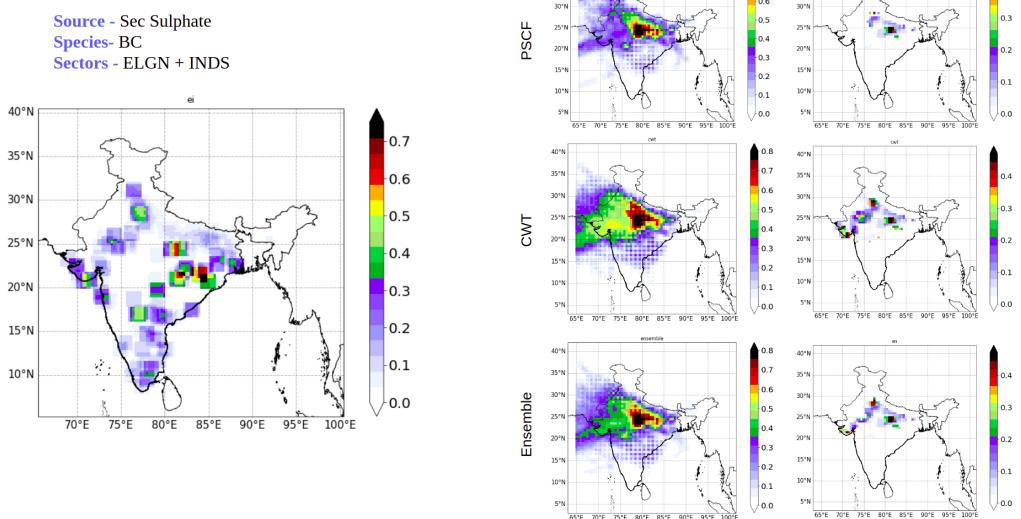


Figure A1.7. Comparisons between trajectory ensemble and EI locations of Black Carbon (BC) from secondary sulphate sources. The large map (leftmost) is the combined EI map of BC emissions from ELGN and INDS sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

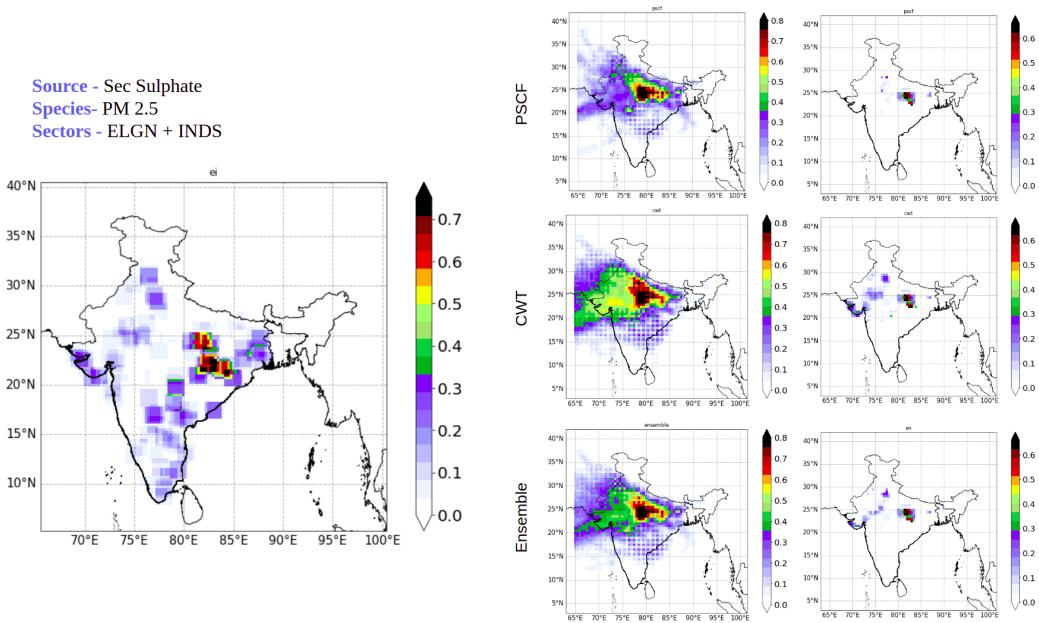


Figure A1.8. Comparisons between trajectory ensemble and EI locations of $\text{PM}_{2.5}$ from secondary sulphate sources. The large map (leftmost) is the combined EI map of $\text{PM}_{2.5}$ emissions from ELGN and INDS sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

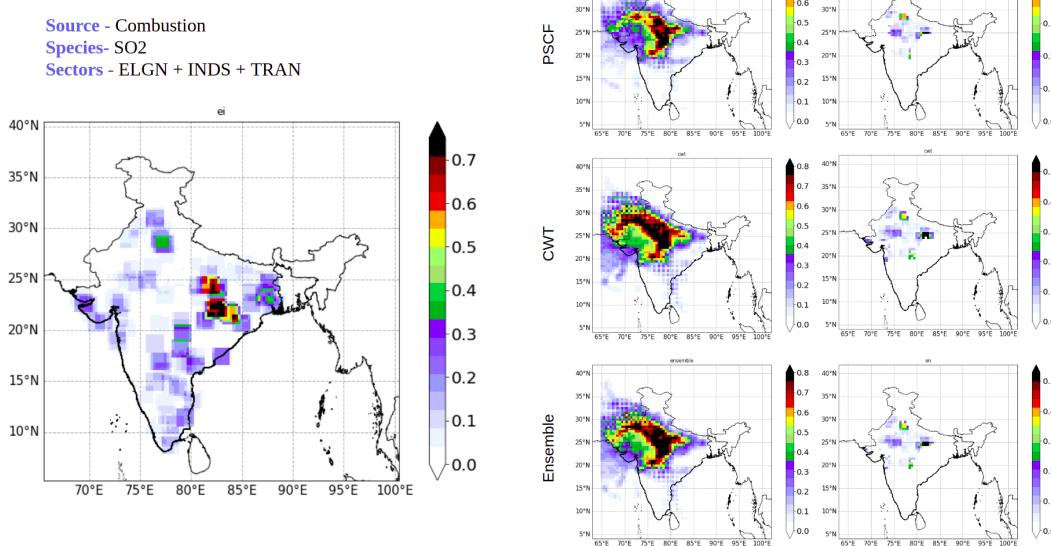


Figure A1.9. Comparisons between trajectory ensemble and EI locations of SO_2 from combustion aerosol sources. The large map (leftmost) is the combined EI map of SO_2 emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

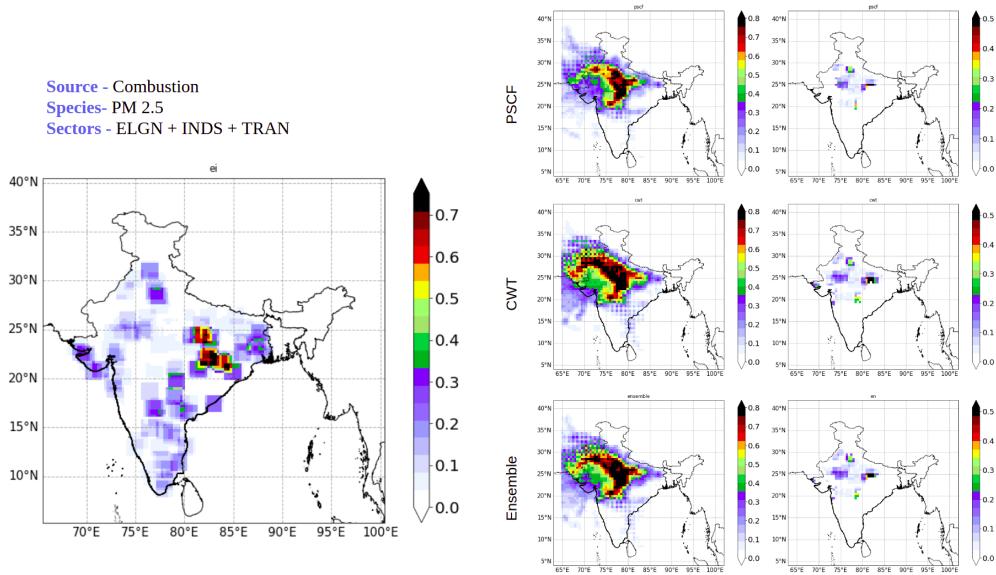


Figure A1.10. Comparisons between trajectory ensemble and EI locations of PM_{2.5} from combustion aerosol sources. The large map (leftmost) is the combined EI map of PM_{2.5} emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps and on the right are their respective overlap maps with the EI (rightmost).

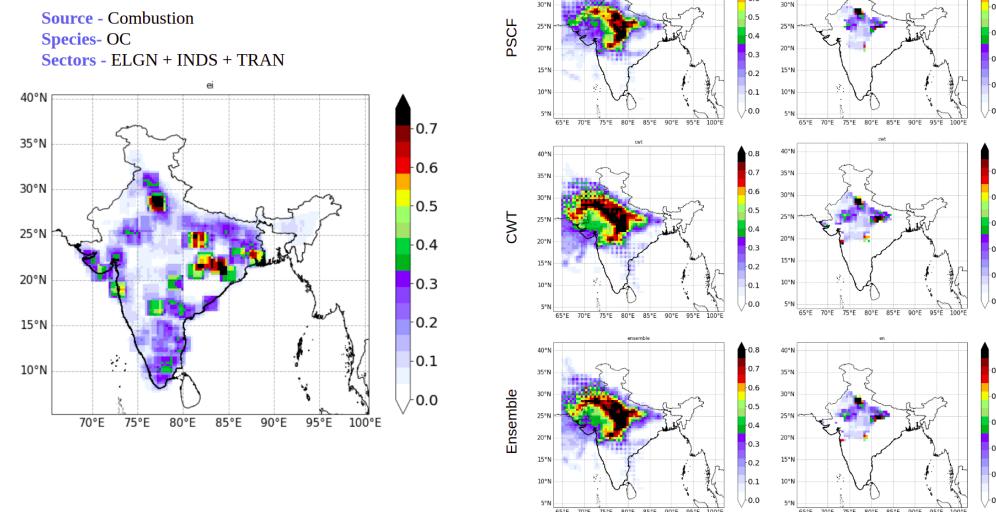


Figure A1.11. Comparisons between trajectory ensemble and EI locations of Organic Carbon (OC) from combustion aerosol sources. The large map (leftmost) is the combined EI map of OC emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

Source - Combustion
Species- BC
Sectors - ELGN + INDS + TRAN

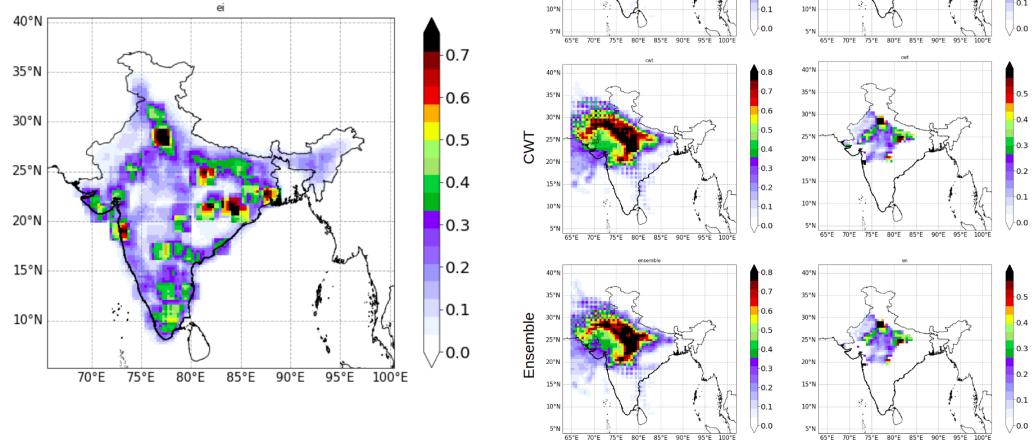


Figure A1.12. Comparisons between trajectory ensemble and EI locations of Black Carbon (BC) from combustion aerosol sources. The large map (leftmost) is the combined EI map of BC emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

Source - Biomass
Species- BC
Sectors - REBM + AGBR

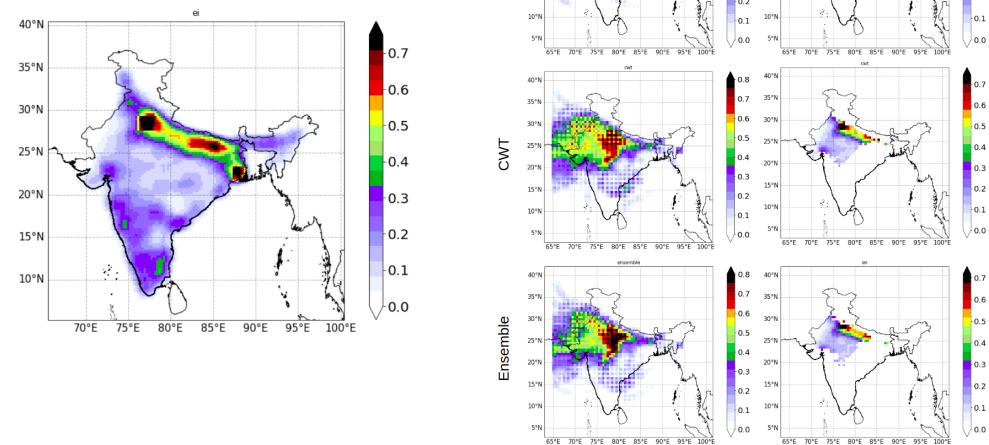


Figure A1.13. Comparisons between trajectory ensemble and EI locations of Black Carbon (BC) from biomass burning aerosol sources. The large map (leftmost) is the combined EI map of BC emissions from REBM and AGBR sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

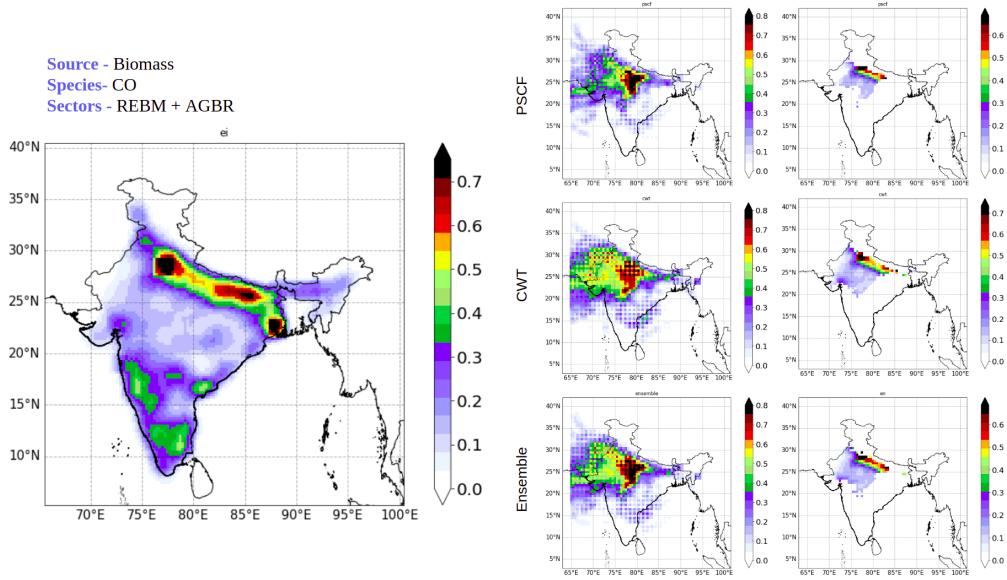


Figure A1.14. Comparisons between trajectory ensemble and EI locations of CO from biomass burning aerosol sources. The large map (leftmost) is the combined EI map of CO emissions from REBM and AGBR sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

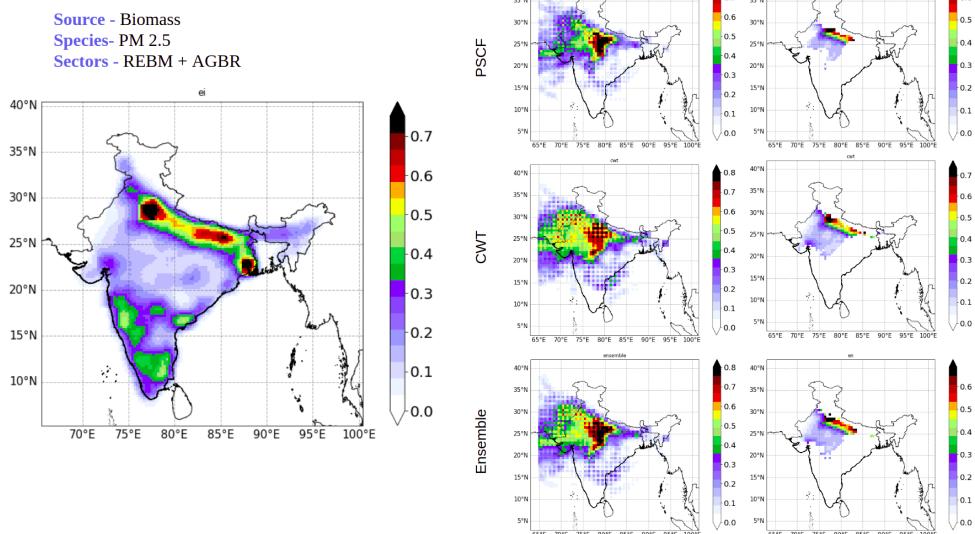


Figure A1.15. Comparisons between trajectory ensemble and EI locations of PM_{2.5} from biomass burning aerosol sources. The large map (leftmost) is the combined EI map of PM_{2.5} emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

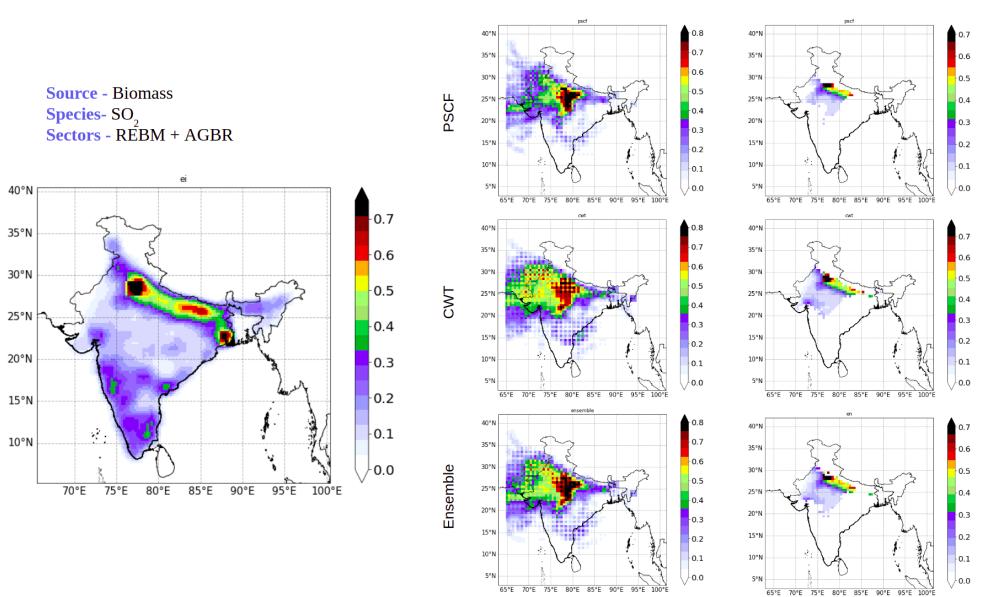


Figure A1.16. Comparisons between trajectory ensemble and EI locations of SO₂ from biomass burning aerosol sources. The large map (leftmost) is the combined EI map of SO₂ emissions from ELGN, INDS and TRAN sectors, maps on the left of the 2x3 map grid are trajectory ensemble maps, and on the right are their respective overlap maps with the EI (rightmost).

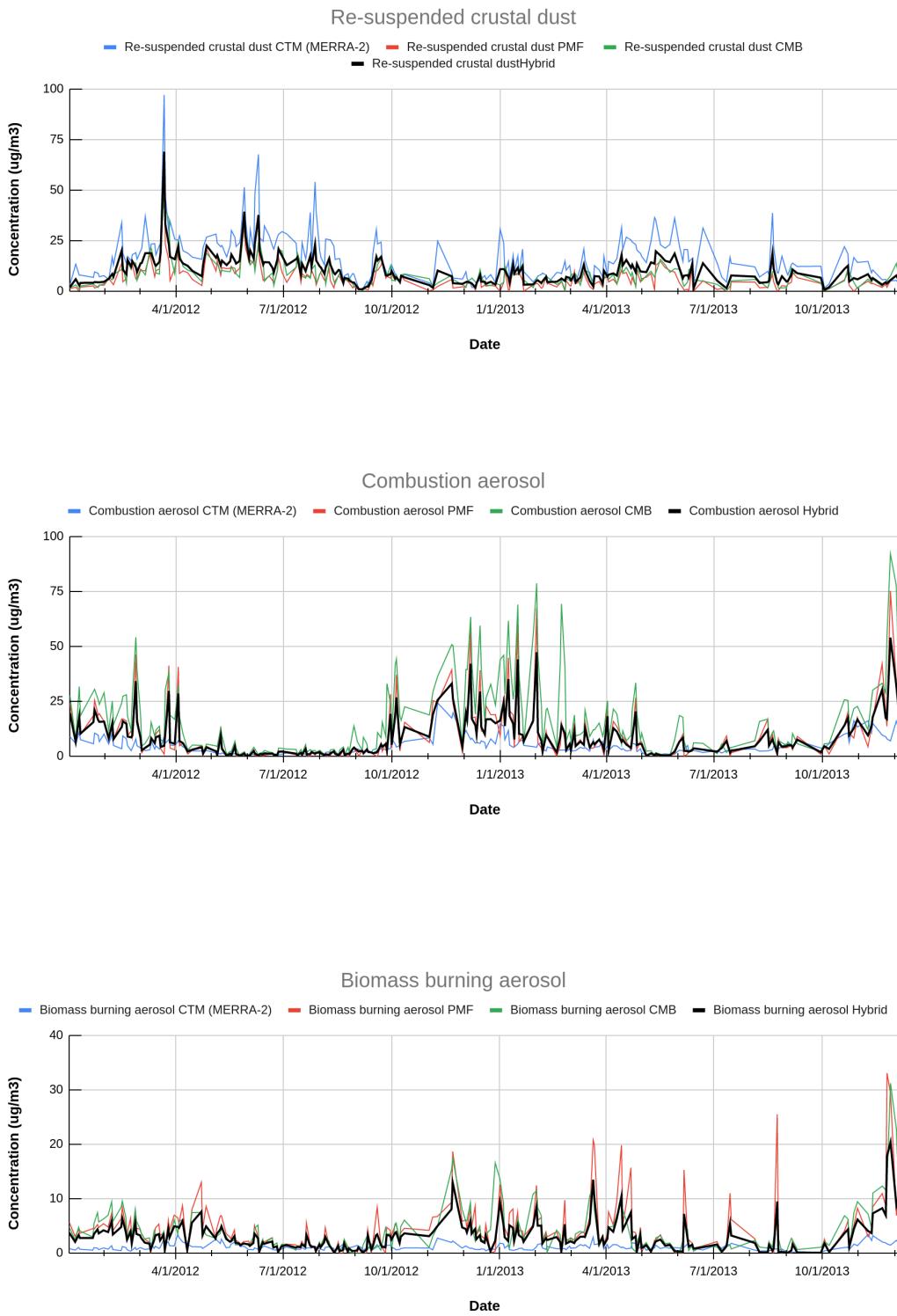
Sources	Sectors	PM			NMVOC			NOX			BC		
		PSCF	CWT	EN									
Biomass burning aerosol	REBM+AGBR	74.12	75.31	75.93	76.44	77.08	77.96	75.41	76.48	77.17	78.43	78.99	79.93
Combustion aerosol	ELGN+INDS+TRAN	62.07	59.45	60.32	57.14	56.49	57.35	75.74	73.52	76.23	83.58	80.78	82.81
Secondary Sulphate	ELGN+INDS	75.38	69.60	73.45	74.51	70.00	76.17	76.73	72.48	77.58	80.80	77.44	80.94
Sources	Sectors	PM			NMVOC			NOX			BC		
Pyrolysis carbon-rich aerosol	REBM	93.00	95.39	95.55	93.44	95.89	96.78	-	-	-	-	-	-
Secondary Nitrate	TRAN+DDSL	93.47	96.85	97.62	-	-	-	95.02	99.09	99.51	-	-	-

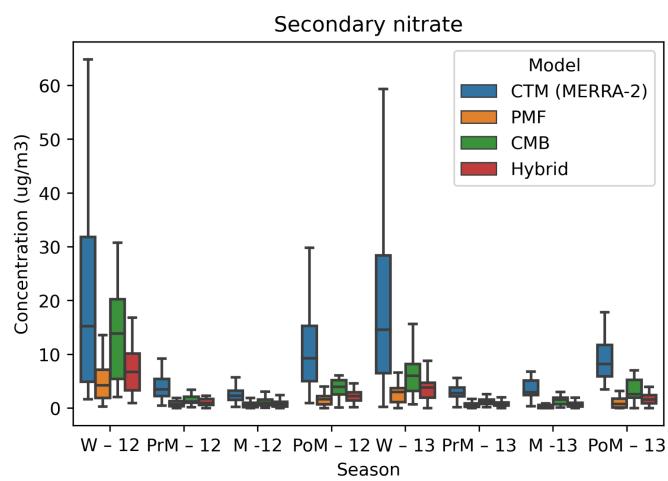
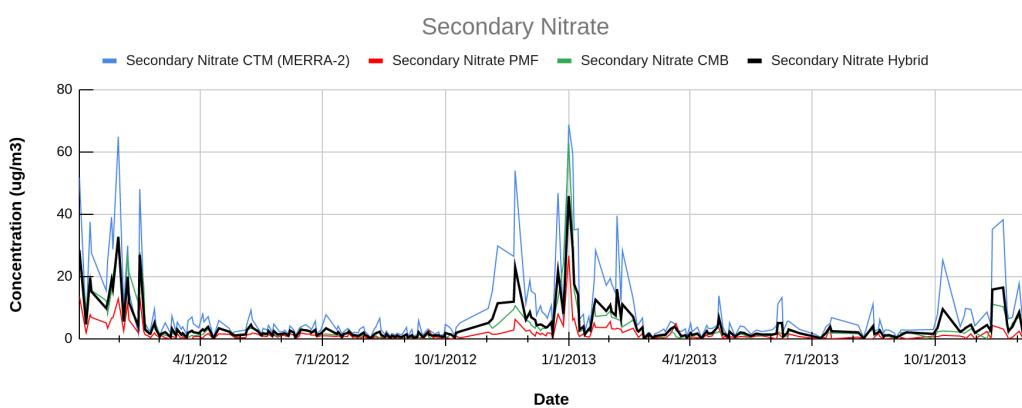
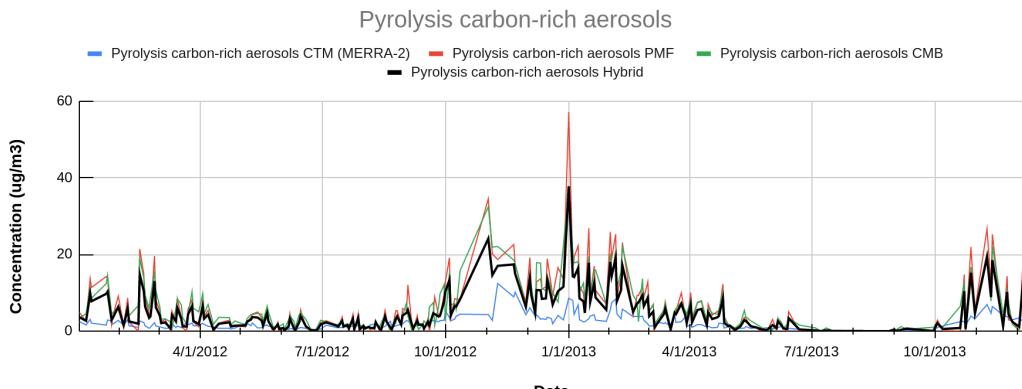
Table A1. The “r-statistic” values for various VVNP factor obtained after statistically comparing the EI species and corresponding sectors and trajectory ensemble maps (PSCF/CWT/Ensemble) are given and the “ - ” represents that the given pair of factors and species are not compared.

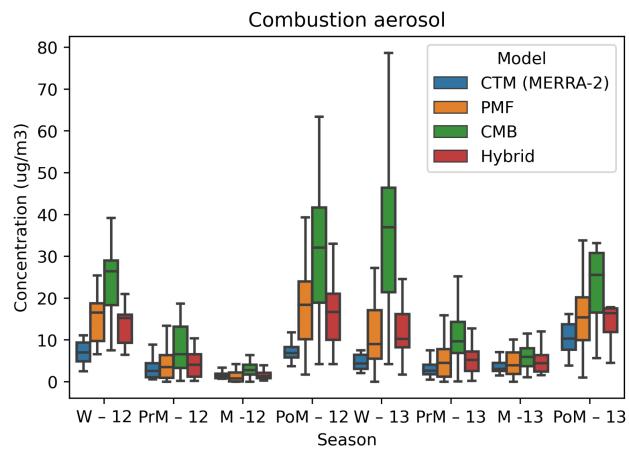
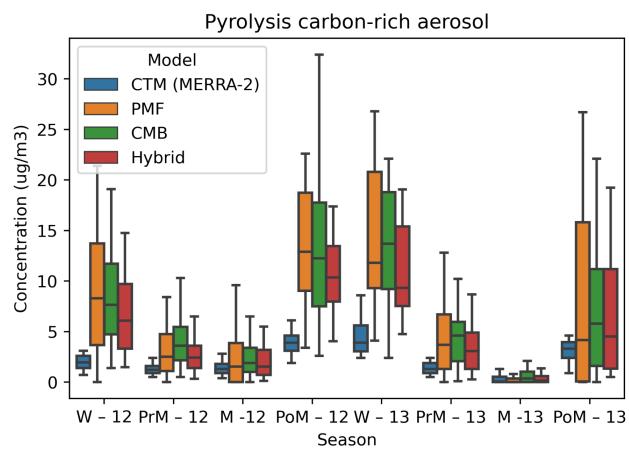
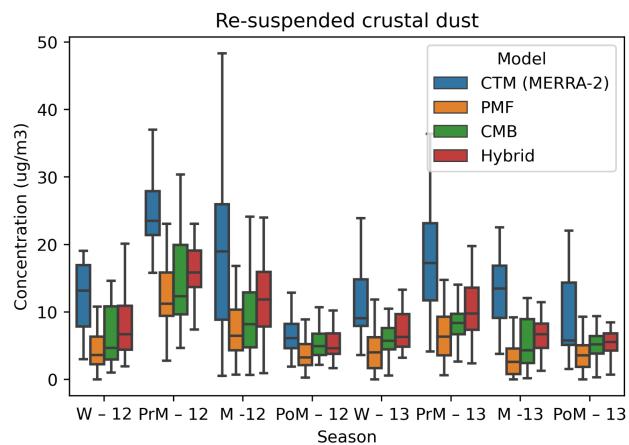
Appendix II- Hybrid Model

Table A2.1. Statistical Parameters used in CMB8.2 to assess the model performance.

Output/Statistic	Abbreviation	Target	Explanation
Std. Error	STD ERR	<< SCE	The standard error of the SCE.
T-statistic	T-STAT	> 2.0	The ratio of the value of the SCE to the uncertainty in the SCE. A T-STAT greater than 2 means that the SCE has a relative uncertainty of less than 50%. T-STAT = SCE/STD ERR
R-square	R-SQUARE	0.8 to 1.0	A measure of the variance of the ambient concentration explained by the calculated concentration. The target range is 0.8 to 1.0, where an r-square of 1.0 is perfect.
Chi-square	CHI-SQUARE	0.0 to 4.0	A term that compares the difference between the calculated and measured ambient concentrations to the uncertainty of the difference. A perfect fit has a chi-square of 0, and a chi-square less than 2 usually indicates a good fit. The target range is 0.0 to 4.0.
Percent Mass Explained	% MASS	100% ± 20%	The ratio of the total calculated to measured mass. The target range is 80% to 120%. % MASS = $M_c/M_m \times 100$
Degrees of Freedom	DF	> 5	The difference between the number of fitting species and the number of fitting sources. This value must exceed 1 and should be greater than 5.
Uncertainty/Similrity Clusters	U/S CLUSTERS	None	A list of sources that were not sufficiently resolved by the CMB analysis. No clustering is preferred.
Ratio of Calculated to Measured	RATIO C/M	0.5 to 2.0	The ratio of the calculated to measured concentration of an ambient species. Ideally, this value should be 1.0, but the target range is 0.5 to 2.0. RATIO C/M = C_c/M_m for each species <i>I</i> .
Ratio of Residual to Uncertainty	RATIO R/U	2.0 to 2.0	The ratio of the residual (calculated minus measured) to the uncertainty of the residual (square root of the sum of squares of the uncertainties). Target range is -2.0 to 2.0







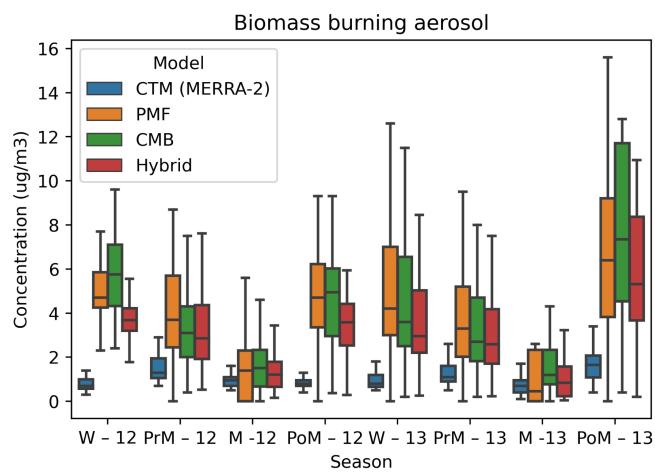
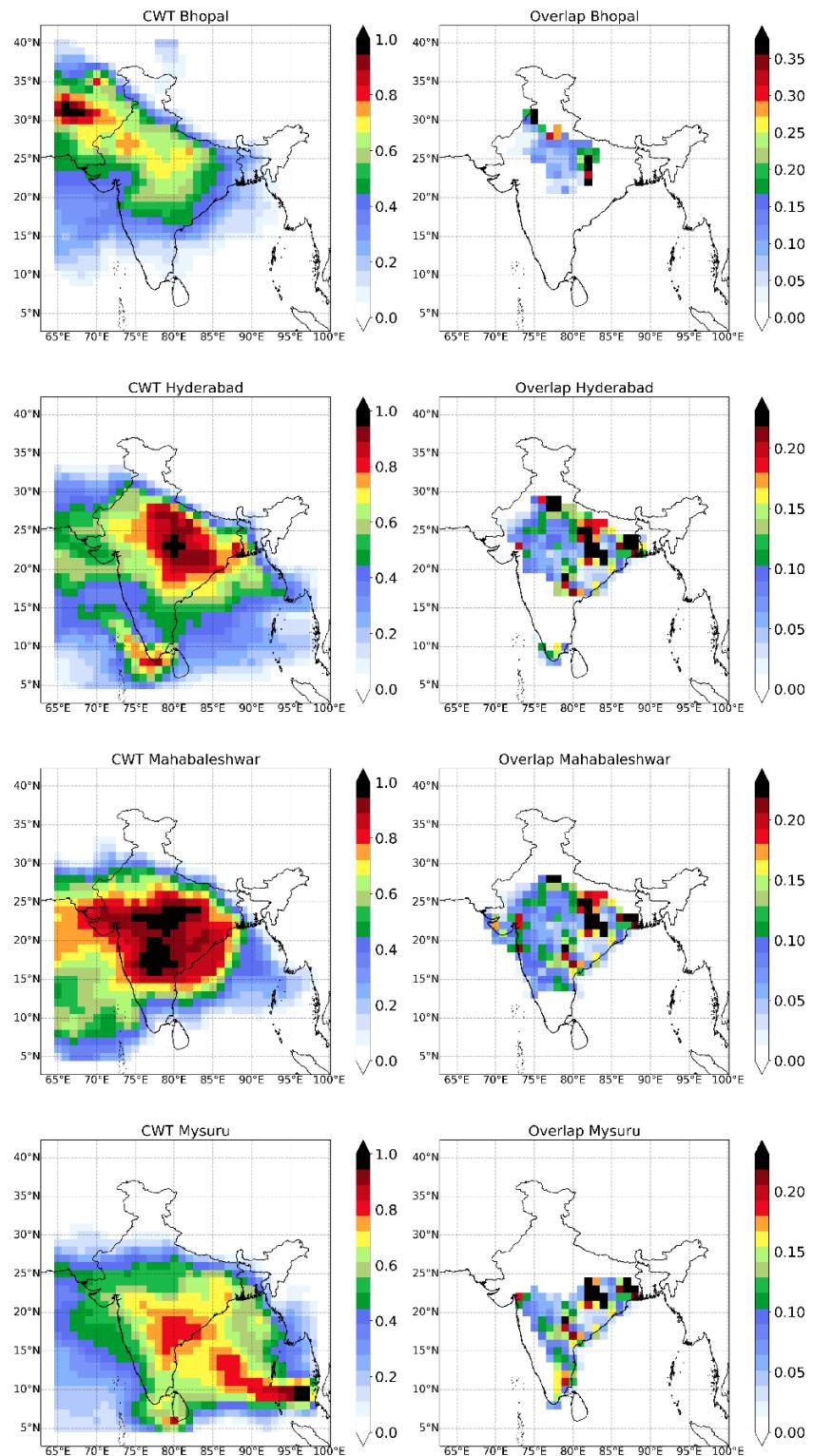
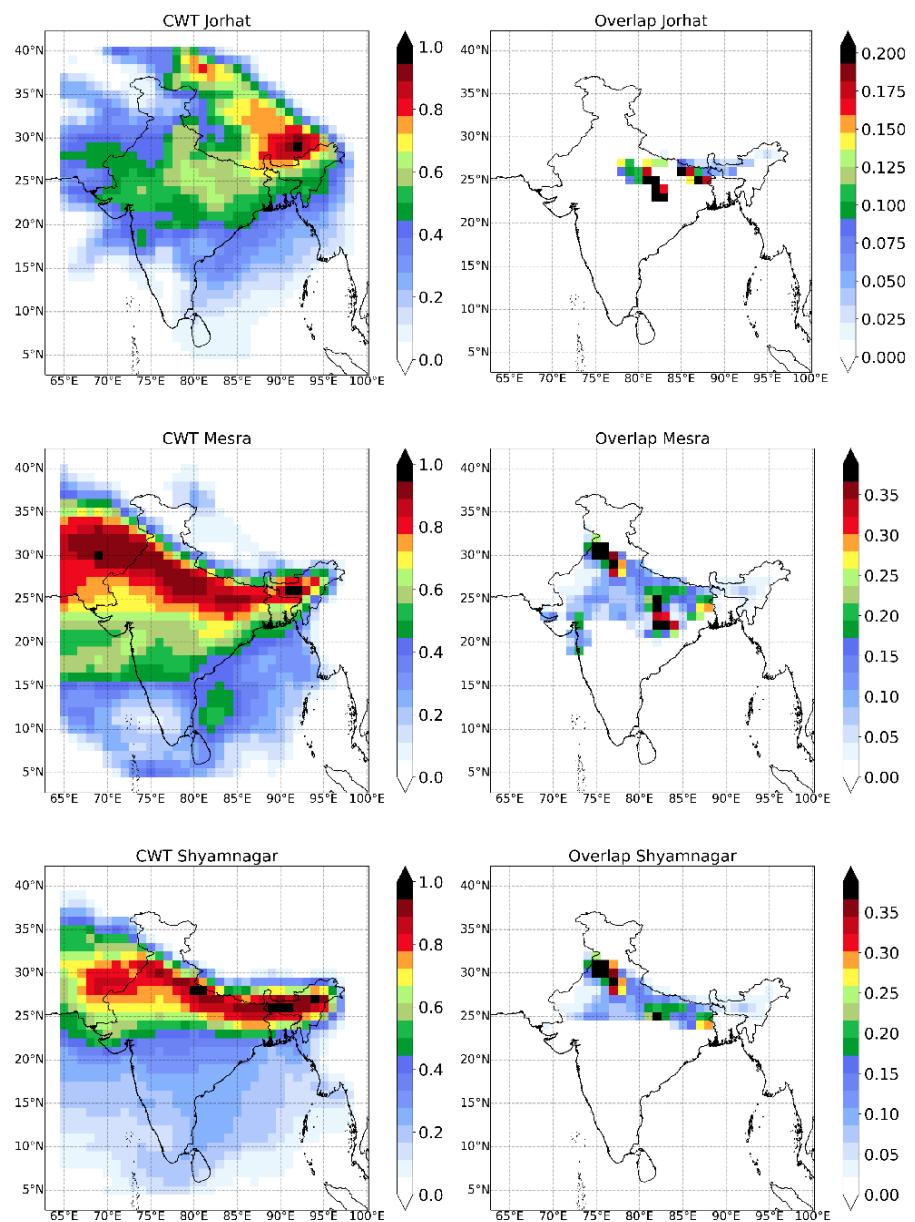


Figure A2.1. Daily time-series plots and seasonal box and whiskers plot showing PM2.5 source contributions (mass concentration) from each model for different sources

Appendix III - Trajectory Ensemble Results - COALASCE Network





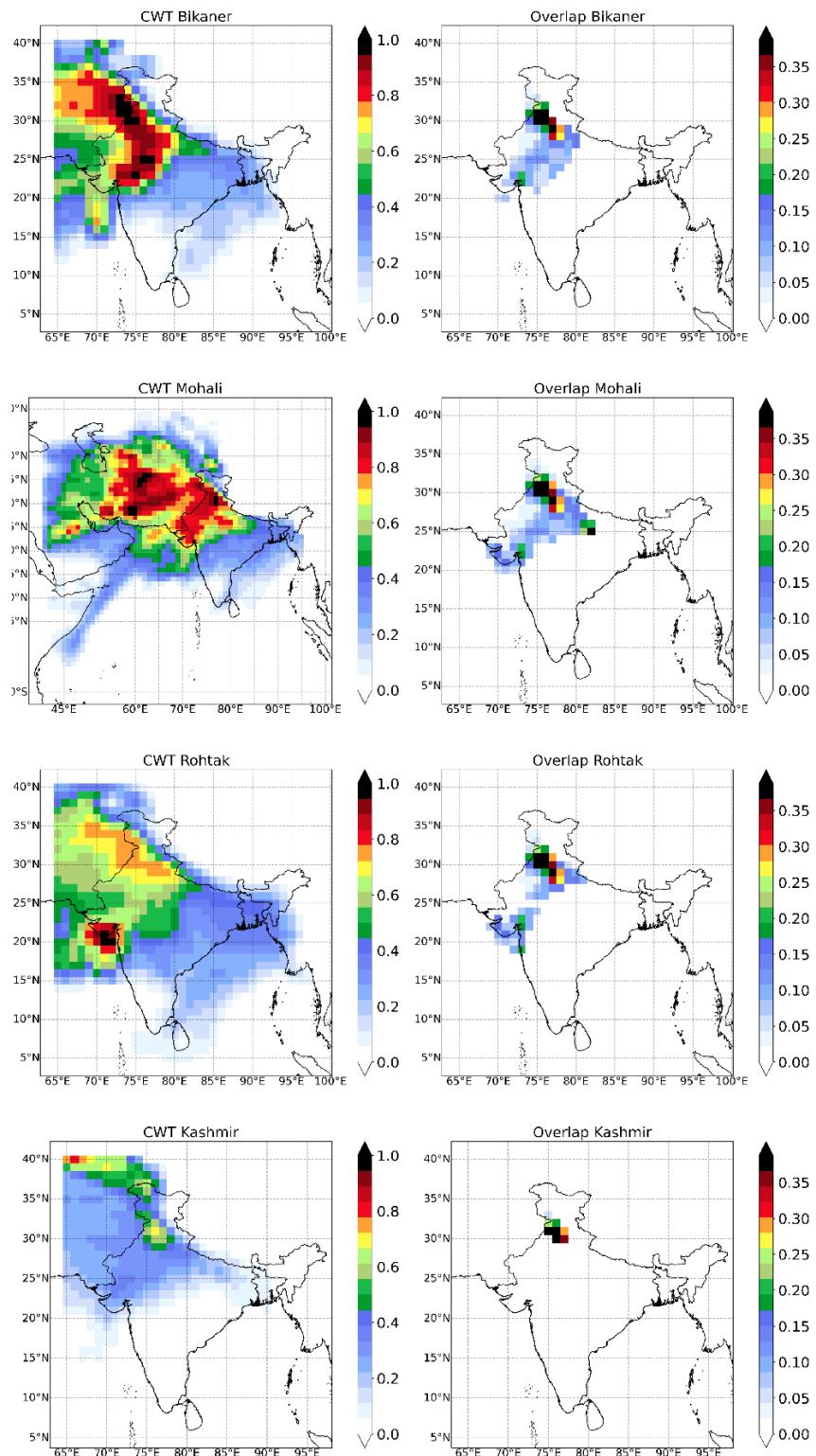


Figure A3.1. CWT Maps and their Overlaps with EI for each site of the COALESCE network regional clusters

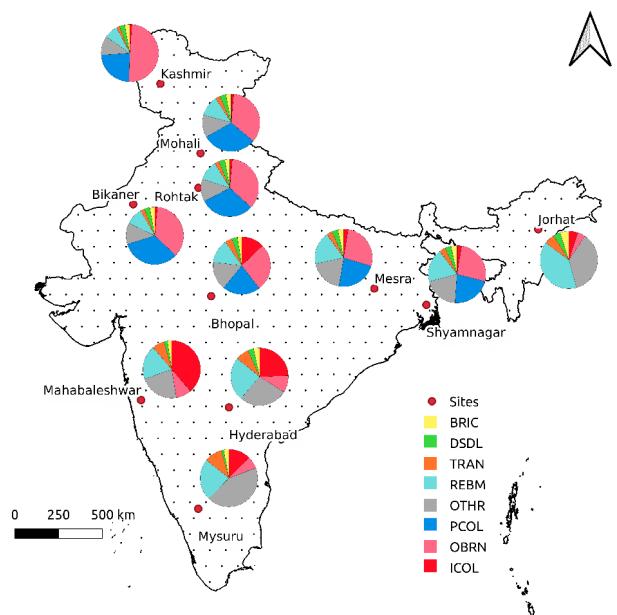
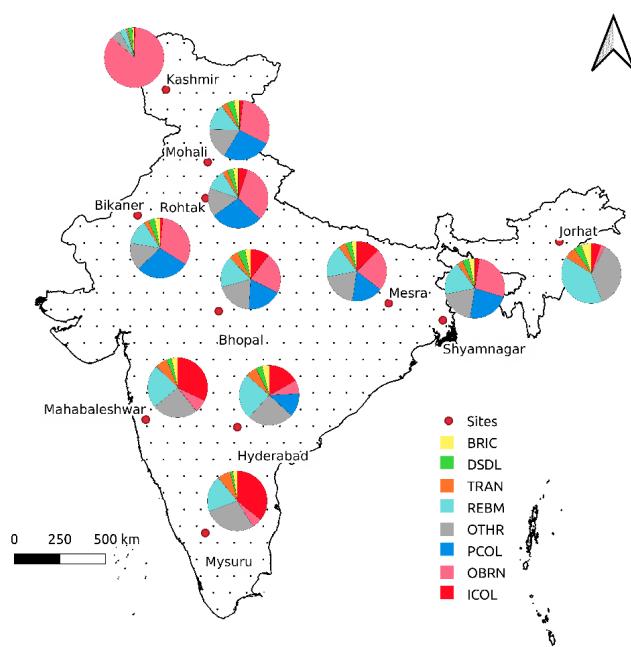


Figure A3.2. Sector-wise contribution to primary PM_{2.5} mass over 11 site COALESCE network across India for 2019 (PSCF-top and CWT-bottom)

References

- Ashbaugh, Lowell L., William C. Malm, and Willy Z. Sadeh. 1985. "A Residence Time Probability Analysis of Sulfur Concentrations at Grand Canyon National Park." *Atmospheric Environment* 19 (8): 1263–70.
- Balachandran, S., J. E. Pachon, Y. Hu, and D. Lee. 2012. "Ensemble-Trained Source Apportionment of Fine Particulate Matter and Method Uncertainty Analysis." *Atmospheric*. <https://www.sciencedirect.com/science/article/pii/S1352231012007030>.
- Bano, Shahina, Shamsh Pervez, Judith C. Chow, Jeevan Lal Matawle, John G. Watson, Rakesh Kumar Sahu, Anjali Srivastava, Suresh Tiwari, Yasmeen Fatima Pervez, and Manas Kanti Deb. 2018. "Coarse Particle (PM₁₀–2.5) Source Profiles for Emissions from Domestic Cooking and Industrial Process in Central India." *The Science of the Total Environment* 627 (June): 1137–45.
- Barregard, Lars, Gerd Sällsten, Pernilla Gustafson, Lena Andersson, Linda Johansson, Samar Basu, and Lennart Stigendal. 2006. "Experimental Exposure to Wood-Smoke Particles in Healthy Humans: Effects on Markers of Inflammation, Coagulation, and Lipid Peroxidation." *Inhalation Toxicology* 18 (11): 845–53.
- Belis, C. A., F. Karagulian, B. R. Larsen, and P. K. Hopke. 2013. "Critical Review and Meta-Analysis of Ambient Particulate Matter Source Apportionment Using Receptor Models in Europe." *Atmospheric Environment* 69 (April): 94–108.
- Belis, C. A., D. Pernigotti, G. Pirovano, O. Favez, J. L. Jaffrezo, J. Kuenen, H. Denier van Der Gon, et al. 2020. "Evaluation of Receptor and Chemical Transport Models for PM₁₀ Source Apportionment." *Atmospheric Environment: X* 5 (January): 100053.
- Brinkman, Gregory, Gary Vance, Michael P. Hannigan, and Jana B. Milford. 2006. "Use of Synthetic Data to Evaluate Positive Matrix Factorization as a Source Apportionment Tool for PM_{2.5} Exposure Data." *Environmental Science & Technology*. <https://doi.org/10.1021/es051712y>.
- Brown, Steven G., Shelly Eberly, Pentti Paatero, and Gary A. Norris. 2015. "Methods for Estimating Uncertainty in PMF Solutions: Examples with

-
- Ambient Air and Water Quality Data and Guidance on Reporting PMF Results.” *The Science of the Total Environment* 518-519 (June): 626–35.
- Carslaw, David C., and Karl Ropkins. 2012. “Openair—an R Package for Air Quality Data Analysis.” *Environmental Modelling & Software* 27: 52–61.
- Choi, Hyun-Deok, Thomas M. Holsen, and Philip K. Hopke. 2008. “Atmospheric Mercury (Hg) in the Adirondacks: Concentrations and Sources.” *Environmental Science & Technology* 42 (15): 5644–53.
- Coulter, Thomas C. 2004. *EPA-CMB8.2 Users Manual*. U.S. Environmental Protection Agency, Office of Air Quality Planning & Standards, Emissions, Monitoring & Analysis Division, Air Quality Modeling Group.
- Draxler, R. R., and G. D. Rolph. 2010. “HYSPPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model Access via NOAA ARL READY Website (<http://ready.Arl.Noaa.gov/HYSPLIT.Php>), NOAA Air Resources Laboratory.” *Silver Spring, MD* 25. https://www.ready.noaa.gov/hysplit_metadata.html.
- Dutton, Steven J., Sverre Vedal, Ricardo Piedrahita, Jana B. Milford, Shelly L. Miller, and Michael P. Hannigan. 2010. “Source Apportionment Using Positive Matrix Factorization on Daily Measurements of Inorganic and Organic Speciated PM_{2.5}.” *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2010.04.038>.
- Feng, Shaolong, Dan Gao, Fen Liao, Furong Zhou, and Xinming Wang. 2016. “The Health Effects of Ambient PM_{2.5} and Potential Mechanisms.” *Ecotoxicology and Environmental Safety* 128 (June): 67–74.
- Gelaro, Ronald, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles, et al. 2017. “The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).” *Journal of Climate*. <https://doi.org/10.1175/jcli-d-16-0758.1>.
- Gordon, G. E. 1988. “Receptor Models.” *Environmental Science & Technology* 22 (10): 1132–42.
- Grell, Georg A., Steven E. Peckham, Rainer Schmitz, Stuart A. McKeen, Gregory Frost, William C. Skamarock, and Brian Eder. 2005. “Fully Coupled ‘online’ Chemistry within the WRF Model.” *Atmospheric Environment* 39 (37): 6957–75.
- Han, Young-Ji, Thomas M. Holsen, and Philip K. Hopke. 2007. “Estimation of

-
- Source Locations of Total Gaseous Mercury Measured in New York State Using Trajectory-Based Models.” *Atmospheric Environment* 41 (28): 6033–47.
- Henry, Ronald C., Charles W. Lewis, Philip K. Hopke, and Hugh J. Williamson. 1984. “Review of Receptor Model Fundamentals.” *Atmospheric Environment* 18 (8): 1507–15.
- Hopke, Philip K. 2016. “Review of Receptor Modeling Methods for Source Apportionment.” *Journal of the Air & Waste Management Association* 66 (3): 237–59.
- Hopke, Philip K., Liming Zhou, and Richard L. Poirot. 2005. “Reconciling Trajectory Ensemble Receptor Model Results with Emissions.” *Environmental Science & Technology* 39 (20): 7980–83.
- Hopke, P. K., L. A. Barrie, S-M Li, M-D Cheng, C. Li, and Y. Xie. 1995. “Possible Sources and Preferred Pathways for Biogenic and Non-Sea-Salt Sulfur for the High Arctic.” *Journal of Geophysical Research* 100 (D8): 16595.
- Hsu, Ying-Kuang, Thomas M. Holsen, and Philip K. Hopke. 2003. “Comparison of Hybrid Receptor Models to Locate PCB Sources in Chicago.” *Atmospheric Environment*. [https://doi.org/10.1016/s1352-2310\(02\)00886-5](https://doi.org/10.1016/s1352-2310(02)00886-5).
- Ivey, Cesunica, Heather Holmes, Guoliang Shi, Sivaraman Balachandran, Yongtao Hu, and Armistead G. Russell. 2017. “Development of PM2.5 Source Profiles Using a Hybrid Chemical Transport-Receptor Modeling Approach.” *Environmental Science & Technology* 51 (23): 13788–96.
- Jaiprakash, Amrita Singhai, Gazala Habib, Ramya Sunder Raman, and Tarun Gupta. 2017. “Chemical Characterization of PM1.0 Aerosol in Delhi and Source Apportionment Using Positive Matrix Factorization.” *Environmental Science and Pollution Research* 24 (1): 445–62.
- Keeler, Gerald Joseph. 1987. “A Hybrid Approach for Source Apportionment of Atmospheric Pollutants in the Northeastern United States.” <https://ui.adsabs.harvard.edu/abs/1987PhDT.....96K>.
- Kim, Eugene. 2004. “Improving Source Identification of Fine Particles in a Rural Northeastern U.S. Area Utilizing Temperature-Resolved Carbon Fractions.” *Journal of Geophysical Research*. <https://doi.org/10.1029/2003jd004199>.
- Kim, In Sun, Daehyun Wee, Yong Pyo Kim, and Ji Yi Lee. 2016. “Development

-
- and Application of Three-Dimensional Potential Source Contribution Function (3D-PSCF)." *Environmental Science and Pollution Research International* 23 (17): 16946–54.
- Kumar, Samresh, and Ramya Sunder Raman. 2020. "Source Apportionment of Fine Particulate Matter over a National Park in Central India." *The Science of the Total Environment* 720 (June): 137511.
- Lee, Dongho, Sivaraman Balachandran, Jorge Pachon, Roshini Shankaran, Sangil Lee, James A. Mulholland, and Armistead G. Russell. 2009. "Ensemble-Trained PM2.5 Source Apportionment Approach for Health Studies." *Environmental Science & Technology* 43 (18): 7023–31.
- Lee, Sangil, Wei Liu, Yuhang Wang, Armistead G. Russell, and Eric S. Edgerton. 2008. "Source Apportionment of PM2.5: Comparing PMF and CMB Results for Four Ambient Monitoring Sites in the Southeastern United States." *Atmospheric Environment* 42 (18): 4126–37.
- Lekinwala, Nirav L., Ankur Bharadwaj, Ramya Sunder Raman, Mani Bhushan, Kunal Bali, and Sagnik Dey. 2020. "Weight-of-Evidence Approach to Identify Regionally Representative Sites for Air-Quality Monitoring Network: Satellite Data-Based Analysis." *MethodsX* 7 (June): 100949.
- Lin, John C., and Deyong Wen. 2015. "A Method to Quantitatively Apportion Pollutants at High Spatial and Temporal Resolution: The Stochastic Lagrangian Apportionment Method (SLAM)." *Environmental Science & Technology* 49 (1): 351–60.
- Maier, Marissa L., Sivaraman Balachandran, Stefanie E. Sarnat, Jay R. Turner, James A. Mulholland, and Armistead G. Russell. 2013. "Application of an Ensemble-Trained Source Apportionment Approach at a Site Impacted by Multiple Point Sources." *Environmental Science & Technology* 47 (8): 3743–51.
- Masiol, M., S. Squizzato, M. D. Cheng, and D. Q. Rich. 2019. "Differential Probability Functions for Investigating Long-Term Changes in Local and Regional Air Pollution Sources." *Aerosol and Air Quality Resarch*. <https://www.osti.gov/biblio/1504018>.
- Matawle, Jeevanlal, Shamsh Pervez, Shippi Dewangan, Suresh Tiwari, Deewan Singh Bisht, Yasmeen F. Pervez, and Others. 2014. "PM2. 5 Chemical Source Profiles of Emissions Resulting from Industrial and Domestic Burning Activities in India." *Aerosol and Air Quality Research* 14 (7):

2051–66.

- Ma, Xiaodan, Peng Yan, Tianliang Zhao, Xiaofang Jia, Jian Jiao, Qianli Ma, Dongqiao Wu, Zhuozhi Shu, Xiaoyun Sun, and Birhanu Asmerom Habtemicheal. 2021. “Evaluations of Surface PM10 Concentration and Chemical Compositions in MERRA-2 Aerosol Reanalysis over Central and Eastern China.” *Remote Sensing*. <https://doi.org/10.3390/rs13071317>.
- Navinya, Chimurkar D., V. Vinoj, and Satyendra K. Pandey. 2020. “Evaluation of PM2.5 Surface Concentrations Simulated by NASA’s MERRA Version 2 Aerosol Reanalysis over India and Its Relation to the Air Quality Index.” *Aerosol and Air Quality Research*. <https://doi.org/10.4209/aaqr.2019.12.0615>.
- Norris, Gary, Rachelle Duvall, Steve Brown, and Song Bai. 2008. “EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide.” *Prepared for the US Environmental Protection Agency, Washington, DC, by the National Exposure Research Laboratory, Research Triangle Park*.
- Ostro, Bart, Wen-Ying Feng, Rachel Broadwin, Shelley Green, and Michael Lipsett. 2007. “The Effects of Components of Fine Particulate Air Pollution on Mortality in California: Results from CALFINE.” *Environmental Health Perspectives* 115 (1): 13–19.
- Paatero, P., S. Eberly, S. G. Brown, and G. A. Norris. 2014. “Methods for Estimating Uncertainty in Factor Analytic Solutions.” *Atmospheric Measurement Techniques* 7 (3): 781–97.
- Paatero, Pentti, Philip K. Hopke, Bilkis A. Begum, and Swapan K. Biswas. 2005. “A Graphical Diagnostic Method for Assessing the Rotation in Factor Analytical Models of Atmospheric Pollution.” *Atmospheric Environment* 39 (1): 193–201.
- Paatero, Pentti, and Unto Tapper. 1993. “Analysis of Different Modes of Factor Analysis as Least Squares Fit Problems.” *Chemometrics and Intelligent Laboratory Systems* 18 (2): 183–94.
- Pandey, Apoorva, Pankaj Sadavarte, Anand B. Rao, and Chandra Venkataraman. 2014. “Trends in Multi-Pollutant Emissions from a Technology-Linked Inventory for India: II. Residential, Agricultural and Informal Industry Sectors.” *Atmospheric Environment* 99 (December): 341–52.
- Patil, Rashmi S., Rakesh Kumar, Ratish Menon, Munna Kumar Shah, and

-
- Virendra Sethi. 2013. "Development of Particulate Matter Speciation Profiles for Major Sources in Six Cities in India." *Atmospheric Research* 132-133 (October): 1–11.
- Pekney, Natalie J., Cliff I. Davidson, Liming Zhou, and Philip K. Hopke. 2006. "Application of PSCF and CPF to PMF-Modeled Sources of PM_{2.5} in Pittsburgh." *Aerosol Science and Technology: The Journal of the American Association for Aerosol Research* 40 (10): 952–61.
- Pervez, Shamsh, Shahina Bano, John G. Watson, Judith C. Chow, Jeevan Lal Matawle, Anjali Shrivastava, Suresh Tiwari, and Yasmeen Fatima Pervez. 2018. "Source Profiles for PM_{10-2.5} Resuspended Dust and Vehicle Exhaust Emissions in Central India." *Aerosol and Air Quality Research* 18 (7): 1660–72.
- Pillon, Silvia, Francesca Liguori, Fabio Dalan, and Giuseppe Maffeis. 2011. "PM Source Apportionment Analysis in the Venetian Area." *International Journal of Environment and Pollution*. <https://doi.org/10.1504/ijep.2011.047330>.
- Polissar, Alexandr V., Philip K. Hopke, Pentti Paatero, William C. Malm, and James F. Sisler. 1998. "Atmospheric Aerosol over Alaska: 2. Elemental Composition and Sources." *Journal of Geophysical Research: Atmospheres*. <https://doi.org/10.1029/98jd01212>.
- Polissar, A. V., P. K. Hopke, P. Paatero, Y. J. Kaufmann, D. K. Hall, B. A. Bodhaine, E. G. Dutton, and J. M. Harris. 1999. "The Aerosol at Barrow, Alaska: Long-Term Trends and Source Locations." *Atmospheric Environment* 33 (16): 2441–58.
- Raman, Ramya Sunder, and S. Ramachandran. 2011. "Source Apportionment of the Ionic Components in Precipitation over an Urban Region in Western India." *Environmental Science and Pollution Research International* 18 (2): 212–25.
- Raman, R. S., M. Bhushan, K. Bali, and S. Dey. 2020. "A Framework for Setting up a Country-Wide Network of Regional Surface PM_{2.5} Sampling Sites Utilising a Satellite-Derived proxy—The COALESCE Project, India." *Atmospheric*. https://www.sciencedirect.com/science/article/pii/S135223102030279X?casa_token=YciH1sXUta8AAAAA:7Wq4ypsnKvehK5X-phVq8UNITrMOy_t_V0dFnN-8mvl2OoK8it23N45x-JPwImzIbAA7yhO2AeQ.
- Randles, C. A., A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R.

-
- Govindaraju, A. Smirnov, et al. 2017. “The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation.” *Journal of Climate*. <https://doi.org/10.1175/jcli-d-16-0609.1>.
- Rotstayn, Leon D., and Ulrike Lohmann. 2002. “Tropical Rainfall Trends and the Indirect Aerosol Effect.” *Journal of Climate*. [https://doi.org/10.1175/1520-0442\(2002\)015<2103:trati>2.0.co;2](https://doi.org/10.1175/1520-0442(2002)015<2103:trati>2.0.co;2).
- Sadavarte, Pankaj, and Chandra Venkataraman. 2014. “Trends in Multi-Pollutant Emissions from a Technology-Linked Inventory for India: I. Industry and Transport Sectors.” *Atmospheric Environment* 99 (December): 353–64.
- Sarnat, Jeremy A., Amit Marmur, Mitchel Klein, Eugene Kim, Armistead G. Russell, Stefanie E. Sarnat, James A. Mulholland, Philip K. Hopke, and Paige E. Tolbert. 2008. “Fine Particle Sources and Cardiorespiratory Morbidity: An Application of Chemical Mass Balance and Factor Analytical Source-Apportionment Methods.” *Environmental Health Perspectives* 116 (4): 459–66.
- Seo, Yong-Seok, Young-Ji Han, Thomas M. Holsen, Eunhwa Choi, Kyung-Duk Zoh, and Seung-Muk Yi. 2015. “Source Identification of Total Mercury (TM) Wet Deposition Using a Lagrangian Particle Dispersion Model (LPDM).” *Atmospheric Environment* 104 (March): 102–11.
- Song, Zijue, Disong Fu, Xiaoling Zhang, Yunfei Wu, Xiangao Xia, Jianxin He, Xinlei Han, Renjian Zhang, and Huizheng Che. 2018. “Diurnal and Seasonal Variability of PM_{2.5} and AOD in North China Plain: Comparison of MERRA-2 Products and Ground Measurements.” *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2018.08.012>.
- Tai, Amos P. K., Loretta J. Mickley, and Daniel J. Jacob. 2010. “Correlations between Fine Particulate Matter (PM_{2.5}) and Meteorological Variables in the United States: Implications for the Sensitivity of PM_{2.5} to Climate Change.” *Atmospheric Environment* 44 (32): 3976–84.
- Tibrewal, Kushal, and Chandra Venkataraman. 2020. “Climate Co-Benefits of Air Quality and Clean Energy Policy in India.” *Nature Sustainability*, December, 1–9.
- Uria-Tellaetxe, Iratxe, and David C. Carslaw. 2014. “Conditional Bivariate Probability Function for Source Identification.” *Environmental Modelling & Software* 59 (September): 1–9.

-
- Venkataraman, Chandra, Michael Brauer, Kushal Tibrewal, Pankaj Sadavarte, Qiao Ma, Aaron Cohen, Sreelekha Chaliyakunnel, et al. 2018. "Source Influence on Emission Pathways and Ambient PM_{2.5} Pollution over India (2015-2050)." *Atmospheric Chemistry and Physics* 18 (11): 8017–39.
- Wang, Jiandong, Shuxiao Wang, A. Scott Voorhees, Bin Zhao, Carey Jang, Jingkun Jiang, Joshua S. Fu, Dian Ding, Yun Zhu, and Jiming Hao. 2015. "Assessment of Short-Term PM_{2.5}-Related Mortality due to Different Emission Sources in the Yangtze River Delta, China." *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2015.05.060>.
- Wang, Y. Q., X. Y. Zhang, and R. Arimoto. 2006. "The Contribution from Distant Dust Sources to the Atmospheric Particulate Matter Loadings at XiAn, China during Spring." *The Science of the Total Environment* 368 (2-3): 875–83.
- Watson, J. G., N. F. Robinson, J. C. Chow, R. C. Henry, B. M. Kim, T. G. Pace, E. L. Meyer, and Q. Nguyen. 1990. "The USEPA/DRI Chemical Mass Balance Receptor Model, CMB 7.0." *Environmental Software* 5 (1): 38–49.
- Watson, J. G., N. F. Robinson, C. Lewis, and T. Coulter. 1997. "Chemical Mass Balance Receptor Model Version 8 (CMB8) User's Manual." *Prepared for US*.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.2872&rep=rep1&type=pdf>.
- Watson, John G., Judith C. Chow, Zhiqiang Lu, Eric M. Fujita, Douglas H. Lowenthal, Douglas R. Lawson, and Lowell L. Ashbaugh. 1994. "Chemical Mass Balance Source Apportionment of PM₁₀ during the Southern California Air Quality Study." *Aerosol Science and Technology*. <https://doi.org/10.1080/02786829408959693>.
- Watson, John G., John A. Cooper, and James J. Huntzicker. 1984. "The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model." *Atmospheric Environment* 18 (7): 1347–55.
- Zhou, Liming, Philip K. Hopke, and Wei Liu. 2004. "Comparison of Two Trajectory Based Models for Locating Particle Sources for Two Rural New York Sites." *Atmospheric Environment* 38 (13): 1955–63.
- Zong, Zheng, Xiaoping Wang, Chongguo Tian, Yingjun Chen, Shanfei Fu, Lin Qu, Ling Ji, Jun Li, and Gan Zhang. 2018. "PMF and PSCF Based Source Apportionment of PM_{2.5} at a Regional Background Site in North China." *Atmospheric Research* 203 (May): 207–15.

