Using Radon-222 as an Indicator of Atmospheric Mixing Depth in ME-2 for PM$_{2.5}$ Source Apportionment

Jagoda Crawford*, Scott Chambers, David D. Cohen, Alan Griffiths, Alastair Williams, Eduard Stelcer

Australian Nuclear Science and Technology Organisation, Locked Bag 2001 Kirrawee DC NSW 2232, Australia

ABSTRACT

We isolated diurnal timescale contributions to a 6-year hourly radon record and incorporated them in ME-2 as a proxy for changes in atmospheric mixing depth in an attempt to improve the source apportionment of fine atmospheric particulate matter (PM$_{2.5}$). Results from this radon-based implementation of ME-2 are directly compared with the more traditional ME-2 implementation where wind speed is used, as a proxy for changes in mixing depth. The radon-based version more accurately reproduced daily PM$_{2.5}$ source contributions, as evidenced by better correlations with the results from the corresponding bi-linear model. The versions of ME-2 employed in this study were modified to account for calm wind conditions separately, and a recently updated solution approach was adopted.

Source apportionment for the radon-based ME-2 implementation was most successful for the finer, primary emissions (Smoke, Autos, Industry) that are more easily suspended and whose concentrations are more directly tied to changes in atmospheric mixing depth. Incorporation of the diurnal radon signal in ME-2 improved the estimated source strength distributions of the Smoke, Autos and Industry sources with respect to the township of Muswellbrook. It also resulted in a more consistent anti-correlation between these 3 source types and atmospheric mixing depth than for the wind speed case. These results confirm that near surface radon concentration is more closely tied to atmospheric mixing depth (and therefore pollutant concentrations) than wind speed.

The measurement site for this study is a small township in a rural setting, with nearby power stations and open-cut coal mines. Consequently, the distribution and characteristics of anthropogenic aerosol sources are very different than for a typical urban or industrial setting. This is reflected in lower correlation between the multi-linear models and the corresponding bi-linear models, indicating that the performance of multi-linear models is affected by the nature of the distribution of sources.

Keywords: Meteorological parameterisation; Receptor modelling; Multi-linear modelling; Rotational ambiguity.

INTRODUCTION

Fine aerosols are known to affect the Earth’s radiative balance (e.g., Charlson et al., 1992; Hubert et al., 2003; Jiang et al., 2013) and contribute to adverse health effects (e.g., Dockery et al., 1993; Moloi et al., 2002; Russell and Brunekreef, 2009). As a result, regulatory guidelines have been set in many regions of the world, and efforts made to reduce atmospheric releases of primary anthropogenic aerosols or precursors to secondary aerosols (e.g., USEPA, 2007; NSW, 2011; VES, 2012).

A first step to implementing effective release control measures is identifying the main aerosol sources and their relative contributions to the observed aerosol concentrations. When aerosol concentration and elemental composition is determined by Ion Beam Analysis techniques (e.g., Cohen, 1996), source characterisation and apportionment can be performed using multivariate receptor modelling techniques (e.g., Kim and Hopke, 2008; Lee et al., 2008; Cohen et al., 2012). Multivariate methods employed to date include: Principle Component Analysis (PCA; e.g., Jollife, 1986), Positive Matrix Factorisation (PMF; e.g., Paatero and Tapper, 1994) and UNMIX (e.g., Henry, 2002). The advantage of PMF over PCA is that non-negativity of the solution is built directly into the solution process. Two programs commonly used to solve the PMF problem are PMF-2 (Paatero and Tapper, 1994) and ME-2 (Multi-linear Engine; Paatero, 1999). Whilst in PMF-2 the factorisation model is pre-defined, ME-2 is more flexible, having been developed such that the user can specify the form of the model using a scripting language.

Factor analysis problems are renowned for rotational
ambiguity in their solutions (i.e., different solutions provide equally good fits, as measured by the optimisation method; Paatero et al., 2005; EPA, 2009; Paatero and Hopke, 2009). To reduce rotational ambiguity, multi-linear models that include other parameters (e.g., speed, wind direction, and temporal factors) have been introduced in ME-2 (Paatero and Hopke, 2002; Kim et al., 2003; Buset et al., 2006). Crawford et al. (2013) introduced hourly Radon-222 (radon) concentrations observed at Richmond, NSW, to a multi-linear model as a combined proxy for (i) the degree of terrestrial influence on an air mass (an air mass fetch effect), and (ii) the degree of dilution in the atmosphere due to the diurnal evolution of the atmospheric boundary layer (ABL). Since fetch effects are already accounted for, at least in part, by the wind direction parameter, we sought a radon-based parameter that eliminated that redundancy.

Chambers et al. (2014b) describe a technique by which a long-term radon time series can be decomposed into its diurnal and longer timescale components. Contributions to radon time series on greater than diurnal timescales are primarily associated with fetch effects and the long-term (2–3 week) air mass history. Contributions on diurnal timescales, however, are primarily associated with changes in atmospheric mixing depth (a term often used synonymously with atmospheric stability) and the strength of the local radon source function.

In this study we compare multi-linear models where wind speed and the diurnal radon component have been used separately as parameters representing diurnal changes in atmospheric mixing depth. A further development to a recent similar study (Crawford et al., 2013) is that the implementation of the multi-linear models has been modified to include calm wind conditions and a refined solution process is used (i.e., a model and a solution process distributed with a more recent version of the ME-2 program is adopted; Paatero 2012, personal communication). The performance of the refined multi-linear model implementation is evaluated for Muswellbrook observations, where large point sources are located in the immediate vicinity. The measurement site for this study consists of a small township in a rural setting, with nearby power stations and open-cut coal mines. Some comparisons with results from a prior study in the western Sydney region (Crawford et al., 2013) are also made.

METHOD AND DATA

Measurement Site and Study Domain

The measurement site (32°16′S, 150°53′E; 144 m above sea level, asl), henceforth referred to as Muswellbrook (Fig. 1), is situated 130 km from the coast in a broad primarily-rural valley that runs WNW-SE. Located near the north-western boundary of the township of Muswellbrook, local urban emissions are prominent for wind directions from north-east through south. In addition to local emissions, three coal fired power stations are located south-east of Muswellbrook and open-cut coal mines are located in close proximity, mainly to the south, such that the site is classified as rural/industrial.

Wind directions at Muswellbrook are strongly influenced by its valley location; with predominantly west to north westerly winds in winter, and south easterly winds in summer. A mixture of wind directions is commonly observed throughout the other seasons.

Aerosol Sampling and Elemental Analysis

Aerosol sampling in Muswellbrook commenced in 1998 and is ongoing. Samples are collected twice a week (Wednesday and Sunday), using an IMPROVE PM$_{2.5}$
cyclone system, by passing air through a 25 mm diameter stretched Teflon filter for 24-hours (midnight-to-midnight) at a flow rate of ~22 L/min (Cohen et al., 1996). Only observations between January 2007 and December 2012 were used for this study, corresponding to the period for which collocated atmospheric radon measurements were also available.

Elemental analysis of each aerosol sample was performed by accelerometer-based ion beam analysis (IBA; Cohen et al., 1996; Cohen et al., 2004), yielding concentrations for 20 elements: H, Na, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Br, Pb; black carbon (BC) concentrations were determined using laser integrating plate method (LIPM) assuming a mass absorption coefficient of 7 m²/g (Taha et al., 2007).

Atmospheric Radon Measurements

The ubiquitous naturally occurring radioactive noble gas radon has a well-constrained terrestrial source (16–26 mBq/m²/s; Schery and Wasiolek, 1998; Goto et al., 2008) that exhibits little spatial and temporal variability over unsaturated, unfrozen surfaces (e.g., Jacob et al., 1997). The sole atmospheric sink of radon is radioactive decay. The half-life of radon (3.8 days) is much greater than turbulent timescales in the atmosphere (~1 hour), is sufficiently long for radon to be considered a conservative tracer over the course of a single night, yet short enough to prevent radon from accumulating in the atmosphere on greater than monthly timescales. These characteristics make radon an ideal tracer for atmospheric mixing or transport studies (e.g., Balkanski et al., 2013). Continuous, direct, hourly radon measurements were assumed for the purpose of this investigation.

The atmospheric radon monitoring program at Muswellbrook was operational between late 2006 and late 2013. Continuous, direct, hourly radon measurements were made using a 1500 L dual flow loop, two filter radon detector (e.g., Whittlestone and Zahorowski, 1998; Chambers et al., 2014a). Sampling was conducted from 2 m agl, at a flow rate of approximately 60 L/min, through a screened “goose-neck” style intake to minimise the ingestion of insects and precipitation. A 400 L delay volume was included as (Paatero, 2010):

\[ X = GF + E \]

where \( X \) is a matrix of measured elements, \( G \) and \( F \) are factor matrices to be determined, and \( E \) is a matrix of residuals. The lower limit of determination, defined here as the radon concentration at which the detector’s measurement uncertainty reaches 30%, was 40 mBq/m³.

The Muswellbrook radon record can be considered to constitute a superposition of signals on hourly, diurnal, synoptic and seasonal timescales. Contributions to the variability on diurnal timescales are closely related to changes in the local atmospheric mixing depth. Contributions on larger timescales, however, are primarily related to wind direction (changes in long-term air mass fetch). Since wind direction is already being employed as an independent parameter of ME-2, before incorporating radon as a parameter we isolated diurnal from longer-timescale contributions to the Muswellbrook radon record using a method outlined by Chambers et al. (2014a). While it is expected that the diurnal radon signal will be superior to wind speed as a parameter to indicate changes in atmospheric mixing depth (closely related to concentrations of fine aerosols and gaseous precursors), it is expected to be poorly related to the mobilisation of heavier aerosols (e.g., soil and sea-spray).

Wind Speed and Direction Estimates

Measurements of wind speed and direction were not available at the site, thus a regional model was used to downscale a global reanalysis to obtain estimates of wind speed at 100 m agl. For this purpose the Advanced Research WRF (WRF hereafter) model version 3.5.1 (Skamarock et al., 2005; Skamarock and Klemp, 2008) was run with boundary-conditions from the half-degree Climate Forecast System (CFS) and CFSv2 (Saha et al., 2010, 2013) with a change-over in 2011. Our configuration is similar to that used by Angevine et al. (2013) and verification against surface observations confirmed that the WRF model was an improvement on using CFS or ERA-interim directly.

Validation of the WRF wind fields (to observations) was limited to ensuring that the regional model provides improvements in the root mean squared error (rmse) and bias when compared with the CFSR or the ERA-Interim (Dee et al., 2011) reanalyses. The improvement in error scores, based on observations within the model domain during the year 2013 is shown in Table 1. For the 15 Australian Bureau of Meteorology (BOM) meteorological observation sites, the wind speeds from the WRF simulations had the smallest rmse, although at night a positive bias was seen. The BOM stations were chosen to be well within the WRF domain and close to the coast.

Bi-Linear Factor Analysis Model (PMF)

The standard (bi-linear) factor analysis problem is specified as (Paatero, 2010):

\[ X = GF + E \]

where \( X \) is a matrix of measured elements, \( G \) and \( F \) are factor matrices to be determined, and \( E \) is a matrix of residuals. The lower limit of determination, defined here as the radon concentration at which the detector’s measurement uncertainty reaches 30%, was 40 mBq/m³.

The Muswellbrook radon record can be considered to constitute a superposition of signals on hourly, diurnal, synoptic and seasonal timescales. Contributions to the variability on diurnal timescales are closely related to changes in the local atmospheric mixing depth. Contributions on larger timescales, however, are primarily related to wind direction (changes in long-term air mass fetch). Since wind direction is already being employed as an independent parameter of ME-2, before incorporating radon as a parameter we isolated diurnal from longer-timescale contributions to the Muswellbrook radon record using a method outlined by Chambers et al. (2014a). While it is expected that the diurnal radon signal will be superior to wind speed as a parameter to indicate changes in atmospheric mixing depth (closely related to concentrations of fine aerosols and gaseous precursors), it is expected to be poorly related to the mobilisation of heavier aerosols (e.g., soil and sea-spray).

Wind Speed and Direction Estimates

Measurements of wind speed and direction were not available at the site, thus a regional model was used to downscale a global reanalysis to obtain estimates of wind speed at 100 m agl. For this purpose the Advanced Research WRF (WRF hereafter) model version 3.5.1 (Skamarock et al., 2005; Skamarock and Klemp, 2008) was run with boundary-conditions from the half-degree Climate Forecast System (CFS) and CFSv2 (Saha et al., 2010, 2013) with a change-over in 2011. Our configuration is similar to that used by Angevine et al. (2013) and verification against surface observations confirmed that the WRF model was an improvement on using CFS or ERA-interim directly.

Validation of the WRF wind fields (to observations) was limited to ensuring that the regional model provides improvements in the root mean squared error (rmse) and bias when compared with the CFSR or the ERA-Interim (Dee et al., 2011) reanalyses. The improvement in error scores, based on observations within the model domain during the year 2013 is shown in Table 1. For the 15 Australian Bureau of Meteorology (BOM) meteorological observation sites, the wind speeds from the WRF simulations had the smallest rmse, although at night a positive bias was seen. The BOM stations were chosen to be well within the WRF domain and close to the coast.

Bi-Linear Factor Analysis Model (PMF)

The standard (bi-linear) factor analysis problem is specified as (Paatero, 2010):

\[ X = GF + E \]

where \( X \) is a matrix of measured elements, \( G \) and \( F \) are factor matrices to be determined, and \( E \) is a matrix of residuals. The lower limit of determination, defined here as the radon concentration at which the detector’s measurement uncertainty reaches 30%, was 40 mBq/m³.

The Muswellbrook radon record can be considered to constitute a superposition of signals on hourly, diurnal, synoptic and seasonal timescales. Contributions to the variability on diurnal timescales are closely related to changes in the local atmospheric mixing depth. Contributions on larger timescales, however, are primarily related to wind direction (changes in long-term air mass fetch). Since wind direction is already being employed as an independent parameter of ME-2, before incorporating radon as a parameter

\[ X = GF + E \]

where \( X \) is a matrix of measured elements, \( G \) and \( F \) are factor matrices to be determined, and \( E \) is a matrix of residuals. The lower limit of determination, defined here as the radon concentration at which the detector’s measurement uncertainty reaches 30%, was 40 mBq/m³.

The Muswellbrook radon record can be considered to constitute a superposition of signals on hourly, diurnal, synoptic and seasonal timescales. Contributions to the variability on diurnal timescales are closely related to changes in the local atmospheric mixing depth. Contributions on larger timescales, however, are primarily related to wind direction (changes in long-term air mass fetch). Since wind direction is already being employed as an independent parameter of ME-2, before incorporating radon as a parameter

\[ X = GF + E \]
represents the concentration of element \( j \) in the \( i \)th sample, \( G \) is an \( n \) by \( p \) matrix of source contributions (reflecting temporal variation in source strengths), \( F \) is a \( p \) by \( m \) matrix of source fingerprints.

While PMF-2 could be used to solve Eq. (1), we chose to use ME-2 here because of its versatility. Based on similar previous studies in the Sydney region (Cohen et al., 2011, 2012; Crawford et al., 2013), solutions for six, seven, and eight factors were examined. In order to obtain a global minimum 50 alternate solutions were obtained for each case; those with the lowest Q values were further examined for identifiable source fingerprints. The \( G \) matrix was examined for correlation between columns (an indication that one source has possibly been split into two factors, and the existence of rotational freedom). The residuals approximated a normal distribution in all cases.

Based on this analysis seven fingerprints were selected as the best solution. For this case (\( n = 449, m = 22 \) and \( p = 7 \)), the expected optimum \( Q \) would be 6581 (the degrees of freedom in the optimisation problem). The final \( Q \) was 5277 and there was a low correlation between columns of the \( G \) matrix (Supplementary Fig. 1).

### Multi-Linear Factor Analysis Model (ME-2)

One way of reducing the rotational ambiguity in PMF problems is to introduce multi-linear equations that include other parameters to constrain the problem, which can then be solved using ME-2. Parameters previously incorporated into ME-2 include: wind speed, wind direction, and temporal factors, such as day of week and season (Paatero and Hopke, 2002; Kim et al., 2003; Buset et al., 2006). Introducing such parameters also provides additional information, e.g., source factor dependence on the selected parameters.

As previously performed in Crawford et al. (2013), here we compare the performance of two multi-linear models: in the first case wind speed is included as a parameter, whereas in the second we include the diurnal radon variability. Unlike for Crawford et al. (2013), however, in this study the form of multi-linear model employed is similar to Kim et al. (2003), where calm conditions are treated separately.

The solution process was as follows: the contributing sector for \( PM_{2.5} \) was determined for each hour of the sampling days from the wind direction (see Table 3). The \( PM_{2.5} \) concentration for a given hour was then estimated by adjusting the contribution from that sector by a factor based on wind speed (case 1) or diurnal radon concentration (case 2). Lastly, temporal factors were applied: time of year (to account for seasonal variations in the source term), day of week (to account for varying emissions over the working days and the weekend), and time of day.

The two variations of the multi-linear models are:

**Case 1: The wind speed based model**

\[
x_{i,j} = \sum_{k=1}^{p} M \left( m_i, k \right) D \left( d_i, k \right)
\left( \sum_{h=1}^{24} T \left( t_{i,h}, k \right) U \left( i, k, h \right) \right) f_{i,h} + \epsilon_{i,j}
\]  

**Case 2: The radon based model**

\[
x_{i,j} = \sum_{k=1}^{p} M \left( m_i, k \right) D \left( d_i, k \right)
\left( \sum_{h=1}^{24} T \left( t_{i,h}, k \right) V \left( i, k, h \right) \right) f_{i,h} + \epsilon_{i,j}
\]

\[
U(i, k, h) = S(\theta_{i,k}, h) W(s_{i,k}, h) \text{ if wind speed for hour } h \text{ of day } i > 1 \text{ m/s}
\]

\[
= C(k) \text{ otherwise; i.e., calm conditions.}
\]

\[
V(i, k, h) = S(\theta_{i,k}, h) R(r_{i,k}, h) \text{ if wind speed for hour } h \text{ of day } i > 1 \text{ m/s}
\]

\[
= C(k) \text{ otherwise; i.e., calm conditions.}
\]

where \( S(\theta_{i,k}, h) \) is an element of matrix \( S \), being the contribution to source \( k \) from sector \( \theta_{i,k} \) (i.e., the sector for hour \( h \) of day \( i \)). Matrix \( D \) corresponds to an adjustment based on whether the sampling was conducted on a weekday or a weekend. \( W(s_{i,k}, h) \) is an adjustment for each factor \( (k = 1,...,p) \) based on wind speed for hour \( h \) of sample \( i \) (for the wind speed based model). Matrix \( R \) corresponds to an adjustment of the concentration (matrix \( S \)) by the diurnal component of radon concentration (for the radon based model). Matrix \( C \) represents the adjustment to each source factor during calm conditions (i.e., calm conditions are not associated with a wind direction sector). The matrix \( T \) relates to the \( PM_{2.5} \) concentration adjustment based on the time of day, while the matrix \( M \) allows for seasonal variations.

Elements of matrices \( M, D, T, S, C, W, R \) and \( \theta \) are to be resolved in the factorisation process. The definitions of the indices are presented in Tables 2 and 3.

Since Eqs. (2) and (3) are ill-posed and multiple solutions are possible, the elements of matrices \( T, M, W, R \) and \( D \) were regularised using auxiliary equations, as detailed in Crawford et al. (2013).

From Eqs. (2) and (3) matrices \( G' \) and \( G'' \) can be calculated which estimate the temporal source contributions from the multi-linear models, which can be compared directly with the \( G \) matrix from the corresponding bi-linear model:

\[
G'_{i,k} = M \left( m_i, k \right) D \left( d_i, k \right) \left( \sum_{h=1}^{24} T \left( t_{i,h}, k \right) U \left( i, k, h \right) \right)
\]  

\[
G''_{i,k} = M \left( m_i, k \right) D \left( d_i, k \right) \left( \sum_{h=1}^{24} T \left( t_{i,h}, k \right) V \left( i, k, h \right) \right)
\]

To solve the models we started with the ME-2 scripts called EnhPMF-200601714.ini, as supplied by Pentti Paatero (personal communication; 2012), and then made the changes outlined above. In Crawford et al. (2013), the bi-linear model and each of the multi-linear models were solved simultaneously but for the multi-linear models the errors between the measured matrix \( X \) and that estimated by the multi-linear models were minimised. In the EnhPMF-200601714.ini implementation, the bi-linear model and multi-linear model would be solved simultaneously using the ME-2 scripts.
Table 2. Notation for index terms in Eqs. (2) and (3). The corresponding ranges of values for each parameter are given in Table 3.

<table>
<thead>
<tr>
<th>Index</th>
<th>Meaning</th>
<th>Range of Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_i$</td>
<td>Time of year for sample $i$</td>
<td>1–6</td>
</tr>
<tr>
<td>$\theta_{i,h}$</td>
<td>Sector for hour $h$ of sample $i$</td>
<td>1–18</td>
</tr>
<tr>
<td>$s_{i,h}$</td>
<td>Wind speed for hour $h$ of sample $i$</td>
<td>1–5</td>
</tr>
<tr>
<td>$r_{i,h}$</td>
<td>Radon stability signal for hour $h$ of sample $i$</td>
<td>1–5</td>
</tr>
<tr>
<td>$t_{i,h}$</td>
<td>Time of day</td>
<td>1–6</td>
</tr>
<tr>
<td>$d_i$</td>
<td>Day of week</td>
<td>1–weekday</td>
</tr>
</tbody>
</table>

Table 3. Definition of indices for Eqs. (2) and (3). The wind information used was based on WRF simulations at 100 m above ground level.

<table>
<thead>
<tr>
<th>Index</th>
<th>Wind Direction ($\theta_i$) (°)</th>
<th>Wind Speed ($s_{i,h}$) (m/s)</th>
<th>Time of Year ($m_i$)</th>
<th>Radon ($r_{i,h}$) (mBq/m³)</th>
<th>Time of day ($t_{i,h}$) (hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0–20</td>
<td>1–2.5</td>
<td>Dec–Jan</td>
<td>&lt; 1000</td>
<td>3–Jun</td>
</tr>
<tr>
<td>2</td>
<td>20–40</td>
<td>2.5–4.0</td>
<td>Feb–Mar</td>
<td>1000–2500</td>
<td>7–Sep</td>
</tr>
<tr>
<td>3</td>
<td>40–60</td>
<td>4.0–6.0</td>
<td>Apr–May</td>
<td>2500–11000</td>
<td>Oct–13</td>
</tr>
<tr>
<td>4</td>
<td>60–80</td>
<td>6.0–8.0</td>
<td>Jun–Jul</td>
<td>11000–30000</td>
<td>14–16</td>
</tr>
<tr>
<td>5</td>
<td>80–100</td>
<td>&gt; 8.0</td>
<td>Aug–Sep</td>
<td>&gt; 30000</td>
<td>17–20</td>
</tr>
<tr>
<td>6</td>
<td>100–120</td>
<td></td>
<td>Oct–Nov</td>
<td></td>
<td>21–02</td>
</tr>
<tr>
<td>7</td>
<td>120–140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>18</td>
<td>340–360</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear models are also solved simultaneously, but errors between the elements of matrices $G$ and $G'$ (or $G''$ for the radon model) are minimised:

$$Q'' = \sum_{i=1}^{n} \sum_{k=1}^{c} \frac{(G(i,k) - G'(i,k))^2}{\sigma_{i,k}^2}$$ (6)

where $\sigma_{i,k}$ is calculated according to a predefined ME-2 model error code (EM code) of –5 (Paatero, 2000). In this case:

$$\sigma_{i,k} = c_1 + c_3 \left| G(i,k) - G'(i,k) \right|$$ (7)

where $c_1$ (= 1.4) and $c_3$ (= 0.35) were varied during the optimisation process and were chosen such that the multi-linear model did not unduly affect the fitting process and the fingerprints were still identifiable.

**Conditional Probability Function**

The Conditional Probability Function (CPF) is used to estimate the probability that the source contribution from a given wind direction sector will exceed a pre-determined threshold. The CPF was defined as (notation adapted from Kim and Hopke, 2004):

$$CPF_{\Delta \theta} = \frac{m_{\Delta \theta}}{n_{\Delta \theta}}$$ (8)

where $n_{\Delta \theta}$ represents the total number of wind observations from angular sector $\Delta \theta$, and $m_{\Delta \theta}$ the number of times the measured concentration exceeded a predefined threshold within sector $\Delta \theta$. In this study, 18 sectors were used ($\Delta \theta = 20^\circ$) and the thresholds were the 50th percentiles of the recovered source fingerprint concentrations. The CPF was applied to the source contributions as identified by the bi-linear model for comparison with the results of matrix $S$ from the multi-linear models (section 2.6).

**RESULTS AND DISCUSSION**

Muswellbrook aerosol and radon data from January 2007 to December 2012 was used in this study. The southern hemisphere seasonal convention has been adopted, and all reported times are in Eastern Standard Time (UTC +10 h). Extreme soil dust events occurring in September 2009 were removed from the data set before the PMF analysis was carried out.

**Source Fingerprints**

Source fingerprints resolved by the bi-linear model and two multi-linear models (wind speed and radon) are presented in Fig. 2, the percentage contribution is presented in Table 4, and the primary fingerprint characteristics summarised below:

- **Smoke**: Dominated by H (from organic sources), K and BC (from biomass burning; Yli-Tuomi et al., 2003; Lee et al., 2008; Cohen et al., 2010), with traces of soil. Traces of Cl were present in the solutions to the bi-linear and radon based multi-linear model.

- **IndSaged**: Dominated by Na, S and BC. The Na to S ratio was 2.7 to 2.1 which is indicative of Na$_2$SO$_4$ (i.e., Na/S ratio of 1.44) most likely formed by the chemical reaction of sea spray with sulfate (Qin et al., 1997; Wu et al., 2009).
Fig. 2. The seven source fingerprints from the standard bi-linear model, and the two multi-linear models (wind speed and radon).

Table 4. The percentage contribution to the total PMF estimated mass from each source for Jan 2007 to Dec 2012, and regression results ($r^2$ values) between the ME-2 models (for wind speed and radon) and their corresponding bi-linear model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Bi-linear % of total PM$_{2.5}$ mass</th>
<th>Wind speed</th>
<th>Radon</th>
<th>Wind speed</th>
<th>Radon</th>
<th>$r^2$</th>
<th>Wind speed</th>
<th>Radon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke</td>
<td>23.42</td>
<td>22.7</td>
<td>22.16</td>
<td>0.32</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autos</td>
<td>19.87</td>
<td>20.08</td>
<td>22.2</td>
<td>0.46</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2ndryS</td>
<td>25.56</td>
<td>26.27</td>
<td>25.28</td>
<td>0.22</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea</td>
<td>3.82</td>
<td>3.75</td>
<td>3.76</td>
<td>0.16</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IndSaged</td>
<td>13.23</td>
<td>13.38</td>
<td>13.81</td>
<td>0.32</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>7.24</td>
<td>7.22</td>
<td>7.08</td>
<td>0.03</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>6.86</td>
<td>6.6</td>
<td>5.71</td>
<td>0.27</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Autos:* Dominated by H and BC, as well as other trace elements associated with motor vehicles (e.g., Zn from tyre wear, Cu from breaks). The solution from the bi-linear model contained traces of S, which was not present in the multi-linear models. In the multi-linear models S had moved to the secondary sulfate fingerprint (2ndryS).

*Soil:* Dominated by Al, Si, Ca, Ti and Fe, the five key elements commonly associated with windblown soils (Wu et al., 2009; Amato and Hopke, 2012). The [Al/Si] ratio was 0.29 which was in the range common for alumina-silicates (0.25–0.35; Cohen et al., 2010).

*2ndryS:* Dominated by S and H, indicative of secondary
sulfate. The [S/H] ratio was ~4.5 from all three models, indicating the majority of this aerosol was fully neutralised ammonium sulfate \((\text{NH}_4)_2\text{SO}_4\). The bi-linear model had traces of Ca, which were not present in the solution of the multi-linear models.

**Industry:** Dominated by H and Fe, with traces of Mn, Ni and Cu, indicating metal industry.

**Sea:** Dominated by Na and Cl, with small amounts of Br (indicative of sea spray; e.g., Qin et al., 1997; Wu et al., 2009). The [Cl/Na] ratio was 4.50, 3.96 and 3.44 (for the bi-linear, wind-speed and radon based multi-linear models, respectively), which was higher than expected for fresh sea salt (1.51 and 1.8 in sea water) indicating the presence of other Cl sources, or the removal of Na in transit. However, the multi-linear models did produce the smaller ratios.

Error bars on the fingerprints were determined using a bootstrapping procedure. The mean and 1.96 times the standard deviation for elemental composition in each fingerprint, from the bootstrapping procedure, are presented in the Supplementary Fig. 4.

**G-Space Plotting**

One way of examining the rotational freedom of a PMF solution is to observe the degree of correlation between columns of the G matrix (i.e., “G-space plotting”; Paatero et al., 2005). Scatter plots between columns of the G, G' and G" matrices from the bi-linear and multi-linear cases, respectively, are supplied in Supplementary Figs. 1–3. G-space plots were similar for all models, indicating no significant improvement over the bi-linear model by either of the multi-linear models. However, as also observed by Crawford et al. (2013), a weak relationship was seen between Autos and Smoke for all three models. A relationship was also observed between Autos and Industry, where low Industry loadings indicated lower Autos loadings. This could well be due to both industry and the suburbs being located to the east of the sampling site.

**Evaluation of the Multi-Linear Model**

To assess the ability of each multi-linear model to reproduce daily \(\text{PM}_2.5\) source contributions a linear least-squares regression was performed between columns of the G matrices of the multi-linear model and their corresponding bi-linear model (source contributions: G' vs. G and G" vs. G). While the resulting correlations were less than those observed by Crawford et al. (2013) at Richmond, better \(r^2\) values were produced for almost all source types when using the radon multi-linear model (with the exception of 2ndryS; see Table 4).

The best results were obtained for the Autos source, where 46% of the variation could be explained by the wind speed model but 57% by the radon model. This performance was closely matched by Smoke and IndSaged. Soil showed the lowest correlation; much lower than the corresponding value for Richmond (0.23 for the wind-speed model and 0.28 for the radon model). On inspection, this result was mainly affected by a small number of high soil days, for which both the wind-speed and radon models underestimated the contribution. These high Soil days might well be due to the close proximity of the open-cut coal mines and the local agricultural region at the Muswellbrook site, as the difference at Richmond was smaller where regional soil dust sources have an influence. Back trajectory analysis showed that the desert regions also had an influence at Muswellbrook.

The largest differences in percentage contributions estimated by the bi-linear and radon based multi-linear model were for Autos, Smoke and Industry, the same fingerprints for which the G-space analysis showed the most significant correlations.

**Estimating the \(\text{PM}_2.5\) Source Strength from Each Sector**

The local meteorology is strongly affected by the Hunter Valley, which channels winds from the west in winter and the east and south-east in summer. In Fig. 3 we compare the source strength for each fingerprint by sector between the three approaches: (i) using the wind speed model (matrix S, Eq. (2)); (ii) using the radon model (matrix S, Eq. (3)); and (iii) the CPF from the bi-linear model. Overall there was a good agreement between the three models.

The radon-based approach predicts slightly higher source strengths of Smoke, Autos and Industry when the wind direction is from the east corresponding to the urban areas.

The largest sources of \(\text{SO}_2\) are the three coal-fired power stations to the south-east of the site. While this is consistent with the indicated source strength for aged sulphur of industrial origin \((\text{IndSaged})\), higher concentrations of 2ndryS occur from the north-east and east sector indicating perhaps local sources of \(\text{SO}_2\) such as vehicles, contribute to 2ndryS.

Fresh sea salt (Sea) has high source strength in the south-east corresponding to the summer onshore winds. Oceanic salt reacts with sulphurous emissions from populated regions en route to the measurement site, contributing to the IndSaged signal.

Larger sources of Soil are seen in the north and to the south-west sectors in the direction of the closer located open-cut coal mines.

**Seasonality of Fingerprints and Source Factors**

Monthly distributions of each bi-linear source fingerprint are presented in Fig. 4, with the time of year factor matrices presented in Fig. 5.

A clear seasonal trend is seen in almost all cases, with Smoke, Autos and Industry showing the highest concentrations in winter and 2ndryS and IndSaged showing the highest concentrations in summer. Smoke shows the highest concentration in winter corresponding to a time of year when fire wood is used for domestic heating. 2ndryS shows the highest concentration in summer corresponding to more suitable conditions for the conversion of \(\text{SO}_2\) to secondary sulphate (Kim et al., 2003 and references therein).

No clear seasonal trend was observed for the fresh sea salt source, Sea, which could be due the inland location of Muswellbrook and the possible local sources of Cl and/or the depletion of Na in transit.

The seasonal trends seen in the bilinear source fingerprints (Fig. 5) are reflected closely in the time of year factors (Fig. 6), with only very small differences between the wind speed and radon based versions of the multi-linear model.
The exception was in the Industry category, for which the radon-based model showed a weaker seasonal variation compared with the wind speed model.

**Comparison of Wind Speed and Radon Factors**

Factors for wind speed (matrix W) and radon (matrix R) are compared in Fig. 6; a key is provided on the figure to indicate the parameter ranges for each factor.

In the case of the fine particle primary emissions (e.g., *Smoke, Autos and Industry*), there is a general tendency for the aerosol concentration to decrease with increasing wind speed. This is expected, as increasing wind speeds tend to increase the depth of the atmospheric mixing volume into which the pollutants are being emitted and also lead to more dilution due to horizontal transport and dispersion. The sense of the trend for the radon factors is opposite, since higher radon concentrations indicate shallow atmospheric mixing depths and less horizontal dilution. It is clear from
these plots that the factors for these three sources change more strongly and consistently with radon indices 1–4 than they do with the wind speed index. In the case of the highest radon concentrations (index 5: 30–80 Bq/m$^3$), the three factors all decrease somewhat. This index corresponds to nights when the atmosphere was extremely thermally stable, for which katabatic drainage flows are known to develop at the head of the Hunter Valley. Such flows would tend to flush some of the locally-generated emissions slowly down the valley. Since the radon source is much more evenly distributed than the aerosol sources, the relationship between radon and fine primary aerosol pollutants is observed to change; this would not be the case at a flat inland site. These results confirm that near surface radon concentration is more closely tied to atmospheric mixing depth (and therefore pollutant concentrations) than wind speed.

In the case of 2ndryS, both wind speed and radon again indicate an overall tendency for increased concentrations with decreasing mixing depth (i.e., reducing wind speeds or increasing radon concentration), but the relationship is not as clear or pronounced as was the case for the primary fine particles.
The tendency for IndSaged to increase for the highest wind speeds, and to reduce with increasing radon concentrations, is consistent with IndSaged being derived from remote sources, i.e., old sea air arriving inland. High radon concentrations develop under stable (inversion) conditions, and low-level inversions act to isolate local observations from the residual layer and lower troposphere, whose concentrations are primarily contributed to by distant sources.

As previously mentioned, some extreme Soil events at Muswellbrook (likely associated with high winds and large-scale dust storms) were removed from this dataset prior to analysis. For the remaining soil events, both wind speed and radon models show increasing Soil contributions with shallower atmospheric mixing depths (and again the radon model provides the stronger signal). This indicates that these Soil events were typically of a fairly local origin (e.g., agricultural or from the open cut mines). The anti-correlation with wind speed may also imply that the soil dust observed at this site is typically mobilised by mechanical means other than wind (e.g., tilling of soil, or mining activity), and remains suspended long enough to be concentrated by changes in mixing depth.

For the highest wind speeds (deepest mixing depths), both the wind speed and radon models show an increased contribution to the Sea factor, as expected given the distance (>100 km) air masses need to travel from the ocean to get to the site. This is consistent with the findings of Wang et al. (2004), and Huang et al. (2009).

Calm Conditions

Under calm conditions, air can stagnate at the head of the Hunter Valley. Higher values were recovered for Smoke, Autos, 2ndryS, Soil and Industry (Table 5), indicating the respective sources are located in fairly close proximity to the measurement site. For Sea and IndSaged, on the other hand, lower values were recovered, consistent with more distant sources and the need for higher wind speed to mobilise and transport the sea spray.

Day-of-Week Factors

The measurement time within a week was categorised into two groups: weekday or weekend. The weekday contribution to each of the source factors was set to 1, and then the relative contribution for the weekend is given in Table 5.

The observed weekend increase in Smoke indicates an increase in domestic heating or other burning activities (e.g., incinerators, clearing) over the weekend. An increased the Sea factor together with a reduction of IndSaged on the weekend is consistent with lower reactions of NaCl with the lower weekend anthropogenic air pollution over the urban region (the Industry and Autos contribution is less over the weekend). Unlike the results for Richmond, a decrease in Soil is seen for weekends, which is likely related to a reduction in coal mine operations and local farming.
Fig. 6. Wind speed factors (matrix $W$) for the wind speed model and radon factor (matrix $R$) for the radon model.

Table 5. Calm conditions factor followed by the relative contribution to each source factor during the weekend, i.e., normalised with the weekday contribution.

<table>
<thead>
<tr>
<th>Source</th>
<th>Smoke</th>
<th>Autos</th>
<th>2ndryS</th>
<th>Sea</th>
<th>IndSaged</th>
<th>Soil</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calm</td>
<td>1.661</td>
<td>1.361</td>
<td>1.538</td>
<td>0.443</td>
<td>0.644</td>
<td>1.566</td>
<td>1.211</td>
</tr>
<tr>
<td></td>
<td>1.279</td>
<td>1.122</td>
<td>1.528</td>
<td>0.592</td>
<td>0.702</td>
<td>1.35</td>
<td>0.816</td>
</tr>
<tr>
<td>Weekend</td>
<td>1.378</td>
<td>0.842</td>
<td>1.069</td>
<td>1.244</td>
<td>0.85</td>
<td>0.733</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td>1.381</td>
<td>0.878</td>
<td>1.083</td>
<td>1.187</td>
<td>0.85</td>
<td>0.728</td>
<td>0.739</td>
</tr>
</tbody>
</table>

**Time of Day Factors**

While the PM$_{2.5}$ measurements are daily integrated values, a time of day factor matrix is determined from the model. The time of day factor showed only small variations between source types (Fig. 7) – with maximum values in the morning (3–9 am) and minimum values in the afternoon, which most likely reflects the change in concentrations due to the boundary layer evolution and not due to the source term variation with the time of day. In many cases the morning concentration increase was better reflected by the radon model, indicating a closer correspondence to diurnal changes in mixing depth than for wind speed. Kim et al. (2003) found little diurnal variations in their fingerprints (other than nitrate-rich secondary aerosols) which they attributed to high source strengths during the day and high concentrations at night due to temperature inversion.

**CONCLUSION**

We compared the performance between two multi-linear models, one using wind speed and the other using diurnal radon variability as parameters representative of atmospheric mixing depth. The radon-based version better reproduced daily PM$_{2.5}$ source contributions, as evidenced by results that were better correlated with the corresponding bi-linear model, although overall correlations at this site (Muswellbrook, NSW), were lower than reported in a similar study by Crawford et al. (2013) at Richmond, NSW. This difference is most likely attributable to contrasting source characteristics and distributions between the sites. The best correlations for the radon-based method were observed for the Smoke, Autos and Industry sources, for which related emissions have fine particle sizes that are more easily
suspended in the atmosphere, and whose concentrations scale more directly with changes in atmospheric mixing depth. Correlations for larger aerosols (e.g., Sea, Soil) and secondary products (e.g., 2ndryS) were lower. While the incorporation of calm conditions into this ME-2 application provided additional information regarding the influences of regional sources, employing a modified solution approach did not have a measurable impact on the final solutions.

While for Richmond some rotational ambiguity was removed between the Autos and Smoke fingerprints, for Muswellbrook there was little improvement by introducing the multi-linear models. However, the radon based model, once again, performed better than the wind speed based model, and the rotational ambiguity between source factors was reflected in the variation of the estimated percentage contribution between the bi-linear model and radon based multi-linear model.

Incorporation of the diurnal radon signal as a parameter in ME-2 improved the estimated source strength distributions of the Smoke, Autos and Industry sources with respect to the nearby township of Muswellbrook. It also resulted in a more consistent anti-correlation between these 3 source types and atmospheric mixing depth than for the wind speed case. These results confirm that near surface radon concentration is more closely tied to atmospheric mixing depth (and therefore pollutant concentrations) than wind speed. The negative correlation between IndSaged and diurnal radon index implied a remote source whereas the positive correlation with Soil implied that local mining and rural activities were the primary soil sources for Muswellbrook. Little difference between the radon and wind speed cases was noted for Sea and 2ndryS sources. The time-of-day factor for the radon implementation of ME-2 was also more consistent with known diurnal ABL behaviour than results for the wind speed case.

ACKNOWLEDGMENTS

The NOAA Air Resources Laboratory (ARL) made available the HYSPLIT transport and dispersion model and the relevant input files for generation of back trajectories used in this paper. We would also like to acknowledge Pentti Paatero and Eugene Kim for supplying the ME-2 scripts which were the starting point of our final scripts. We would also like to thank the Mesoscale and Microscale Meteorology (MMM) Division of NCAR for supporting WRF development.

SUPPLEMENTARY MATERIALS

Supplementary data associated with this article can be found in the online version at http://www.aaqr.org.
REFERENCES


Figure 1: Correlation between columns of the G matrix for the standard bi-linear model.
Figure 2: Correlation between columns of the G matrix corresponding to the wind speed model.
Figure 3: Correlation between columns of the $G$ matrix corresponding to the radon model.
Figure 4: Average fingerprints generated from a bootstrapping procedure, with errors calculated as 1.96 times the standard deviation.