Environmental Science Advances

View Article Online
DOI: 10.1039/D5VA00495K

This study provides a methodology to estimate the long-term contribution of residential solid fuel burning (RSFB) to ambient fine particulate matter ($PM_{2.5}$), thus closing a gap of knowledge worldwide. This methodology does not require air quality models or emission inventories. It uses ambient concentration and meteorological data to identify the spatiotemporal pattern associated with RSFB contributions. Outliers -frequently linked to peak pollution events - are explicitly included in the methodology.

The material under study is PM2.5 from RSFB, a critical air pollutant in cities affected by residential heating sources. The results are useful for assessing air quality management policies, evaluating long-term mitigation measures, and improving evidence-based decision-making.

View Article Online
DOI: 10.1039/D5VA00495K

A novel approach to long-term source contribution of ambient $PM_{2.5}$ from residential solid fuel burning using the FUSTA methodology

Carolina Estuardo-Norambuena [1,2*]; Ana María Villalobos[1]; Hector Jorquera[1,2]

[1]Departmento de Ingeniería Química y Bioprocesos, Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860, Santiago, 7820436, Chile

[2]Centro de Desarrollo Urbano Sustentable (CEDEUS), Los Navegantes 1963, Santiago, Providencia, 7520246, Chile

*Corresponding author, Email: caestuardo@uc.cl

Abstract

Managing ambient fine particulate matter ($PM_{2.5}$) requires determining how much different $PM_{2.5}$ sources contribute over time. This study uses FUSTA (Fuzzy SpatioTemporal Apportionment) to estimate how much residential solid fuel burning (RSFB) contributes to long-term $PM_{2.5}$ concentrations. FUSTA has been applied to multiyear data for two cities in Chile and one in Poland, at which ongoing impacts of RSFB on ambient $PM_{2.5}$ are reported. These results were validated by comparison with short-term receptor model (RM) estimates for the three cities, leading to Pearson correlations r=0.84, 0.97, and 0.89, and linear slopes within 10% of unity, showing methodological consistency. The spatiotemporal pattern (STP) associated with RSFB is qualitatively similar in the three cities, with high seasonality, peak values in winter, contributions that come from all wind directions, and a diurnal cycle with peak values around midnight. The novelty of the proposed methodology is the estimation of long-term contributions of RSFB to ambient $PM_{2.5}$ for the three cities, without relying on emission inventories nor long-term RM results. The trends

View Article Online
DOI: 10.1039/D5VA00495K

observed in the long-term contribution of RSFB to $PM_{2.5}$ concentration were significant but small on an annual basis, suggesting that current RSFB regulations should be reconsidered.

Keywords: Residential solid fuel burning, source apportionment, ambient $PM_{2.5}$, fuzzy clustering, air quality management

1. Introduction

Population exposure to airborne particles with an aerodynamic size of less than 2.5 µm ($PM_{2.5}$) is estimated to cause 7 million premature deaths worldwide [1]. The health risks of $PM_{2.5}$ exposure have been extensively studied worldwide [2-4], and several studies have focused on residential heating emissions [5-7], attributing the toxicity of $PM_{2.5}$ mainly to residential fuel compounds [8-11].

In many countries, the switch to liquid or gaseous household fuels has been already implemented, yet household solid fuels like biomass or coal are still being used. Biomass alone contributes 75.1% to ambient $PM_{2.5}$ in Australia and New Zealand, 17.8% in Eastern Europe, and 16.6% in South America, without considering coal, which is widely used in Eastern Europe and can contribute more to $PM_{2.5}$ than biomass [12]. Thus, solid residential fuel PM is a global problem, and it is crucial to quantify its contribution to reduce its impact on public health and evaluate the performance of regulatory policies in many nations.

Long-term $PM_{2.5}$ levels in Chile exceed the 20 $\mu g/m^3$ ambient air quality limit in several cities in the central and southern regions, that is, substantially above the WHO's long-term average of 5 $\mu g/m^3$ [13]. This scenario is not unique to Chile. Krakow

View Article Online
DOI: 10.1039/D5VA00495K

has the greatest air pollution in Europe [14]. Poland is one of the most polluted countries in the EU [15]. In this city, the annual average limit of 25 µg/m³ is significantly surpassed [16].

The estimation of the contribution of any emission source to a given pollutant can be carried out through a Chemical Transport Models (CTM) or a receptor model (RM). The former includes a propagated uncertainty (emissions – meteorology – dispersion modeling) that is difficult to quantify. On the other hand, RM are expensive because they need to measure chemical speciation in ~ 100 samples per site under study to achieve statistical significance. In addition, if any RM is only applied for short periods of time, they do not get long-term trends [17, 18]. Thus, these methods do not provide the necessary long-term information unless expensive long-term campaigns are carried out. It is also common that available RM studies for a city have been conducted with different techniques (PMF vs. CMB for instance) and using different sets of chemical species per study, making it difficult to construct a consistent long-term source contribution record. In this regard, emission inventories could contribute to estimating long-term emissions in a different way; however, emissions, especially those from biomass burning, represent a major source of uncertainty in the inventories [19, 20]. In Chile, surveys are used to estimate biomass consumption since it is traded in an informal market. Various studies in the same location yield varying emission results [21].

So, there is a knowledge gap about how residential emissions have historically affected $PM_{2.5}$ levels in many areas worldwide.

View Article Online
DOI: 10.1039/D5VA00495K

In Chile and Poland, short-term $PM_{2.5}$ measurement campaigns have been conducted in cities with high $PM_{2.5}$ pollution from residential solid fuel burning — henceforth abbreviated as RSFB — [22-26]. However, these studies are not representative of long-term conditions because they have applied a receptor model during a short period.

An alternative way of looking at the source apportionment problem is to regard total ambient $PM_{2.5}$ as sum of spatiotemporal patterns (STP), each one corresponding to a major source (or mixture of sources) contribution. Then, a method is needed to find STP in ambient data. Cluster analysis has been applied in air pollution pattern recognition for decades. These tools have been used to study the temporal and spatial variation of pollutants and their relationship with synoptic weather [27-29], and air pollution monitoring sites' spatial similarity, but not for source apportionment. Traditional cluster analysis cannot explain ambient concentrations as a total of distinct contributions from each observation since it assigns each data point to a single cluster.

Fuzzy k-means, a fuzzy clustering method, has been recently used to look at how gaseous industrial sources contribute to ambient SO_2 [30] and the method that came out of it was named FUSTA, which stands for "*FUzzy SpatioTemporal Apportionment*". This methodology consists of a fuzzy clustering of the data, generating spatiotemporal patterns (STP) that represent the behavior of a given pollutant as a sum of contributions from different STPs, which are associated with the various emission sources present in the study area. The flexibility of fuzzy clustering consists in that each observation may be explained by a finite sum of fuzzy

View Article Online
DOI: 10.1039/D5VA00495K

contributions, emulating the way source contributions add up to total ambient concentrations. Good results have been obtained for SO_2 emitted from point sources, identifying the contribution from the largest emission sources in an industrial area [30]. This method considers both pollutant concentrations and meteorology at the measurement site. Since meteorology exerts a strong influence on the spatial variations of SO_2, the results allow for a source apportionment result with lower uncertainty in the results compared to CTM [30].

Thus, the present work has extended FUSTA to $PM_{2.5}$ to find the STP that represents $PM_{2.5}$ from RSFB, with the aim of isolating for the first time the contribution to $PM_{2.5}$ generated by this source without relying on emission inventories and CTM nor upon long-term receptor model results. In this sense, this study provides a tool for long-term monitoring of emissions from solid residential fuels, particularly in areas lacking long-term RM or CTM results, and seeks to assist in the monitoring and evaluation of public policies designed to reduce the impacts of particulate matter (PM) on public health. This present tool is meant to complement, not to replace RM for source apportionment.

This new methodology was applied in three cities that differ in the relative contribution of RSFB to the total $PM_{2.5}$, with good agreement with short-term RM results, demonstrating the robustness of FUSTA to isolate the contribution to $PM_{2.5}$ from RSFB. This is the main contribution of this new approach, and the novelty is that it allows analyses that have not been possible until now without historical RM or CTM results.

2. Methodology

View Article Online
DOI: 10.1039/D5VA00495K

2.1 Case studies and data

The city of Santiago is located at 33°25' S, 70°33' W, with a population of 5.24 million inhabitants. The city of Temuco is located at 38°45' S and 72°40' W, with a total population of 358,000 inhabitants [31]. Krakow is located at 50°3′ N, 19°56′ E, with a population of 780,000 inhabitants [32]. Santiago is classified within a Mediterranean climate (Csb) and a cold semi-arid climate (BSk); Temuco is classified as Mediterranean climate (Csb) according to the Köppen-Geiger system [33]. Krakow climate, on the other hand, falls under the temperate oceanic (Cfb) classification [34].

The Chilean Air Quality Monitoring Network (https://sinca.mma.gob.cl/) provided meteorological and air quality data for the cities of Santiago and Temuco. For the city of Santiago, the data was extracted from the "Parque O'Higgins" station, and for the city of Temuco, the "Las Encinas" station was used. For the city of Krakow, the air quality data was extracted from the Environmental Protection Inspection of Poland page (https://powietrze.gios.gov.pl/pjp/archives); the chosen station was Krasinski, which is in the center of Krakow. The meteorological data from Krakow airport was obtained from USA National Oceanic and Atmospheric Administration (https://www.ncei.noaa.gov/maps/hourly/). The chosen regulatory monitoring stations were located close to the respective sites at which RM results were available.

These cities were chosen as case studies because they consistently exceed the regulations regarding $PM_{2.5}$ and differ in their emission inventories. For cities in Chile, Santiago is a city where the contribution of RSFB is relatively low — less than 13% [35] — and Temuco is a city where the use of residential fuel for space heating is

View Article Online
DOI: 10.1039/D5VA00495K

the largest source of $PM_{2.5}$ emissions [21, 23]. For its part, Krakow is among the 30 most polluted cities in Europe [15], and 70% of households have a solid fuel boiler [16]; the emissions from combustion can reach 53% of total emissions in winter [36]. Hourly ambient concentrations of $PM_{2.5}$ and PM_{10} were collected, along with meteorological data such as wind speed and direction, temperature, and pressure. The use of meteorological variables was considered since it has been studied that atmospheric variables influence the distribution of PM in each area [37, 38]. The database spans the years 2013–2022 from Chilean cities and 2012–2019 from Krakow.

2.2 Fuzzy Clustering and visualization of the results

The applied fuzzy clustering algorithm is shown in equation (1). The second and third terms represent the modifications made to the original fuzzy k-means algorithm [39]. The second part of the equation can be thought of as a measure of information entropy for the fuzzy partition. The third part of the equation is the noise cluster, which is made up of all the outliers in the observations, and it does not consist of objects with similar properties [40]. The entropy and noise term improve stability by acting as regulators and stabilizing the partition. The entropic term allows measuring the degree of indefiniteness of a fuzzy set, avoiding hard solutions [40]. The incorporation of the noise term reduces the influence of outliers, providing more reliable clusters and minimizing the appearance of groups with similar characteristics [41].

For choosing the number of clusters to be analyzed (p), it is assumed that the clusters are associated with the number of sources found by a receptor model. A

View Article Online
DOI: 10.1039/D5VA00495K

typical receptor model application usually identifies between 3 and 7 sources, so this was the number of clusters analyzed.

For the application of the algorithm, the PM concentrations were logarithmically transformed to obtain more normal distributions. The wind speed and direction were transformed to (u,v), the Cartesian components of the wind. By standardizing them, they all contribute the same scale, and their influence on the distances in the fuzzy algorithm depends only on their relative variations, not on their original physical magnitudes[42]. Outliers were not removed, and the concentrations cannot be negative, as the minimum value of these is given by the detection limit of the measuring instrument. Finally, the chosen fuzzy clustering algorithm standardized all variables before processing them.

$$\min_{U,C} J_{FKMNE} = \sum_{i=1}^{n}\sum_{k=1}^{p} u_{ik}\cdot\|x_i - c_k\|^2 + t\cdot\sum_{i=1}^{n}\sum_{k=1}^{p} u_{ik}\cdot log(u_{ik}) + \sum_{i=1}^{n}\delta^2\left(1-\sum_{k=1}^{p} u_{ik}\right)^2$$

$$u_{ik} \in [0,1]; \sum_{k=1}^{p+1} u_{ik} = 1 \quad (1)$$

Equation (1) was solved using the *fclust* package available in the open software environment R (routine *fkm.ent.noise*). The solutions of (1) were estimated using different cluster numbers (*p*) as described above. After solving the previous equation (for a fixed value of *p*), the concentration of $PM_{2.5}$ for any observation "*i*" can be written as follows:

$$PM_{2.5,i} = \sum_{k=1}^{p+1} PM_{2.5,i}\cdot u_{ik} = \sum_{k=1}^{p} PM_{2.5,k} + PM_{2.5,\,noise} (2)$$

View Article Online
DOI: 10.1039/D5VA00495K

The above equation implicitly means that we assume the equivalence of probability of belonging to a cluster with the mass fraction of total $PM_{2.5}$ being apportioned to that same cluster. Thus, starting with equation (2), the STP features for the different $PM_{2.5}$ clusters found through FUSTA are inspected using the *openair* package in R environment software to make temporal variation graphs and two bivariate graphs: (wind speed, wind direction, $PM_{2.5}$) and (temperature, wind direction, $PM_{2.5}$) to characterize the different STP contributions. Then, the resulting STP (clusters) were visualized in their temporal variability. The clusters that best showed how biomass contributions vary in each city were qualitatively chosen based on how they behaved and how similar they were to patterns found in previous studies.

2.3 Validation of the FUSTA results

To validate the new methodology proposed, the different fuzzy clusters determined by applying FUSTA (equations 1 and 2) were compared with quantitative estimates of the RSFB contribution of $PM_{2.5}$ using the receptor model obtained from previous campaigns. Two campaigns were conducted in Santiago and Temuco in 2013 and 2014 [26, 43], respectively, using the Chemical Mass Balance (CMB) RM, and for Krakow from 2018 and 2019, the rolling Positive Matrix Factorization (PMF) RM was used, along with molecular markers for coal and wood combustion, obtained with an ACSM equipment [44]. The "Lmodel2" statistical program in the open software R was used to fit major axis linear regressions to determine whether there was a significant correlation between the concentrations in the fuzzy clusters and those obtained by the above-described receptor model campaigns. Since all fuzzy clusters resolved by FUSTA have little correlation among themselves (Supplementary Tables S1-S3),

View Article Online
DOI: 10.1039/D5VA00495K

this suggests that the identification of which fuzzy cluster corresponds to RSFB should be unique.

To compare the results of FUSTA with RM (discrete or continuous sampling), the data corresponding to the date range when the sampling campaigns for RM were conducted were matched with the hourly results from FUSTA, ensuring that FUSTA and RM were compared on the same temporal window.

In a quantitative way, we compared all fuzzy clusters with the RSFB receptor model results using statistical metrics such a correlation coefficient, mean fractional bias and mean fractional error[45], to confirm the qualitative choice mentioned above. A parsimony criterion was used to choose the solution with the fewest clusters that allowed the RSFB contribution (or fuzzy cluster) to be identified. Using the principle of parsimony is the most recommended practice when using cluster structures, as it mitigates the complexity of the results and preserves homogeneity [46], since an excessively high number of clusters tends to introduce errors [47]. In this context, and given that no improvements were observed in the models by increasing *p,* no additional sensitivity analysis was conducted.

3. Results and discussion

3.1 Identification of the STP corresponding to RSFB contributions

The fuzzy algorithm (1) was applied to the PM and the meteorological data for the study period. The number of clusters (p) varied between 3 and 7 in all cases.

View Article Online
DOI: 10.1039/D5VA00495K

Below are the selected results for the city of Santiago (5 clusters + 1 noise), Temuco (4 clusters + 1 noise), and Krakow (5 clusters + 1 noise), as well as the analysis used to identify the respective clusters in each case.

It is important to mention that in all the cases studied, the *p* clusters obtained with FUSTA do not exhibit similar characteristics, as shown in tables S1 to S3. These tables display the correlation coefficient between the selected cluster in each case and the remaining obtained clusters. The results show that there are no similarities between them. The same happens when correlating the rest of the clusters with each other. This is due to the use of the noise group, which aims to strengthen the algorithm by eliminating bias and stabilizing the partition of the clusters [41, 48, 49], reducing the likelihood of similar clusters that occur in the implementation of other approaches.

Figure 1(a) shows the five clusters found in Santiago. A previous study [50] published a daily profile for aerosols from RSFB (denoted by biomass burning organic aerosol, BBOA) in this city, so the chosen profile must follow this pattern, and this source contribution magnitude is below 13% [35]. Following these criteria, the RSFB cluster was found to be number 5 ($PM_{2.5}$_5 in Figure 1(a)), as it has all the required features and a strong seasonal increase, showing most of its contribution in the winter (June, July, and August) when more residential solid fuel is used because of the lower ambient temperatures. The bivariate wind direction/temperature graphs in Figure S1 of the supplementary material also show that this source increases when temperatures are lower. The bivariate wind direction/wind speed graph for this cluster (Figure S2) was looked at, and its contribution was arriving at the receptor

View Article Online
DOI: 10.1039/D5VA00495K

site from all directions, increasing at low wind speeds, which is a distinctive behavior of area source emissions. This is similar to what other studies have found for this source in other parts of the world [51, 52] and supports our choice of this cluster as the source corresponding to RSFB. On the other hand, a correlation analysis among the five clusters shows that they are weakly correlated to each other (Table S1, supplementary material), suggesting that the FUSTA method can find STP that are close to being statistically independent to each other. It is also important to observe the noise pattern, as it may include a contribution from RSFB. In this case for Santiago, the noise cluster obtained represents 1.3% of the total $PM_{2.5}$ concentration, and its temporal behavior does not match that of RSFB; therefore, qualitatively, it is ruled out that in this case it contributes to the concentration. Quantitatively, this was also analyzed, as detailed further in section 3.3.

Figure 1(b) shows the solution of four clusters found for the city of Temuco. No published temporal profile exists for this pollutant in this city, but we know some of its previously reported characteristics, which helped us to identify it. About 90% of all emissions in this city come from residential fuel [23, 53], and its concentration accumulates between 8 PM and 11 PM due to the lower temperatures during this time, which influence the use of domestic heating [54]. Additionally, the chosen profile must exhibit seasonal behavior, increasing its presence during winter due to higher fuel consumption during this period. Considering all the above criteria, the chosen profile in this case is cluster 2 ($PM_{2.5}$_2 in Figure 1(b)), as it meets all the above features. Also, it is important to note that the bivariate graphs for this cluster show a uniform spatial distribution with a higher concentration at low speeds (Figure S3),

View Article Online
DOI: 10.1039/D5VA00495K

typical of area sources dispersion, and an increase in magnitude at low temperatures (Figure S4), typical of increased RSFB emissions driven by increased household heating demand. We again find little statistical correlation among the clusters found in this city (Table S2). For Temuco, the noise cluster is greater than in Santiago, approximately 5% of the total $PM_{2.5}$ concentration. Regarding its behavior, it follows a pattern like the chosen RSFB cluster, so it seems appropriate to include it into the chosen RSFB pattern. This decision will be quantitatively corroborated in section 3.3.

Figure 1(c) shows the solution found in the city of Krakow. Previous studies have shown the temporal profiles of solid fuels [16], which were the basis for the selection of RSFB contribution in this city. It is known that the burning of solid fuels has a seasonal behavior in this city [25, 36], showing its maximum values in the winter season, and its prevalence during the night should be observed, as has already been reported in previous studies [16]. With all these characteristics, the chosen cluster is number 4 ($PM_{2.5\_4}$ in Figure 1(c)). The bivariate graphs confirmed this source comes from many wind directions, with a higher concentration at low speeds (Figure S5) and at low temperatures (Figure S6). Again, there is little statistical correlation among the clusters identified in this city, as in the previous two cases (Table S3). In this city, the noise cluster is approximately 3% of the total $PM_{2.5}$ concentration. In this case, the noise contribution behavior is like the RSFB cluster, so adding it to the RSFB concentration is suggested to improve accuracy. This decision will be quantitatively corroborated in section 3.3.

3.2 Temporal variability of residential solid fuel burning contributions

View Article Online
DOI: 10.1039/D5VA00495K

Figure 2 shows the temporal variability obtained for RSFB contributions once the fuzzy clustering algorithm was applied to the cities of Santiago (a), Temuco (b), and Krakow (c). In Figure 2, we can observe the hourly and monthly behavior of the $PM_{2.5}$ concentration originating from RSFB contributions in the three cities.

The specific characteristics of each city, particularly the meteorological ones, explain the difference in behavior of the RSFB contribution. At midnight in Santiago, the planetary boundary layer is at its minimum and the wind speed drops to less than half of what it is during the day, and sometimes it is almost zero [37, 55]. The above condition causes an accumulation of $PM_{2.5}$ from the residential burning of solid fuels throughout the night, and its dispersion begins around mid-morning. The monthly variability in Santiago is as expected; the largest contribution is in the winter, when biomass is used more because of the cooler weather, and it tends to go down in the summer, when temperatures rise, and residential solid fuel burning is less used. Finally, the concentration that shows the temporal pattern corresponds with the contributions reported for the city, around 13% [35].

There is a different hourly variability in Temuco when compared with Santiago. This is because residential solid fuel burning is the main source of $PM_{2.5}$ in Temuco, accounting for about 90% of all emissions [23, 53]; both cities have different meteorological conditions, and residential solid fuel burning has a strong cultural component in this city[56]. The behavior observed in Figure 2(b) matches what has been stated in previous studies [23, 53, 54]. In the hourly variability, an accumulation of $PM_{2.5}$ is observed between 8:00 p.m. and 11:00 p.m. due to the lower temperatures during this time, which promote the use of residential heating [54]. In another study

View Article Online
DOI: 10.1039/D5VA00495K

that looked at specific compounds in PM, like levoglucosan or soluble potassium, which are tracers of residential solid fuel burning emissions, these compounds were found to be higher at night in the same city [53], and showed an increase between 8 and 11 am, after its observed decrease in the early morning, matching the temporal variability in Figure 2(b). Another study from Temuco [23], which looked at levoglucosan over the course of a year, showed that it behaved in a way like the monthly changes shown in Figure 2(b), with its highest level observed in the winter and decreasing levels in the summer. The difference between the late nighttime peak observed in Santiago and Temuco can be explained by the meteorological conditions in both cities. In Santiago, there is a katabatic recirculation of air masses from the mountains to the valley, which results in a longer residence time for pollutants such as $PM_{2.5}$ [37, 54]. This does not occur in the city of Temuco, located in a flat basin at which higher wind speeds are observed at night as compared with Santiago [54].

Krakow, for its part, is in a valley surrounded by mountains and often shrouded in fog, which forms a barrier that prevents large circulations and favors the accumulation of pollutants [24, 57]. In the afternoon and evening, as the temperature decreases, the use of residential solid fuels increases, which raises the presence of $PM_{2.5}$ from this source, as shown in Figure 3. At night, thermal inversions and calm winds are frequent in the city [14, 57], which explains the persistence of $PM_{2.5}$ and its dominance during the night, as various studies have already pointed out [16, 24, 36]. During the day, the dispersion of $PM_{2.5}$ is driven by vertical mixing [16]. Several studies have already looked at the seasonal changes in $PM_{2.5}$ from residential solid fuels

View Article Online
DOI: 10.1039/D5VA00495K

and found similar results [24, 36], for example, they found that biomass increases its contribution from 5% in the summer to 17% in the winter [25]. In this way, the pattern found for solid fuels in the city of Krakow matches the mentioned characteristics.

Based on these results, it is possible to indicate that the methodology proposed in this study allows for a qualitative visualization of the behavior of $PM_{2.5}$ originating from biomass, in cities where it is a secondary source (Santiago and Krakow) and in cities where it is the largest source (Temuco). This is a desirable characteristic of any technique that allows for distinguishing low source contributions from high source contributions on air quality.

3.3 Comparison of FUSTA and receptor model results

The FUSTA results were compared with those from two receptor model campaigns that were held in Santiago and Temuco [26, 43] and one in Krakow [16].

To find out if there is a relationship between the data predicted by the biomass STP presented in Section 3.2 above and those found in those studies, a simple linear regression was performed using the statistical package "lmodel2" from the open software "R". For the cities of Santiago, Temuco and Krakow, the respective correlation coefficients were r=0.84 (p-value = 0.018), r=0.97 (p-value = 0.0001), and r=0.89 (p-value = 0.0000), see Figure 3. Due to the uncertainty of both the FUSTA data and the receptor model, the most suitable regression for this analysis in a type II regression model approach is primarily [SMA/RMA] [58]. However, four regression methods were analyzed to facilitate comparisons with other studies. The confidence intervals and coefficients for the different regression methods computed with

View Article Online
DOI: 10.1039/D5VA00495K

package "lmodel2" are shown in Tables S4 through S9 in the supplementary material. Owing to the good linear fit results (r values close to 1), there is a significant linear relationship between the RSFB concentrations found by the receptor model and those predicted by the FUSTA methodology. Observing the slopes for Santiago, Temuco and Krakow, the average slope from the regressions is 1.04, 0.94, and 1.44, respectively, suggesting a 1 to 1 relationship between the receptor model's results and those obtained through FUSTA. It is important to mention that the chemical speciation data reported for Krakow and used in this study belongs to PM_1, which results in the slope being greater than 1 (1.44). The Tobler et al. study's supplementary material provides the linear relationship between $PM_{2.5}$ and PM_1 at the reference station, yielding a slope value of 0.77. Introducing this correction in the slope results in a value of 1.1. These slope results close to 1.0 provide quantitative support for the implicit assumption in equation (2) of assuming probability of belonging to any given cluster equals mass fraction apportioned to that same cluster. It should be mentioned that in the three case studies, in addition to the correlation coefficient, the values of Mean Fractional Bias (MFB), Mean Fractional Error (MFE), and Root Mean Square Error (RMSE) were computed, which are shown in Figure 3. In all three cases, the parameters were found to be within the accepted criteria for air quality models [45].

In the city of Santiago, the addition of the noise cluster resulted in a lower correlation coefficient, which corroborates the analysis in section 3.1, ruling out that the noise group is part of the RSFB. For Temuco and Krakow, adding the noise group resulted in a better correlation than the one obtained without considering this group, which

View Article Online
DOI: 10.1039/D5VA00495K

aligns with the analysis conducted in the previous section, where the noise group was assigned in this case as part of the RSFB concentration.

As an additional analysis, the RM data for RSFB was compared with the rest of the clusters for the three cities. The results are shown in tables S10 through S11. In all cases, the R values are negative or show a significantly smaller correlation compared to the one chosen in each analysis.

Supplementary Figure S7 shows a detailed comparison of the time variability of FUSTA results along with an estimate of total RSFB contributions from the RM model applied in Krakow. To do this, the organic carbon (OC) contributions to coal and wood combustion obtained by Tobler et al (2021) were multiplied by $PM_{2.5}$/OC ratios for those two fuels published in the literature [59-61], and then the concentration of ammonium chloride (from coal combustion) and the elemental carbon contribution from those two solid fuels measured by Tobler et al (2021) were added up to estimate the total RM RSFB contribution in Krakow. As can be seen in Figure S7, the similitude between RM and FUSTA estimates is remarkable.

The above results show that the FUSTA approach accurately estimates the RSFB contributions to the total $PM_{2.5}$ in these cities. Therefore, the resulting long-term FUSTA results provide a timeline of those contributions, from which trends may be estimated, and the effectiveness of air quality regulations may be assessed as well.

3.4 Long-term RSFB behavior

With the FUSTA results validated for the temporal and spatial profiles of RSFB $PM_{2.5}$, it is possible to investigate its long-term trend for the three cities. Figures 4(a) and

View Article Online
DOI: 10.1039/D5VA00495K

(b) show the percentage of the monthly average of $PM_{2.5}$ coming from RSFB for the period 2013–2022 for Santiago and Temuco, respectively, and Figure 4(c) shows the result of Krakow for the period 2012–2019. In Figure S8 of the supplementary material, the temporal linear trends of these contributions, estimated using the TheilSen function from the openair package in R, are presented.

For the city of Santiago (Figure 4a), a decrease in biomass contribution is observed. In Figure S8(a), the trend of decrease is 0.11 $\mu g/m^3$ per year (95% CI: [-0.14, -0.08]). This trend is significant yet small in magnitude, which means it would take decades for the data in Figure 4(a) to decrease below 5 $\mu g/m^3$ (for example) if this trend remains unchanged. In Santiago, a regulation bans the use of wood stoves during periods of environmental contingency. However, enforcing compliance with this regulation is difficult, which explains in part the small downward trend estimate.

In the case of Temuco Figure S8(b) of the supplementary material shows a significant downward trend of 0.45 $\mu g/m^3$ per year (95% CI: [-0.52, -0.38]), which is larger in magnitude than the one estimated for Santiago. Temuco was one of the first Chilean cities to start regulating and implementing initiatives to control and lower RSFB $PM_{2.5}$ [13]. In fact, the number of households subsidies for home refurbishment and wood stove changeouts granted in Santiago is considerably lower (as % of total households) than those granted in Temuco. This explains the larger downward trend found in Temuco. However, like in the case of Santiago, if the trend magnitude remains the same for the times ahead, it would take decades for the city to reach the Chilean ambient air quality standard of 20 $\mu g/m^3$.

View Article Online
DOI: 10.1039/D5VA00495K

Krakow turns out to be the city where the largest decrease is estimated, as shown in Figure 4(c). On average, during the entire study period, the downward trend is 1.28 μg/m³ per year (95% CI: [-1.46, -1.15]), see Figure S8(c). The decrease observed in Figure 4(c) agrees with the regulation implemented in the city of Krakow that bans the burning of solid fuels, wood, and heavy oil in domestic stoves beginning in September 2019 [57]. Since our study period ends in 2019, so it is not possible to estimate the regulatory effects upon this RSFB contribution. However, studies conducted after the regulation show that the environmental conditions of the city are still poor [32] and that the problem is partly due to its location, which receives pollution from the surrounding areas [14].

An important point to remark here is that, in the summer, when temperatures peak, the pattern of RSFB contributions in the three studied cities is near zero. This is explained by the fact that residential heating is strongly linked to temperature, unlike its other uses, such as cooking, which does not have a meteorological component so influential that it allows for differentiation through FUSTA. This is a model limitation that should be considered. However, for Chile, the use of RSFB for cooking does not exceed 5.4% [62], and for the city of Krakow, RM studies have found that the contribution of RSFB to PM2.5 varies between 3% and 6% [36] of the total. So, for these case studies, this limitation does not seem relevant. Summarizing the above results, they are useful for monitoring the RSFB impacts and estimating their effectiveness. In Chile, previous research has shown that the Atmospheric Decontamination Plans (ADP) are not being carried out properly and are not working efficiently to reduce residential wood consumption [63]. This can be attributed to their

View Article Online
DOI: 10.1039/D5VA00495K

emphasis on visible actions rather than the most effective ones. Other studies show how important it is to think about the fact that many households in southern Chile are currently energy poor. This is because updating homes to make them more energy efficient might only make a small difference in how much wood is used, because people care more about their comfort than lowering their heating needs [64]. In Poland, there seem to be similarities with what has been observed in Chile; a large part of the population cannot afford to change their energy source or maintain adequate temperatures in their homes. Additionally, it has been observed that the Polish population does not identify residential heating as a significant source of pollution, indicating that not only are environmental policies needed, but also more education on the topic [32].

4. Conclusions

We have found that it is possible to identify the ambient $PM_{2.5}$ concentrations that come from RSFB using the new methodology proposed. It was possible to validate the new methodology results by comparison with previously conducted short-term receptor model campaigns, confirming a significant linear relationship between the RM and FUSTA results — Pearson correlations r = 0.84, 0.97 and 0.89 for Santiago, Temuco and Krakow, respectively — and that the concentrations predicted by the model correspond directly to the concentrations found by the RM (regression slopes of 1.04, 0.94 and 1.1 for Santiago, Temuco and Krakow, respectively). These slope values near 1.0 support the implicit assumption in equation (2) that the probability of a $PM_{2.5}$ observation belonging to a cluster is equal to the PM2.5 mass fraction associated with RSFB.

View Article Online
DOI: 10.1039/D5VA00495K

This new methodology provides a long-term estimate of how RSFB contribution changes over time and space in cities at which biomass is not a dominant source (like Santiago, Chile) and in cities where it is a dominant source (like Temuco, Chile or Krakow, Poland).

Another relevant result is that all three RSFB patterns identified by FUSTA share similar qualitative features: high seasonality with highest contributions in winter and essentially zero contributions in summer, concentrations that increase as temperature and wind speed decrease and that come from all wind directions, typical of area sources. This suggests that in any urban area with RSFB emissions, the proposed methodology would be useful to estimate RSFB contributions even if no receptor model results are available for calibration. The information requirements are modest and correspond to the information typically monitored in urban areas: ambient $PM_{2.5}$ and meteorological measurements of wind speed and direction, air temperature and pressure. Hence, the proposed methodology helps in closing the gap in long-term source apportionment estimates for RSFB emissions, particularly in developing countries. The resulting information is useful for assessing the effectiveness of past air quality regulations, as input in epidemiological studies aimed at estimating RSFB health impacts, and for constraining CTM simulations to estimate the trends in RSFB emissions.

One limitation of the methodology is that it estimates zero RSFB contributions in summer season, when temperatures are highest. However, in the cities at which RSFB is relevant, there is a small summer contribution, which may include cooking and industrial use of solid fuels.

View Article Online
DOI: 10.1039/D5VA00495K

For the three cities analyzed, significant downward trends were estimated; however, they are rather small in magnitude, especially for the two Chilean cities, suggesting more regulations need to be implemented at those cities to comply with ambient air quality standards.

Acknowledgments

This research was financially supported by the Becas de Doctorado Nacional Doctoral Scholarship program, grant ANID-PFCHA/2022–21220090 granted by Carolina Estuardo-Norambuena, and by grant ANID CEDEUS CIN250009.

Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM- 02).

Author contributions

Carolina Estuardo-Norambuena: Investigation, Conceptualization, Methodology, Formal analysis, Visualization, Model validation, Writing – original draft, Writing – review & editing, Founding acquisition.

Hector Jorquera: Conceptualization, Methodology, Supervision, Writing – review & editing.

Ana Maria Villalobos: Formal analysis, Model validation.

View Article Online
DOI: 10.1039/D5VA00495K

Figures

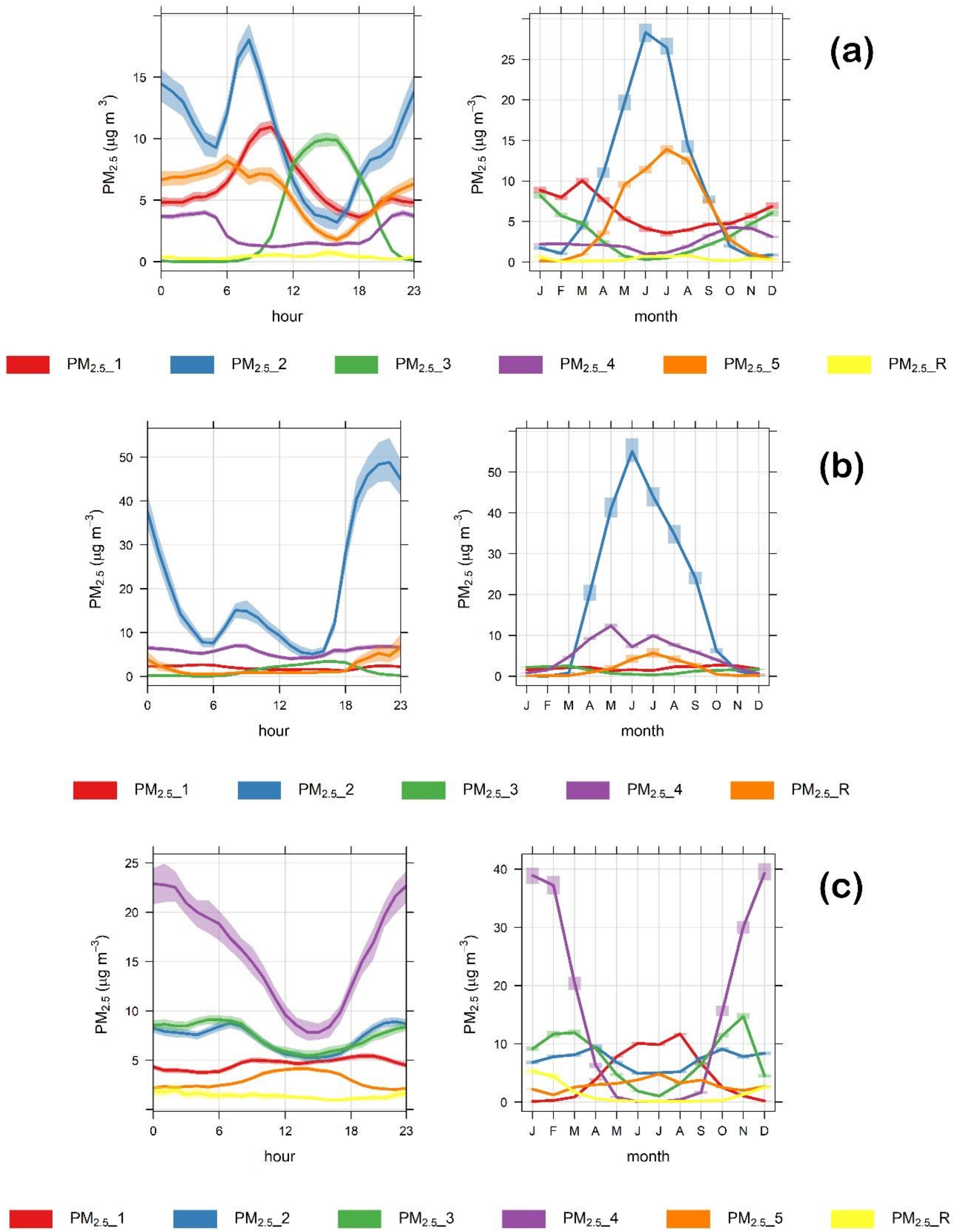

Figure1. Visualization of temporal variation for the solution of clusters found for the city of Santiago (a), Temuco (b) and Krakow (c).

View Article Online
DOI: 10.1039/D5VA00495K

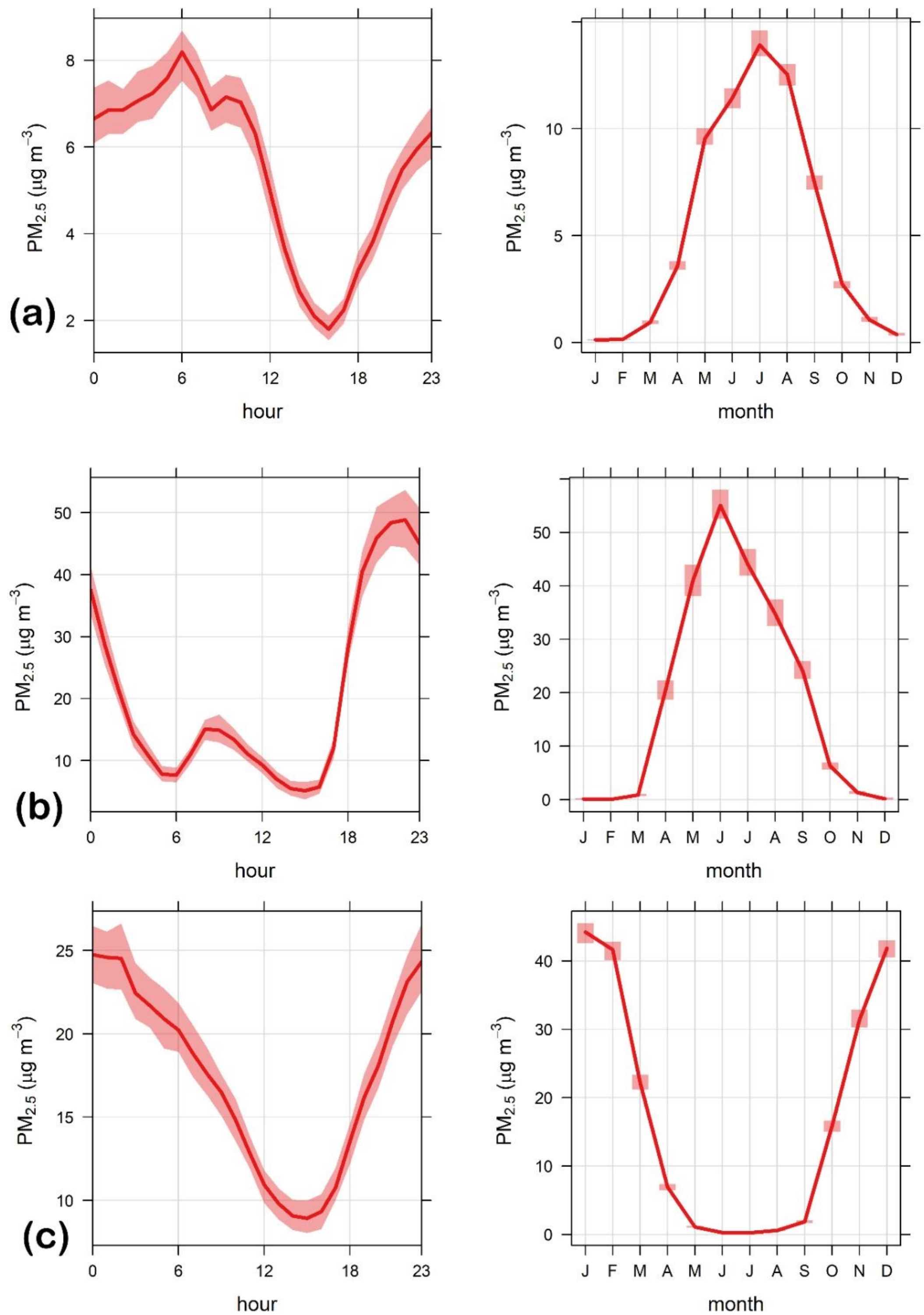

Figure 2. Temporal pattern of $PM_{2.5}$ concentration from biomass Santiago (a), Temuco (b) and Krakow (c).

View Article Online
DOI: 10.1039/D5VA00495K

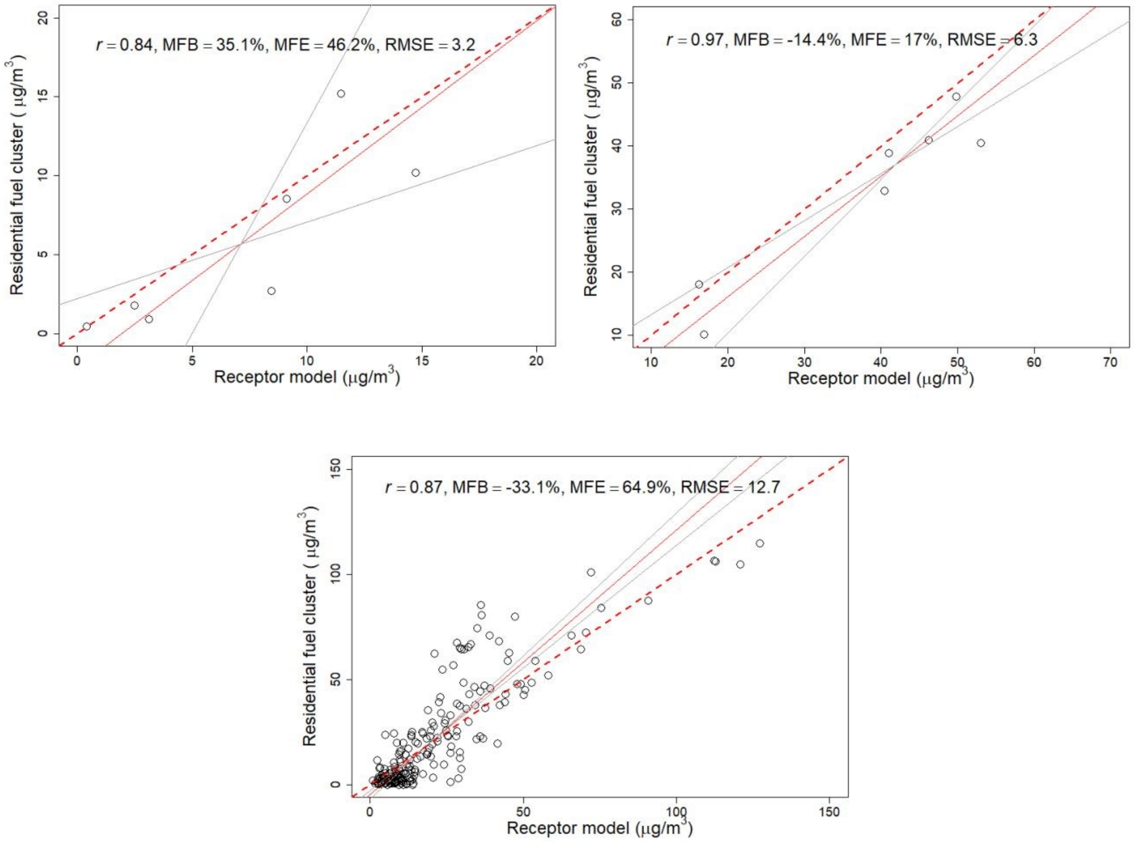

Figure 3: Major axis regressions for the results of receptor model campaigns and the results of Fuzzy Clustering for the cities of Santiago (a), Temuco (b) and Krakow (c)

View Article Online
DOI: 10.1039/D5VA00495K

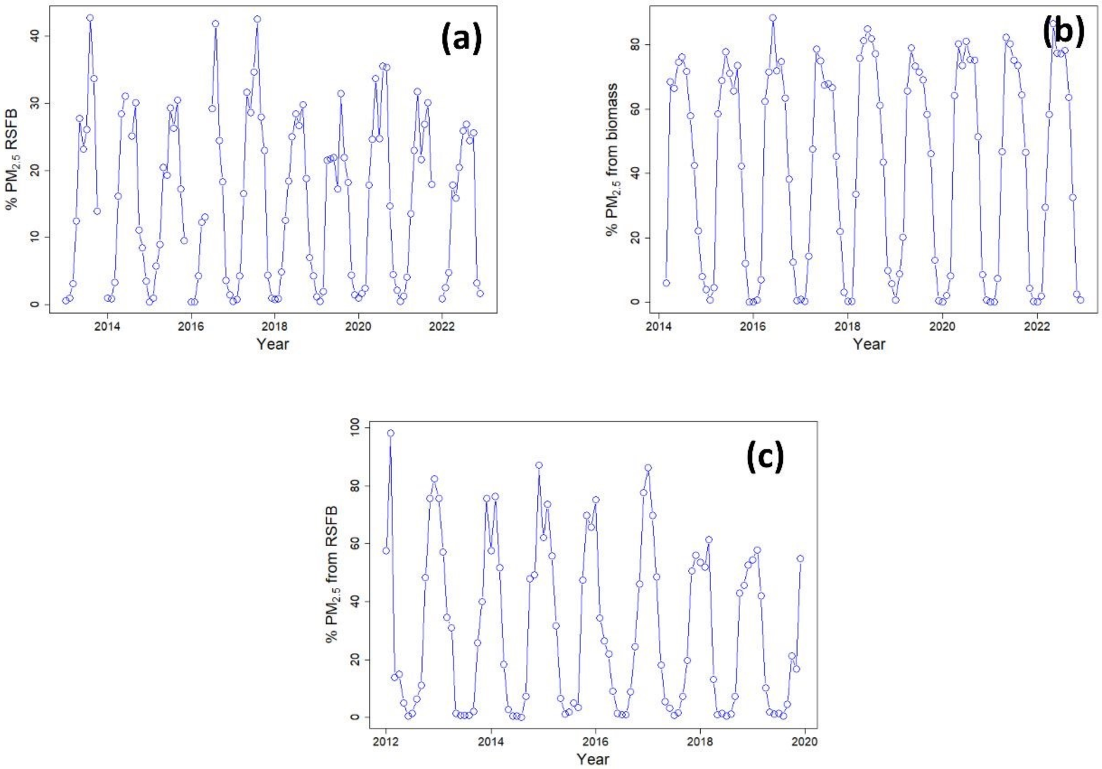

Figure 4: Long-term % of $PM_{2.5}$ coming from RSFB for Santiago (a), Temuco (b), Krakow(c).

View Article Online
DOI: 10.1039/D5VA00495K

References

1. WHO, https://www.who.int/publications/i/item/9789240034228, 2021.
2. K. Balasubramani, K. A. Prasad, N. K. Kodali, N. K. Abdul Rasheed, S. Chellappan, D. K. Sarma, M. Kumar, R. Dixit, M. M. James, S. K. Behera, S. Shekhar and P. Balabaskaran Nina, *Frontiers in Public Health*, 2022, **10**.
3. L. Gilardi, M. Marconcini, A. Metz-Marconcini, T. Esch and T. Erbertseder, *International Journal of Health Geographics*, 2023, **22**, 11.
4. A. S. Kagabo, B. Safari, Y. T. Brou, J. Gasore and B. K. Mutai, *Air Quality, Atmosphere & Health*, 2024, DOI: 10.1007/s11869-024-01672-6.
5. H. Orru, H. Olstrup, J. Kukkonen, S. López-Aparicio, D. Segersson, C. Geels, T. Tamm, K. Riikonen, A. Maragkidou, T. Sigsgaard, J. Brandt, H. Grythe and B. Forsberg, *BMC Public Health*, 2022, **22**, 1286.
6. S. Yin, *Journal of Environmental Management*, 2023, **330**.
7. F. Liang, Z. Xu, L. Ding, Z. Zhu, M. Chen, H. Shu, X. Huang, Z. Su, X. Wang, Y. Xiao, S. Huang, D. Mai, E. Yi, P. Xu and W. Zhang, *Journal of Hazardous Materials*, 2025, **494**, 138569.
8. Y. A. Cipoli, C. Alves, M. Rapuano, M. Evtyugina, I. C. Rienda, N. Kováts, A. Vicente, F. Giardi, L. Furst, T. Nunes and M. Feliciano, *Atmospheric Environment*, 2023, **303**, 119771.
9. E. D. Vicente, D. Figueiredo, C. Gonçalves, I. Lopes, H. Oliveira, N. Kováts, T. Pinheiro and C. A. Alves, *Science of The Total Environment*, 2021, **782**, 146820.
10. I. Tsiodra, G. Grivas, K. Tavernaraki, D. Paraskevopoulou, C. Parinos, M. Tsagkaraki, E. Liakakou, A. Bougiatioti, E. Gerasopoulos and N. Mihalopoulos, *Journal of Hazardous Materials*, 2025, **494**, 138431.
11. C. Zhang, M. Yang, J. Li, H. Wang, L. Song, L. Shen, L. Bai, Y. Lin, J. Liu and B. Wang, *Journal of Hazardous Materials*, 2024, **480**, 135764.
12. P. K. Hopke, Q. Dai, L. Li and Y. Feng, *Science of The Total Environment*, 2020, **740**, 140091.
13. H. Jorquera, *Urban Climate*, 2021, **35**, 100764.
14. T. Danek, E. Weglinska and M. Zareba, *Scientific Reports*, 2022, **12**, 11050.
15. EEA, 2022.
16. A. K. Tobler, A. Skiba, F. Canonaco, G. Močnik, P. Rai, G. Chen, J. Bartyzel, M. Zimnoch, K. Styszko, J. Nęcki, M. Furger, K. Różański, U. Baltensperger, J. G. Slowik and A. S. H. Prevot, *Atmospheric Chemistry and Physics*, 2021, **21**, 14893-14906.
17. S. Faridi, F. Yousefian, V. Roostaei, R. M. Harrison, F. Azimi, S. Niazi, K. Naddafi, F. Momeniha, M. Malkawi, H. A. Moh'd Safi, M. K. Rad and M. S. Hassanvand, *Environmental Pollution*, 2022, **310**, 119889.
18. M. J. Skiles, A. M. Lai, M. R. Olson, J. J. Schauer and B. de Foy, *Environmental Pollution*, 2018, **237**, 366-376.
19. S. Azmi and M. Sharma, *Chemosphere*, 2023, **336**, 139195.
20. S. Lopez-Aparicio, H. Grythe and M. Markelj, *Journal*, 2022, **15**.

View Article Online
DOI: 10.1039/D5VA00495K

21. N. Álamos, N. Huneeus, M. Opazo, M. Osses, S. Puja, N. Pantoja, H. Denier van der Gon, A. Schueftan, R. Reyes and R. Calvo, *Earth Syst. Sci. Data*, 2022, **14**, 361-379.
22. H. Jorquera, A. M. Villalobos and J. J. Schauer, *Atmospheric Pollution Research*, 2021, **12**, 50-59.
23. M. E. Quinteros, E. Blanco, J. Sanabria, F. Rosas-Diaz, C. A. Blazquez, S. Ayala, J. P. Cárdenas-R, E. A. Stone, K. Sybesma, J. M. Delgado-Saborit, R. M. Harrison and P. Ruiz-Rudolph, *Atmospheric Environment*, 2023, **294**.
24. L. Samek, Z. Stegowski, L. Furman, K. Styszko, K. Szramowiat and J. Fiedor, *Water, Air, & Soil Pollution*, 2017, **228**, 290.
25. L. Samek, A. Turek-Fijak, A. Skiba, P. Furman, K. Styszko, L. Furman and Z. Stegowski, *Atmosphere*, 2020, **11**.
26. A. M. Villalobos, F. Barraza, H. Jorquera and J. J. Schauer, *Environmental Pollution*, 2017, **225**, 514-523.
27. Z. Chen, D. Chen, C. Zhao, M.-p. Kwan, J. Cai, Y. Zhuang, B. Zhao, X. Wang, B. Chen, J. Yang, R. Li, B. He, B. Gao, K. Wang and B. Xu, *Environment International*, 2020, **139**, 105558.
28. E. Gramsch, F. Cereceda-Balic, P. Oyola and D. von Baer, *Atmospheric Environment*, 2006, **40**, 5464-5475.
29. L. Zhang and G. Yang, *Environmental Research*, 2022, **204**, 112009.
30. H. Jorquera and A. M. Villalobos, *Journal of Hazardous Materials*, 2023, **443**, 130335.
31. I. d. E. INE, *Santiago: Instituto Nacional de Estadísticas*, 2018.
32. W. Zgłobicki and B. Baran-Zgłobicka, *Environmental Research*, 2024, **240**, 117497.
33. P. Sarricolea, M. Herrera-Ossandon and Ó. Meseguer-Ruiz, *Journal of Maps*, 2017, **13**, 66-73.
34. M. Kottek, J. Grieser, C. Beck, B. Rudolf and F. Rubel, *Meteorologische Zeitschrift*, 2006, **15**, 259-263.
35. F. Barraza, F. Lambert, H. Jorquera, A. M. Villalobos and L. Gallardo, *Atmospheric Chemistry and Physics*, 2017, **17**, 10093-10107.
36. L. Samek, Z. Stegowski, K. Styszko, L. Furman and J. Fiedor, *Environmental Pollution*, 2018, **241**, 406-411.
37. I. Olivares, J. Langner, C. Soto, E. A. Monroy-Sahade, A. Espinosa-Calderón, P. Pérez, M. A. Rubio, A. Arellano and E. Gramsch, *Atmospheric Environment*, 2023, **310**, 119947.
38. I. A. Pérez, M. Á. García, M. L. Sánchez, N. Pardo and B. Fernández-Duque, *Journal*, 2020, **17**.
39. J. C. Bezdek, R. Ehrlich and W. Full, *Computers & Geosciences*, 1984, **10**, 191-203.
40. M. B. Ferraro and P. Giordani, *Fuzzy Sets and Systems*, 2015, **279**, 1-16.
41. R. N. Dave, *Pattern Recognition Letters*, 1991, **12**, 657-664.
42. I. K. Khan, H. Daud, N. Zainuddin and R. Sokkalingam, *Alexandria Engineering Journal*, 2025, **118**, 246-260.
43. A. M. Villalobos, F. Barraza, H. Jorquera and J. J. Schauer, *Science of the Total Environment*, 2015, **512-513**, 133-142.
44. A. K. Tobler and A. S. H. Prevot, 2021.

View Article Online
DOI: 10.1039/D5VA00495K

45. J. W. Boylan and A. G. Russell, *Atmospheric Environment*, 2006, **40**, 4946-4959.
46. H. Tao, P. Youquan, Y. Jinhong and Z. Wenyang, *The Annals of Applied Statistics*, 2025, **19**, 1362-1381.
47. D. Frolov, S. Nascimento, T. Fenner and B. Mirkin, *Information Sciences*, 2020, **512**, 595-615.
48. M. B. Ferraro, *Econometrics and Statistics*, 2024, **30**, 110-123.
49. O. Koc, F. Baser and A. S. Selcuk-Kestel, *Applied Soft Computing*, 2024, **167**, 112430.
50. S. Carbone, S. Saarikoski, A. Frey, F. Reyes, P. Reyes, M. Castillo, E. Gramsch, P. Oyola, J. Jayne, D. R. Worsnop and R. Hillamo, *Aerosol and Air Quality Research*, 2013, **13**, 462-473.
51. I. Stavroulas, A. Bougiatioti, G. Grivas, D. Paraskevopoulou, M. Tsagkaraki, P. Zarmpas, E. Liakakou, E. Gerasopoulos and N. Mihalopoulos, *Atmospheric Chemistry and Physics*, 2019, **19**, 901-919.
52. I. Tsiodra, G. Grivas, A. Bougiatioti, K. Tavernaraki, C. Parinos, D. Paraskevopoulou, K. Papoutsidaki, M. Tsagkaraki, F.-A. Kozonaki, K. Oikonomou, A. Nenes and N. Mihalopoulos, *Science of The Total Environment*, 2024, **951**, 175416.
53. F. Reyes, S. Ahumada, F. Rojas, P. Oyola, Y. Vásquez, C. Aguilera, A. Henriquez, E. Gramsch, C.-M. Kang, S. Saarikoski, K. Teinilä, M. Aurela and H. Timonen, *Aerosol and Air Quality Research*, 2021, **21**, 210110.
54. M. A. Yáñez, R. Baettig, J. Cornejo, F. Zamudio, J. Guajardo and R. Fica, *Atmospheric Environment*, 2017, **161**, 221-234.
55. R. C. Muñoz, R. Garreaud, J. A. Rutllant, R. Seguel and M. Corral, *Atmosphere*, 2023, **14**.
56. B. Álvarez Escobar and Á. Boso Gaspar, *Revista internacional de contaminación ambiental*, 2018, **34**, 527-540.
57. A. Ziernicka-Wojtaszek, Z. Zuśka and J. Kopcińska, *Sustainability*, 2024, **16**, 4797.
58. M. S. Dhanoa, S. López, R. Sanderson, S. J. Lister, R. J. Barnes, J. L. Ellis and J. France, *Journal of Chemometrics*, 2023, **37**, e3511.
59. P. M. Fine, G. R. Cass and B. R. Simoneit, *Environ Sci Technol*, 2001, **35**, 2665-2675.
60. R. J. Sheesley, J. J. Schauer, M. Zheng and B. Wang, *Atmospheric Environment*, 2007, **41**, 9050-9063.
61. Y. Zhang, J. J. Schauer, Y. Zhang, L. Zeng, Y. Wei, Y. Liu and M. Shao, *Environmental Science & Technology*, 2017, **51**, 4734-4734.
62. M. d. E. Corporación de Desarrollo Tecnológico, *Medición del consumo nacional de leña y otros combustibles sólidos derivados de la madera*, Corporación de Desarrollo Tecnológico (CDT), 2015.
63. A. Schueftan, C. Aravena and R. Reyes, *Resource and Energy Economics*, 2021, **66**, 101265.
64. R. Reyes, A. Schueftan, C. Ruiz and A. D. González, *Energy Policy*, 2019, **124**, 301-311.

View Article Online
DOI: 10.1039/D5VA00495K

Data for this article, including long-term ambient PM2.5 concentration data and associated meteorological variables, are available at Zenodo at https://doi.org/10.5281/zenodo.18078493, with restricted access during peer review and open access upon article acceptance.