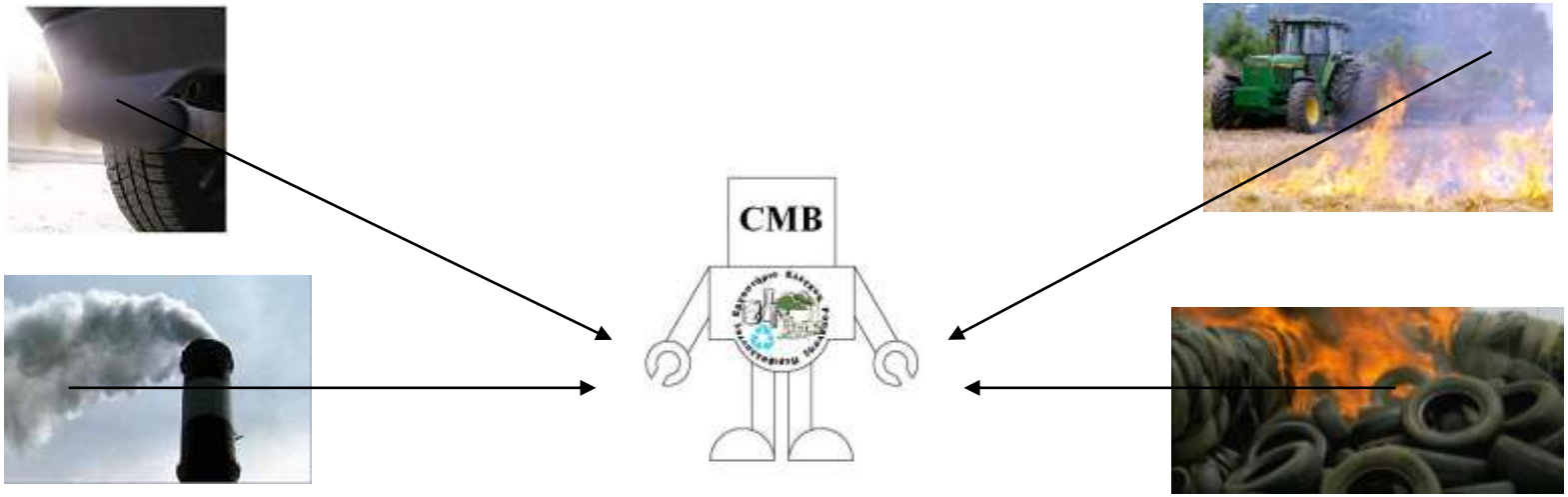
ARISTOTLE UNIVERSITY
OF THESSALONIKI

SCHOOL OF CHEMISTRY / ENVIRONMENTAL
POLLUTION CONTROL LABORATORY

The Robotic CMB: an advanced computational procedure for source apportionment of atmospheric PM

G. Argyropoulos, C. Samara



FAIRMODE Technical Meeting
28th-29th April 2014 / NILU Conference Center, Kjeller, Norway

Introduction-Fundamental Concepts of CMB modeling

Chemical Mass Balance (CMB) modeling is realized by solving an over determined system of linear equations which express ambient concentrations of chemical species measured at the receptor site as sums of contributions from individual sources:

$$C_i = \sum_{j=1}^{n} a_{ij} S_j$$

$$i = 1, 2, .., m \qquad m > n$$
$$j = 1, 2, \ldots, n$$

Where C_i is the mass concentration of chemical species *i* in ambient PM, α_{ij} is the mass fraction of chemical species *i* in the PM emitted from source *j*, and S_j is the contribution of source *j* to the total mass of ambient PM.

The main assumptions on which CMB models rely can be summarized as follows:

- **All the sources, contributing significantly to a receptor site, have been identified and have had their emissions chemically characterized.**
- **Chemical species do not react with each other, i.e. they add linearly.**
- **Compositions of source emissions remain constant during ambient and source sampling.**
- **Source compositions are linearly independent of each other.**
- **Measurement uncertainties are random, uncorrelated and normally distributed.**

Least Squares (LS) Estimators of Source Contributions (S_j)

- ***Ordinary Least Squares (OLS)***

The OLS fitting method can estimate a set of probable values for the source contributions *Sj*, by minimizing the following likelihood function:

$$\chi^2 = \sum_{i=1}^{m}\left(C_i - \sum_{j=1}^{n} a_{ij}S_j\right)^2$$

System of Normal Equations

$$\frac{\partial \chi^2}{\partial S_j} = 0 \Rightarrow A^T \cdot A \cdot S = A^T \cdot C \Rightarrow$$

$$\Rightarrow S = (A^T \cdot A)^{-1} \cdot A^T \cdot C$$

Where $C = \begin{bmatrix} C_1 \\ \dots \\ C_m \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$ $S = \begin{bmatrix} S_1 \\ \dots \\ S_n \end{bmatrix}$

and superscript *T* denotes the transpose matrix

- ***Ordinary Weighted Least Squares (OWLS)***

The OWLS fitting method can estimate a set of probable values for the source contributions (*Sj*), by minimizing the following likelihood function, in which the heteroscedasticity of the receptor's chemical data is reflected as well (Friedlander, 1973):

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(C_i - \sum_{j=1}^{n} a_{ij} S_j \right)^2}{\sigma_{C_i}^2} \qquad \frac{\partial \chi^2}{\partial S_j} = 0 \Rightarrow \ldots \Rightarrow S = (A^T \cdot W \cdot A)^{-1} \cdot A^T \cdot W \cdot C$$

Where $W = \begin{bmatrix} \sigma_{C_1}^{-2} & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \sigma_{C_m}^{-2} \end{bmatrix}$

And σ_{C_i} denotes the typical error in the measurement of C_i.

- ***The Britt and Luecke Algorithm***

The Britt and Luecke algorithm consists of an iterative procedure that may estimate a set of probable values for the source contributions S_j, in which all the measurement uncertainties are reflected, by minimizing the following likelihood function (Britt and Luecke, 1973;Watson et al., 1984):

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(C_i - \sum_{j=1}^{n} \overline{a_{ij}} S_j \right)^2}{\sigma_{C_i}^2} + \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{a_{ij} - \overline{a_{ij}}^2}{\sigma_{a_{ij}}^2}$$

Where the over bars indicate the “real” values of the mass fractions

Iteration Steps of the Britt and Luecke Algorithm

- **All the estimates for the source contributions S_j are initially set equal to zero.**
- **The diagonal elements of the effective weighting matrix $(V_e^k)^{-1}$ are determined according to the following relationship:**

$$\left(V_{e,ii}^{k}\right)^{-1} = \left(\sigma_{C_i}^2 + \sum_{j=1}^{n} \sigma_{\alpha_{ij}}^2 \cdot \left(S_j^k\right)^2\right)^{-1}$$

Where superscript *k* indicates the iteration's number.

- **The new estimates for the "real" values of the mass fractions are calculated for each source profile (*j=1,2,…, n*) by:**

$$\overline{A_j^{k+1}} = \overline{A_j^k} + S_j^k \cdot V_{A_j} \cdot \overline{V_e^k}^{-1} \cdot \left[I - \overline{A^k} \cdot \overline{A^{kT}} \cdot \overline{V_e^k}^{-1} \overline{A^k}^{-1} \cdot \overline{A^{kT}} \cdot \overline{V_e^k}^{-1} \right] \cdot \left(C - AS^k \right)$$

Where V_{A_j} is one *m x m* diagonal matrix with elements on the diagonal and *I* is the *m x m* identity matrix.

- **Finally, the new estimates for the source contributions S_j are calculated by:**

$$S^{k+1} = S^k + \overline{A^{kT}} \cdot \overline{V_e^k}^{-1} \overline{A^k}^{-1} \cdot \overline{A^{kT}} \cdot \overline{V_e^k}^{-1} \cdot C - AS^k$$

- ***The Effective Variance Weighting Least Squares (EFWLS)***

The least squares fitting method of Britt and Luecke can be simplified substantially if the differences between the "true" values of the mass fractions and the measured ones are considered as negligible, allowing for the following likelihood function to be minimized (Watson et al., 1984):

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(C_i - \sum_{j=1}^{n} a_{ij} S_j \right)^2}{\sigma_{C_i}^2 + \sum_{j=1}^{n} \sigma_{\alpha_{ij}}^2 \cdot S_j^{k\,2}}$$

This approximation (EFWLS) is currently the official method suggested by the Environmental Protection Agency of United States (US EPA) for CMB modeling.

Application of CMB models for the SA of ambient PM

- **Although CMB modeling is founded upon the "working hypothesis" that every source, which contributes significantly to the receptor site, has been identified, it is most often applied without any definite knowledge about the ones that actually do, since the identification of contributing sources may indeed be one of the major goals of a Source Apportionment (SA) study, under normal circumstances.**

- **CMB modeling also requires that source compositions remain constant over the period of ambient and source sampling, which is, nonetheless, unlikely to occur.**

- **There is virtually nothing to do too, in order to predict an occurrence of collinearity among the columns of the effectively weighted source profile matrix during run-time, in case that the algorithm of Britt and Luecke, or the EFWLS approximation has been adopted for the solution of the CMB problem.**

- **Due to those common violations of CMB assumptions, CMB modeling involves in practice first some test applications of the desired LS fitting method to over determined linear systems defined by different chemical species and/or source profiles, which have been all considered as equally probable for reflecting the true emissions at the receptor site, according to the personal judgment of the modeler.**

- **According to the US EPA Protocol for Applying and Validating the CMB model (Watson, 2004), trial CMB tests should first be realized for an averaged ambient sample, in order to obtain the so-called "initial source contribution estimates", i.e. to select a default combination of source profiles and fitting species for the ambient data.**

- **According to the same protocol, the initial source contribution estimates should then be optimized separately for each daily ambient sample, again by trial CMBs involving addition, depletion or substitution of source profiles, after taking into account additional factors, such as wind direction or the presumed temporal variation of sources such as biomass burnings.**

- **The US EPA has also established a standard set of statistical performance measures for the evaluation of trial applications, which are given in the following table.**

Diagnostic criteria of the US EPA CMB 8.2 model

Performance measure(s)	Target value(s) (US EPA)
S_j	>0
$R^2 = 1 - \frac{\chi^2}{\sum_{i=1}^{m} C_i^2 \left(\sigma_{e,ii}^k\right)^{-1}}$	≥ 0.8
$\chi^2_{red.} = \frac{\chi^2}{m-n}$	≤ 4
$\%\ mass = \frac{\sum_{j=1}^{n} S_j}{C_{mass}} \cdot 100$	$100\% \pm 20\%$
$FracEst = \frac{E}{n}$	1
$Tstat_j = \frac{S_j}{\sqrt{\mathrm{var}(S_j)}}$	≤ 2
$(Res/Uncer)_i = \frac{\left(\sigma_{e,ii}^k\right)^{-1} \cdot S_i - C_i}{\sqrt{\sigma_{C_i}^2 + \left(\sum_{j=1}^{n} S_j \cdot \sigma_{\alpha_{ij}}\right)^2}}$	$\left\| (Res/Uncer)_i \right\| \leq 2$

Overall Fitting Index

$$FitMeasure = \frac{wf_1 \cdot \left(\frac{1}{\chi^2}\right) + wf_2 \cdot R^2 + wf_3 \cdot \left(\frac{\%mass}{100}\right) + wf_4 \cdot FracEst}{wf_1 + wf_2 + wf_3 + wf_4}$$

Limitations of conventional CMB modeling

- **A major drawback of conventional CMB modeling arises from the fact that standard trial-and-error procedures are strongly subjected to the personal judgment of the modeler and his/her choices of fitting species/source profiles.**

- **The trial CMBs of standard procedures are also limited to a total number far less than the ones that can possibly be defined by a typical set of input data, usually a few hundred or so.**

- **The US EPA CMB 8.2 model, in particular, operating in Best Fit Mode, is capable of ranking, according to the Fit Measure index, a maximum of only 10 over determined linear systems, whose fitting species and source profiles must have been manually selected by the CMB modeler, using 10 pairs of species and profiles selection arrays that are provided for this purpose, by the model's graphical user interface (Coulter, 2004).**

Application of combinatory analysis for the determination of all possible choices between source profiles and/or fitting species of CMB input data

The total number $P_{N,M}$ of possible choices between the source profiles and/or fitting species of CMB input data, including M measured chemical species, and N source profiles ($M>N$) can be calculated by the following equation (Argyropoulos and Samara, 2010):

$$P_{N,M} = \sum_{J=1}^{N} \frac{N!}{J!(N-J)!} \sum_{I=J+1}^{M} \frac{M!}{I!(M-I)!}$$

- **According to the above equation, there is an astronomic total of 1, 613, 294, 846, 589 possible choices between the source profiles and/or fitting species of a typical set of CMB input data, including 23 chemical species, and 18 source profiles.**

- **585,711,642,651 of these choices also define LS that possess 5 degrees of freedom (*K*) or more, according to the following equation (Argyropoulos and Samara, 2010):**

$$P_{N,K} = \sum_{J=1}^{N} \frac{N!}{J!(N-J)!} \sum_{I=J+K}^{M} \frac{M!}{I!(M-I)!}$$

- **Even if trial applications are further limited to a sub range involving only over determined linear systems that consist of specific fitting species, there are still 262,143 that can be defined by all the possible choices between 18 source profiles, according to the following equation (Argyropoulos and Samara, 2010):**

$$P_N = \sum_{J=1}^{N} \frac{N!}{J!(N-J)!} = 2^N - 1$$

N	P_N	N	P_N
1	1	15	32767
2	3	16	65535
3	7	17	131071
4	15	18	262143
5	31	19	524287
6	63	20	1048575
7	127	21	2097151
8	255	22	4194303
9	511	23	8388607
10	1023	24	16777215
11	2047	25	33554431
12	4095	26	67108863
13	8191	27	134217727
14	16383	28	268435455
15	32767	29	536870911

- **It is apparent from the above combinatory analysis that standard trial-and-error procedures of conventional CMB modeling, such as the manually-driven Best Fit Mode of the US EPA CMB 8.2 model not only are considerably laborious, but can also be strongly susceptible to personal prejudices about the study area, because they can never rule out the mathematical probability that combinations of source profiles may fit ambient data better than the relatively few ones, tested (Argyropoulos and Samara, 2010) .**

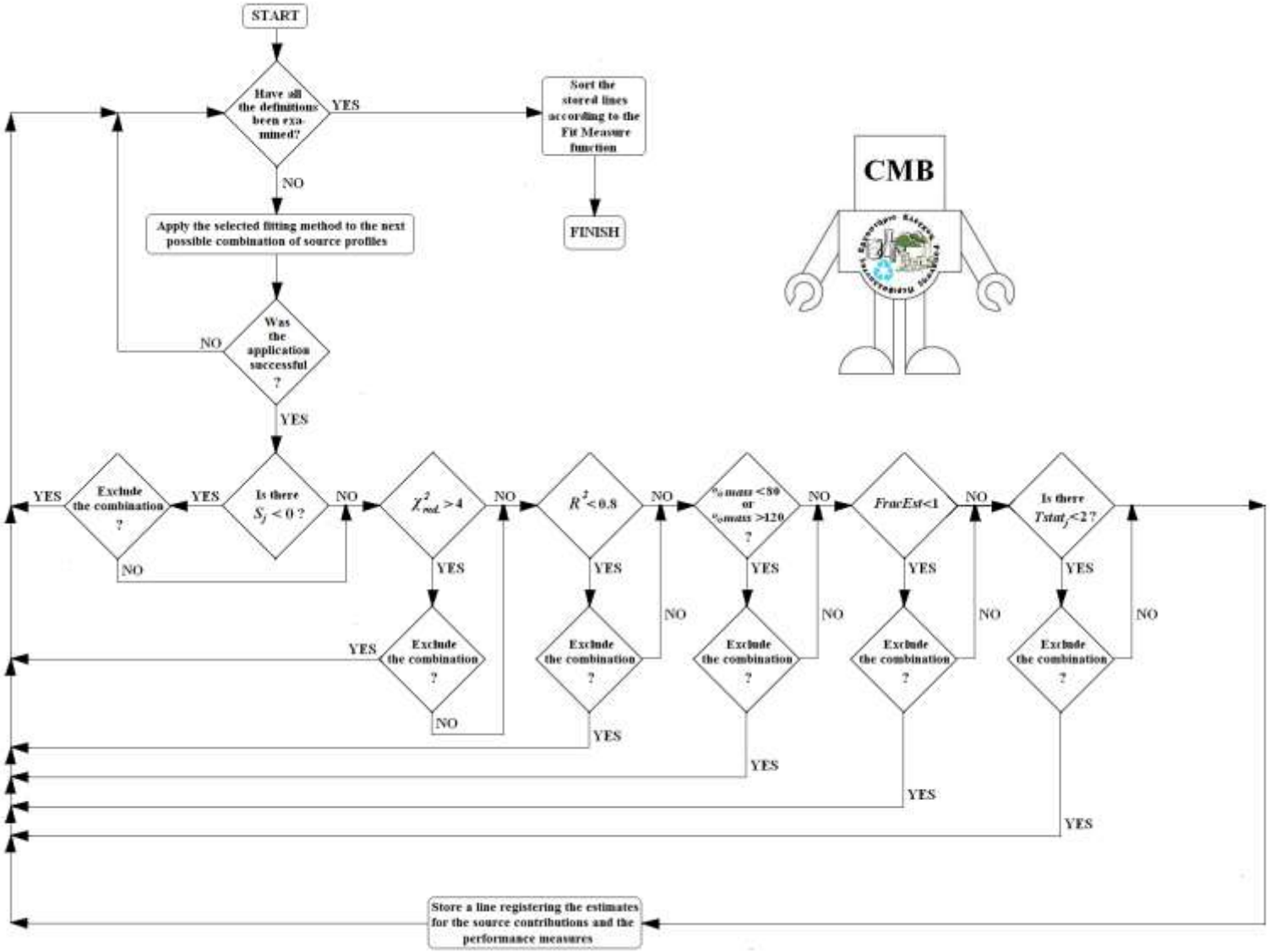
- **The latter one is indeed a fact that has indeed been acknowledged by the US EPA as well, since it is clearly stated in their protocol that "it is possible that more that one subset of source types and source profiles will fit the receptor data equally well" (Watson, 2004).**

The Robotic CMB
(Argyropoulos and Samara, 2010)

- **The Robotic CMB is realized by applying a fitting method to each and every one of the over determined linear systems that can be defined by the possible choices between the source profiles of given input data (P_N) .**

- **Each trial application that successfully converges to a solution is validated according to standard performance measures of the US EPA CMB 8.2 model, such as the absence of any negative values among the estimates for the source contributions, the values of fit indices $\chi^2_{red.}$, and R^2, the value of *%mass*, the value of fraction *FracEst*, and the values of the T-Statistic ratios. The diagnostic criteria employed for the validation of converging trials, are selected by the user, before the beginning of the computational procedure.**

- **After the fitting method has been applied to all the possible combinations of source profiles, if there are any successful applications that met the individual performace measures, they are ranked according to the overall fitting index *FitMeasure*.**



Logic Diagram of RCMB

The Boolean Function maximized by RCMB

Dependent Variable
Real Number

Independent Variables
Boolean (Either *TRUE* or *FALSE*)

$$FitMeasure = f(Choice_1, Choice_2, \ldots, Choice_N)$$

- **The explicit advantage of RCMB is that the best-fit combination of source profiles, deriving straight-forwardly from the maximization of *FitMeasure*, provides a mathematically unique solution to the conventional CMB problem, which cannot be questioned readily, unless additional information becomes available for the study area (Argyropoulos and Samara, 2010) .**

Test Case-Application of RCMB to the Crows, California PM2.5 data from the San Joaquin Valley Air Quality Study (Argyropoulos and Samara, 2010)

- **The ambient data of this set included the $PM_{2.5}$ concentrations of mass, organic carbon (OC) and elemental carbon (EC), nitrate (NO_3^-), sulfate (SO_4^{2-}), ammonium (NH_4^+), soluble sodium (Na^+) and potassium (K^+), and elemental species (Al, Si, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Br, and Pb), registered for 33 24-h aerosol samples that had been collected at Crows Landing, between June 1988 and June 1989 (Chow et al., 1990; Coulter, 2004).**

- **18 source profiles had been selected by Chow et al. (1990) as input data for trial CMBs, in order to establish "initial source contribution estimates". Estimates of source contributions had also been calculated independently for each ambient $PM_{2.5}$ sample, with manual addition, deletion, or substitution of source profiles (Chow et al., 1990).**

- **All source contribution estimates had been determined from the EFWLS estimator, by using the USEPA/ DRI CMB 7.0 model (Chow et al., 1990).**

Source Profiles Applied in the SJVAQS (Chow et. al., 1990)

- **Because geological profiles (SOIL1, SOIL3, SOIL16, and SOIL17) were too collinear to be distinguished from one another, only a single profile of this source type had been used for SA of each daily PM2.5 sample (Chow et al., 1990). The same was also true for vegetative burning profiles (BAMAJC and STAGBC), motor vehicle exhaust profiles (MOVES2 and WHDIEC), and oil combustion profiles (SFCRUC and CHCRUC).**

Profile ID	Profile Mnemonic	Description of Source Profile
01	SOIL01	Stockton Agricultural (Peat) Soil
02	SOIL03	Fresno Paved Road Dust
15	SOIL16	Bakersfield Unpaved Road Dust (Residential)
16	SOIL17	Taft Unpaved Road Dust
17	BAMAJC	Wood smoke emissions, Bakersfield Cordwood Using Majestic Fireplace
35	STAGBC	Stockton Agricultural (Wheat) Burning
37	MOVES2	South Coast Motor Vehicle Emissions, MOVES-SS (NEAEWOB, WOT, TVMT)
29	WHDIEC	Wheeler High Station Diesel Truck Emissions
30	MOTIBC	Modesto Tire Fueled Power Plant Emissions
27	SFCRUC	Santa Fe Crude Oil Boiler Emissions
39	CHCRUC	Chevron Racetrack Crude Oil Boiler Emissions
42	SCRRFC	Stanislaus County Municipal Waste Fueled Power Plant Emissions
51	AMSUL	Ammonium Sulfate, Secondary Aerosol
54	AMNIT	Ammonium Nitrate, Secondary Aerosol
56	NANO3	Sodium Nitrate, Reacted Marine Aerosol
35	MARINE	Primary Marine Aerosol
61	LIME	Primary Construction Emissions (Limestone)
60	OC	Secondary Organic Carbon

- **Resembling the original CMB analysis, RCMB was also applied for the SA of each ambient $PM_{2.5}$ sample seperately, by using the EFWLS fitting method, the same fitting species, and the same source profiles (*N = 18*) that had been employed by Chow et al. (1990) for trial CMBs.**

- **The performance measures, which were utilized for the automatic validation of successful convergences, included the absence of any negative values among estimates for source contributions, the fit indices R^2 and χ^2_{red}, and the *T-stat* ratios.**

- **Similarly to Chow et al. (1990), *% mass* was included to the performance measures of RCMB only for those ambient $PM_{2.5}$ samples, whose mass concentrations had been measured to be above 10 $\mu g/m^3$, since lower values were within a few percent precision intervals of the $PM_{2.5}$ mass measurements.**

- **The rest of the performance measures were inspected manually after the end of each running session.**

Summarization of the output of RCMB for the PM2.5 Crows Data

Apparently, almost all the sets of input data defined a plethora of over determined linear systems, converging to solutions that meet the diagnostic criteria set by the US EPA.

Sampling date	Failures of convergence	Negative estimate(s)	Low R^2	High χ^2	High or low *%mass*	Low T-Stat value(s)	"Good Fits"
20/06/88	26604	90474	50041	94648	11	351	14
02/07/88	19299	127343	29068	84755	145	1483	50
14/07/88	9447	197485	8079	44144	59	2927	2
26/07/88	22534	131458	16667	88117	6	3025	336
07/08/88	28313	113921	32654	84124	146	2943	42
19/08/88	29145	122686	22126	86783	35	1303	65
25/08/88	23712	153009	1794	81254	92	2215	67
31/08/88	26226	108371	23204	102767	8	1422	145
06/09/88	23977	125474	23986	86260	41	2345	60
12/09/88	15393	157746	1825	83178	216	3732	53
18/10/88	27657	114845	22072	96066	23	1457	23
30/10/88	33434	92837	28560	106838	27	399	48
11/11/88	15779	121004	21458	102106	N/A	1646	150
17/11/88	2559	102699	116452	38646	N/A	1772	15
23/11/88	24057	90142	19711	127474	0	679	80
29/11/88	32508	88232	18742	122586	0	47	28
05/12/88	28378	101358	18657	112737	0	828	185
11/12/88	32806	84878	15775	127799	0	783	102
17/12/88	31989	101186	18523	109331	36	767	311
23/12/88	12138	111562	20573	115405	N/A	2440	25
29/12/88	19058	109007	18647	113666	54	1649	62
04/01/89	29828	92670	19337	119358	0	880	70
10/01/89	21724	94830	18995	124377	0	2178	39
16/01/89	32394	92731	17867	118081	0	742	328
22/01/89	35384	72244	16458	137529	0	469	59
28/01/89	31069	80933	13165	135573	0	1210	193
03/02/89	1806	126134	120163	13564	N/A	464	12
09/02/89	31621	85784	18999	123885	0	1768	86
15/02/89	33264	88230	17141	122827	0	652	29
21/02/89	30671	90817	19169	119865	0	1594	27
23/03/89	20734	111295	20587	109017	N/A	447	63
10/04/89	10355	135386	31377	81531	N/A	3295	199
10/05/89	10743	114201	64800	64427	N/A	7913	59

Best Fits of RCMB (Rank 1) for each ambient $PM_{2.5}$ sample at Crows Landing

Date	**20/06/88**	**02/07/88**	**14/07/88**	**26/07/88**	**07/08/88**	**19/08/88**	**25/08/88**	**31/08/88**	**06/09/88**	**12/09/88**	**18/10/88**
SOIL01			2.6								
SOIL03											
SOIL16		1.44		4.38	1.29	3.61	8.71	4.94	4.22		
SOIL17											
BAMAJC		1.1					4.21	1.69		3.26	
STAGBC				2.51	1.8	2.69					
MOVES2	2.61	1.94		3.38	1.64	5.71	3.61	4.34	4.98		3.48
WHDIEC			1.77								
MOTIBC											
SFCRUC				1.07				1.92			
CHCRUC											
SCRRFC	3.55	1.23	4.59	2.56		2.45		4.91	4.25	17.15	15.73
AMSUL	1.96	2.04		3.81	3.49	2.88	2.83	4.27	1.61	0.93	2.59
AMNIT											2.86
NANO3	1.45	0.91		0.91	0.73	1.34	1.03	3.93	1.11	0.86	
MARINE											
OC		1.94			1.57					6.6	
LIME											
R^2	0.88	0.97	0.84	0.97	0.93	0.97	0.98	0.98	0.95	0.93	0.93
χ^2_{red}	1.38	0.9	2.37	1.05	1.61	0.95	0.84	0.73	1.32	0.93	1.15
% mass	85	77	77	88	76	78	77	87	83	83	87
Fit Measure	**0.8168**	**0.9478**	**0.6753**	**0.9364**	**0.7721**	**0.9331**	**0.9805**	**1.0749**	**0.8459**	**0.9465**	**0.8887**

Best Fits of RCMB (Rank 1) Continued

Date	**30/10/88**	**11/11/88**	**17/11/88**	**23/11/88**	**29/11/88**	**05/12/88**	**11/12/88**	**17/12/88**	**23/12/88**	**29/12/88**	**04/01/89**
SOIL01											
SOIL03							0.67				
SOIL16	1.04	0.92				0.46		0.3			
SOIL17											
BAMAJC	3.26	2.06	0.71	0.85	1.36	2.39	4.63	2.7	0.49	0.7	0.56
STAGBC											
MOVES2	4.5	2.08		1.66	2.23	4.58	3.82	2.13	1.05	1.3	1.4
WHDIEC											
MOTIBC											
SFCRUC	0.42			0.71			0.59	0.31		0.29	0.28
CHCRUC					0.11						
SCRRFC				1.82							
AMSUL	7.35	1.01	0.83	0.83	3.43	1.51	3.58	2.35	1.01	1.56	3.57
AMNIT	6.15	1.28		6.59	16.42	13.12	28.87	9.27	3.35	4.99	12.77
NANO3		1.02	1.17	0.86	0.49			0.76	1.03	0.7	0.65
MARINE											
OC						5.7	3.14	1.68			
LIME											
R^2	0.97	0.97	0.91	0.92	0.98	0.96	0.97	0.96	0.97	0.94	0.99
χ^2_{red}	0.87	0.63	0.48	1.97	0.37	0.9	1.03	1.15	0.41	1.14	0.26
% mass	78	92	70	93	86	98	89	83	80	78	91
Fit Measure	**0.9661**	**1.1566**	**1.2329**	**0.7853**	**1.5165**	**1.0187**	**0.9431**	**0.8874**	**1.4047**	**0.8653**	**1.9333**

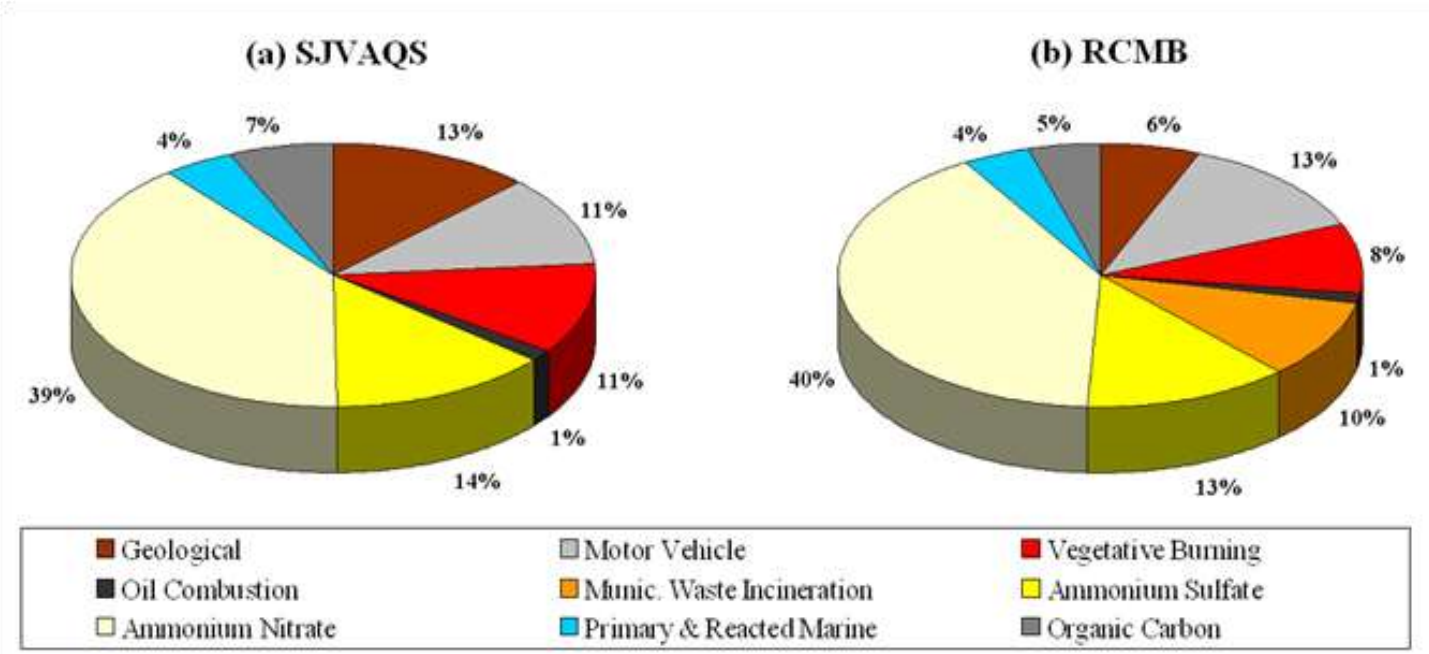
- **It was also apparent that source profiles, which had not been resolved by the original CMB analysis, due to collinearity, such as BAMAJC and STAGBC, were estimable by the preference of RCMB for the one that maximizes Fit Measure.**

Best Fits of RCMB (Rank 1) Continued 2

Date	**10/01/89**	**16/01/89**	**22/01/89**	**28/01/89**	**03/02/89**	**09/02/89**	**15/02/89**	**21/02/89**	**23/03/89**	**10/04/89**	**10/05/89**
SOIL01										0.98	0.51
SOIL03						0.77			0.46		
SOIL16			0.36	0.26							
SOIL17		0.47									
BAMAJC	1.27	2.29	2.36	2.78	0.41	1.34	2	1.42	0.57	1.02	0.41
STAGBC											
MOVES2	2.01	2.13	4.59	3.58		1.71	1.49	1.48	1.73	1.46	
WHDIEC					0.78						
MOTIBC											0.05
SFCRUC	0.44	0.75	0	0.77							
CHCRUC			0.31	0							
SCRRFC										3	
AMSUL	1.43	1.78	3.61	6.17	0.54	3.46	2.93	2.38	1.34	1.3	1.2
AMNIT	13.22	11.53	31.47	39.64	0.85	18.95	18.18	12.89	2.07		
NANO3		0.62		0.78	0.45	0.63	0.54	0.51	0.93	1.08	0.82
MARINE			0.27								0.22
OC		1.74		2.99			1.77	1.57			
LIME											
R^2	0.94	0.98	0.94	0.98	0.96	0.99	0.99	0.98	0.98	0.99	0.98
χ^2_{red}	1.12	0.53	1.59	0.58	0.21	0.22	0.23	0.44	0.43	0.23	0.28
% mass	84	93	90	89	72	91	93	105	89	90	70
Fit Measure	**0.8925**	**1.2641**	**0.8222**	**1.1981**	**2.1715**	**2.1244**	**2.1014**	**1.4068**	**1.3967**	**2.0790**	**1.7698**

- **Moreover, the temporal variation of these source contributions seemed to be reasonable, since wood smoke emissions (BAMAJC) had been expected to occur more frequently during the cold period, from fireplaces and woodstoves, while agricultural burnings (STAGBC) were common during the warm period, due to prescribed burns, set by farmers (Chow et al, 1990).**

Average source contributions to ambient PM2.5 at Crows Landing



- **In contrast with the original CMB analysis, significant contributions were estimated by RCMB for municipal waste incineration, although Chow et al (1990) reported that contributions from such a source type were not detected in any ambient sample.**

Detailed output of RCMB for the ambient $PM_{2.5}$ sample of 02/07/88

Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
SOIL01			1.35		1.47					2.10
SOIL03										
SOIL16	1.44	1.35								
SOIL17						1.28		1.39		
BAMAJC	1.10			1.24	1.40			1.16	1.46	1.63
STAGBC		0.43	0.60			0.43				
MOVES2	1.94	2.03	1.86	1.67	1.76	1.97	1.86	1.87	2.25	1.83
WHDIEC										
MOTIBC										
SFCRUC										
CHCRUC										
SCRRFC	1.23	1.50	1.74	4.38	1.44	1.87	4.88	1.55	4.33	
AMSUL	2.04	2.00	1.98	1.80	2.01	1.97	1.78	2.00	1.77	2.20
AMNIT										
NANO3	0.91	0.90	0.90	0.82	0.90	0.90	0.81	0.90	0.81	0.93
MARINE										
OC	1.94	2.19	2.10	1.67	1.79	2.18	2.06	1.92		1.79
LIME										
% mass	77	76	77	84	79	77	83	79	77	76
χ^2_{red}	0.9	0.9	1.01	1.15	1.18	1.27	1.31	1.32	1.23	1.57
R^2	0.97	0.96	0.96	0.92	0.95	0.95	0.9	0.95	0.9	0.94
Fit Measure	**0.9478**	**0.9448**	**0.9066**	**0.8797**	**0.8633**	**0.8352**	**0.8312**	**0.8301**	**0.8297**	**0.7793**

Detailed output -Continued

Rank	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
SOIL01			1.05							
SOIL03										
SOIL16		1.37								
SOIL17				1.23	2.09					
BAMAJC	1.25				1.25		1.47			
STAGBC										0.72
MOVES2	1.70	2.08	1.96	2.03	1.99	1.89	2.30		1.97	
WHDIEC								1.79		1.58
MOTIBC								0.15		
SFCRUC										
CHCRUC										
SCRRFC	4.37	1.83	2.75	2.32		4.88	4.32	5.12	5.02	5.27
AMSUL	1.76	1.98	1.91	1.94	2.19	1.75	1.74	1.72	1.77	1.75
AMNIT	0.61					0.61	0.59			
NANO3		0.90	0.87	0.89	0.93			0.82		0.82
MARINE										
OC	1.65	2.29	2.19	2.26	1.99	2.04		2.04	1.99	1.87
LIME										
% mass	83	76	78	78	76	81	76	85	78	88
χ^2_{red}	1.7	1.64	1.75	1.78	1.79	1.76	1.74	2.26	1.93	2.53
R^2	0.88	0.93	0.9	0.91	0.93	0.86	0.86	0.87	0.83	0.86
Fit Measure	**0.7663**	**0.7657**	**0.7524**	**0.7501**	**0.7499**	**0.75**	**0.73273**	**0.72193**	**0.7115**	**0.70926**

Detailed output –Continued 2

Rank	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
SOIL01		1.51								
SOIL03										
SOIL16										
SOIL17						1.43				
BAMAJC	1.61	1.42	1.41			1.17				2.04
STAGBC							0.72		0.85	
MOVES2		1.82				1.93				
WHDIEC			1.98	1.74	1.80		1.59	2.58		
MOTIBC	0.14		0.14		0.15			0.15		
SFCRUC										
CHCRUC										
SCRRFC	4.58	1.35	4.69	6.38	5.10	1.46	5.26	5.38	5.42	5.44
AMSUL	1.80	1.91	1.71	1.70	1.69	1.90	1.73	1.67	1.81	1.79
AMNIT		0.63			0.65	0.63	0.64			
NANO3	0.82		0.82	0.80				0.81	0.82	0.81
MARINE										
OC	2.28	1.76		1.93	2.04	1.89	1.87		2.58	2.00
LIME										
% mass	82	76	78	92	83	76	86	77	84	88
χ^2_{red}	2.38	2.43	2.38	2.95	2.72	2.55	2.99	2.48	2.8	3.07
R^2	0.87	0.91	0.87	0.81	0.85	0.9	0.83	0.85	0.83	0.81
Fit Measure	**0.70453**	**0.69238**	**0.69149**	**0.68659**	**0.68401**	**0.68375**	**0.67557**	**0.67371**	**0.67328**	**0.67254**

Detailed output –Continued 3

Rank	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
SOIL01		1.17	1.07							
SOIL03										
SOIL16							0.96			
SOIL17								1.26		
BAMAJC					1.62	1.83				1.41
STAGBC	0.74	0.99					0.89			
MOVES2			2.01					2.09		
WHDIEC	2.26	1.58				1.69	1.68		1.78	1.99
MOTIBC				0.15	0.14				0.15	0.14
SFCRUC										
CHCRUC										
SCRRFC	5.48	2.37	2.69	5.44	4.56	5.55	2.64	2.24	5.31	4.67
AMSUL	1.71	1.92	1.85	1.77	1.77	1.72	1.90	1.87	1.71	1.69
AMNIT			0.61		0.64			0.62		0.64
NANO3	0.82	0.90		0.81		0.81	0.90			
MARINE										
OC		2.03	2.17	2.88	2.28		2.06	2.24	2.02	
LIME										
% mass	80	80	76	81	80	85	80	75	80	77
χ^2_{red}	2.64	3.1	2.64	2.78	2.91	3.03	3.25	2.76	2.88	2.89
R^2	0.83	0.88	0.86	0.83	0.85	0.81	0.87	0.86	0.83	0.85
Fit Measure	**0.67118**	**0.66836**	**0.66594**	**0.66577**	**0.66482**	**0.66202**	**0.66159**	**0.65977**	**0.65771**	**0.65292**

Detailed output –Continued 4

Rank	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
SOIL01							1.17	1.16		1.43
SOIL03										
SOIL16									0.81	
SOIL17										
BAMAJC					1.63					
STAGBC	0.70	0.85		0.74			1.15		1.03	0.77
MOVES2										
WHDIEC	1.58		2.59	2.26				1.82		1.70
MOTIBC			0.15		0.14	0.16		0.14		
SFCRUC										
CHCRUC										
SCRRFC	5.50	5.42	5.36	5.48	4.73	5.43	2.45	2.59	3.12	1.43
AMSUL	1.74	1.78	1.65	1.69	1.79	1.75	1.97	1.88	1.93	1.85
AMNIT		0.64	0.64	0.64		0.64				0.66
NANO3							0.90	0.88	0.88	
MARINE										0.43
OC	1.85	2.57			2.25	2.88	2.74	2.25	2.77	2.16
LIME										
% mass	83	82	76	79	77	79	76	78	77	76
χ^2_{red}	3.14	3.21	2.88	3.05	3.04	3.18	3.47	3.86	3.57	3.7
R^2	0.8	0.8	0.82	0.81	0.83	0.81	0.86	0.86	0.84	0.89
Fit Measure	**0.6506**	**0.64484**	**0.64247**	**0.64161**	**0.64088**	**0.6383**	**0.63363**	**0.63349**	**0.62949**	**0.63941**

Conclusions

- **From the illustrated test case it becomes evident that a typical set of input data, gathered for CMB modeling, can often define a plethora of least squares systems, for which standard LS fitting methods converge successfully, to solutions that meet common statistical criteria.**

- **The above test case also confirms the well-established fact (Cheng and Hopke, 1986) that two different solutions, both having acceptable performance measures, can often be found for the same CMB problem by two different people.**

- **RCMB minimizes personal judgment, because it is capable of leading straight-forwardly to the best-fit combination of source profiles that can possibly be obtained by a set of input data, from a statistical point of view.**

- **Nevertheless, RCMB, like any other receptor model, is rather explanatory than predictive, thus, it should not be considered as a statistical black box.**

Thank you for your attention!

Bibliography

- **Argyropoulos G, Samara C. Development and application of a robotic chemical mass model for source apportionment of atmospheric particulate matter. Environmental Modelling and Software 2010; 26, 469-481.**

- **Cheng, M. D. and Hopke, P. K., 1986. Investigation on the Use of Chemical Mass Balance Receptor Model: Numerical Computations, Chemometrics and Intelligent Laboratory Systems 1, 33-50.**

- **Chow, J.C., Watson, J.G., Lowenthal, D.H., Pritchett, L.C., and Richards, L.W., 1990. San Joaquin Air Quality Study, Phase 2 : PM10 modelling and analysis, Volume I: Receptor Modeling Source Apportionment, Final Report, DRI Document No. 8929.1F, JPA Contract #88-1.**

- **Coulter, C.T., 2004. EPA-CMB8.2 Users Manual, Report No. EPA-452/R-04-011, U.S. Environmental Protection Agency, Research Triangle Park, N.C.**

Bibliography-Continued

- **Friedlander, S.K., 1973. Chemical element balances and identification of air pollution sources, Environmental Science and Technology 7 (3), 235–240.**

- **Watson, J.G., Cooper, J.A., Huntzicker, J.J., 1984. The effective variance weighting for least squares calculations applied to the mass balance receptor model, Atmospheric Environment 18 (7)**

- **Watson, J.G., 2004. Protocol for Applying and Validating the CMB Model for PM2.5 and VOCs, Report No. EPA-451/R-04-001, U.S. Environmental Protection Agency, Research Triangle Park, N.C.**