A supervised machine learning approach to classify traffic-derived PM sources based on their magnetic properties.

Sarah Letaïef, Pierre Camps, Claire Carvallo

A supervised Machine Learning approach to classify traffic-derived PM sources based on their magnetic properties.

Sarah Letaïef [a,*], *Pierre Camps* [a,], *and Claire Carvallo*[b]

[a] Géosciences Montpellier, Université de Montpellier, CNRS, Montpellier, France.

[b] Sorbonne Université, UMR 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, F-75005, Paris, France.

KEYWORDS

Magnetic properties, kNN classification, machine learning, traffic related PM.

ABSTRACT

Environmental magnetism techniques are increasingly used to map the deposition of particulate pollutants on any type of accumulative surfaces. The present study is part of a collective effort that begun in recent years to evaluate the efficiency of these techniques involving a large range of measurements to trace the source signals. Here we explore the possibilities provided by the very simple but robust k-near-neighbors algorithm (kNN) for classification in a source-to-sink approach. For this purpose, in a first phase, the magnetic properties of the traffic-related sources of particulate matter (tire, brake pads, exhaust pipes, etc.) are used to parameterize and train the model. Then, the magnetic parameters measured on accumulating surfaces exposed to a polluted air as urban plant leaves and passive filters are confronted to the model. The results are very encouraging. The algorithm predicts the dominant traffic-related sources for different kinds of accumulative surfaces. The model predictions are generally consistent according to the sampling locations. Its resolution seems adequate since different dominant sources could be identified within one street. We demonstrate the possibility to trace traffic-derived pollutants from sources to sinks based only on magnetic properties, and to eventually quantify their contributions in the total magnetic signal measured. Because magnetic mapping has a high-resolution efficiency, these results open the opportunity to complement conventional methods used to measure air quality and to improve the numerical models of pollutant dispersion.

1. INTRODUCTION

Despite various actions set up by governments and a record number of more than 6,000 cities in 117 countries that are now continuously monitoring air quality, nearly all the world's population (99%) still breathe air that exceeds the limits prescribed by the World Health Organization (World Health Organization, 2022). Indeed, the increase of urban

population (Department of Economic and Social Affairs, 2019) leads to an increase of human activities and consequently, an increase of various types of atmospheric pollutants emissions, including particulate matter (PM) and ultrafine particles (UFP <30 nm). Since the anthropogenic particulate matter often contains Fe-bearing nanoparticles (Gonet and Maher, 2019; Hofman et al., 2017; Hunt et al., 2013) magnetic monitoring of air pollution by measuring dust depositions on accumulative surfaces are more and more used over the past 20 years. As a matter of fact, this approach provides a high-resolution mapping of magnetic pollutants, including the UFPs (Hofman et al., 2017; Maher et al., 2008; Matzka and Maher, 1999) It can be very precise and consequently useful for citizens and policy makers if the survey is carried out at a micro-scale as was done for highway infrastructures (Letaïef et al., 2020) or canyon streets (Muxworthy et al., 2022) in order to pinpoint the major sources of pollution. Thus, it could be perceived as an alternative but also as a complement to help conventional air-monitoring by state-approved agencies. Magnetic measurements could be performed on various types of exposed accumulative surfaces: passive filters (Cao et al., 2015), plant leaves (Hofman et al., 2014; Letaïef et al., 2020; Sagnotti et al., 2009), lichens (Winkler et al., 2022, 2020) or moist paper swabs taken from television/computer screens to evaluate outdoor or indoor PM pollution (Maher et al., 2013). These accumulative surfaces are generally sampled after few months of polluted air exposure to collect enough dust to be measured. Therefore, this technique provides information on the chronic exposure, which is an important parameter in the public health issues. Environmental Scanning Electronic Microscopy images (Figure 1A) of plant leaves illustrate well the heterogeneity of dust deposition, in terms of nature and shape. This is in relation to the variety of PM sources (Figure 1B), which have been sorted based on the nesting emission stratification levels including long-range transport, regional and urban sources, right down to the street level (Kakosimos et al., 2010). Because many sources

exist, Letaïef et al. (2020) proposed to isolate the main source of the study area by expressing magnetic measurements as the relative change between the measurements of the accumulative surfaces concerned and several other measurements from an area not impacted by the targeted pollutant source. In urban areas, the main source is undoubtedly vehicular traffic (Gonet and Maher, 2019; Hama et al., 2017) via exhaust (Liati et al., 2012), wear products (Ingo et al., 2022; Kukutschová et al., 2011; Rahimi et al., 2021) from vehicle and pavement (Mathissen et al., 2011), or road dust resuspension (Vlasov et al., 2022; Yang et al., 2016) induced by moving vehicles. Because not all the pollutants are specifically magnetic, it is essential to understand what is responsible for creating the magnetic signal measured on the accumulative surfaces.

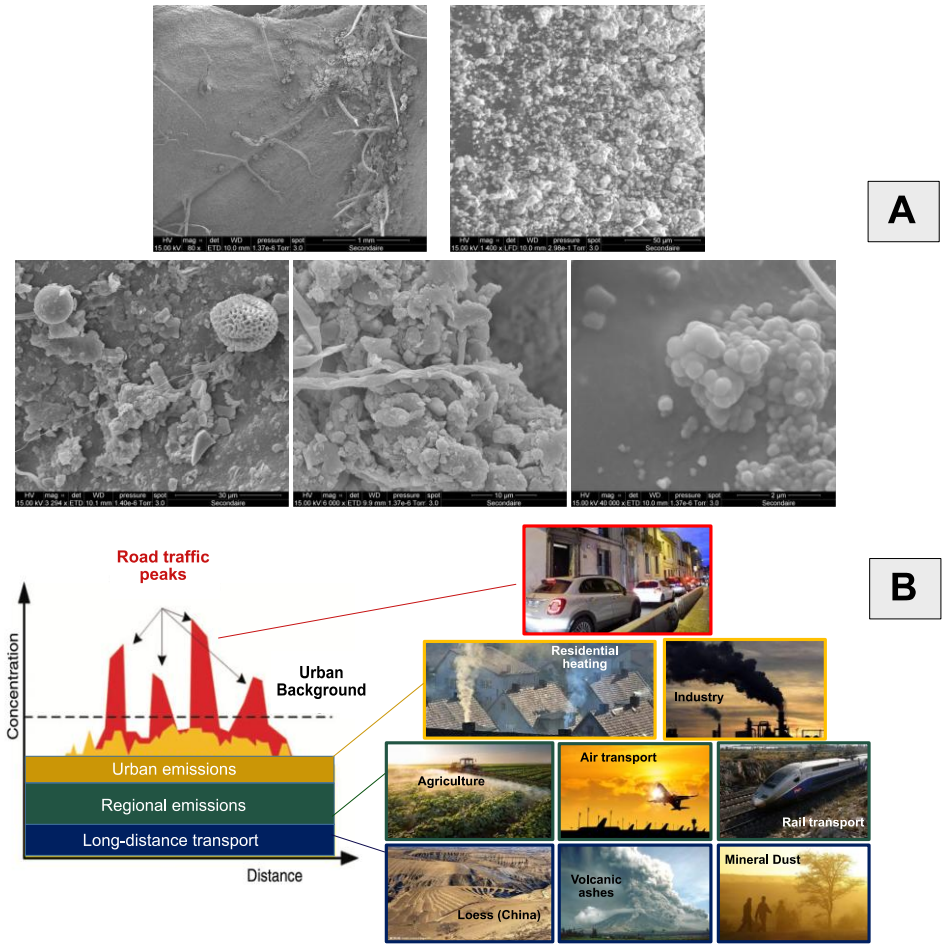

Figure 1. Illustrations for (A) Natural and anthropogenic deposits accumulated on plant leaves and observed on SEM images at different scales from 1 mm to 2 μm. These images were taken at the microscopy department of the University of Montpellier with a Quanta 200 Scanning Electron Microscope (SEM) (B) the stratification levels of PM sources in street canyon. The bottom left diagram is redrawn from (Kakosimos et al., 2010).

One question that emerges from these results is whether environmental magnetism techniques can provide more refined knowledge about the nature and the source of pollutants. This is a critical information that is needed to design efficient mitigation policies to increase air quality in urban environments. For instance, for a given location, do traffic-related particulate emissions come from the braking system or from exhaust fumes? To this date, this level of analysis with only magnetic parameters has not been investigated. Therefore, this study aims to classify PM depositions on plant leaves or passive filters in one to the five PM traffic-related sources categories by using a machine learning algorithm only parameterized with magnetic parameters. This would provide an idea of the magnetic method robustness, when applied to air quality monitoring. A similar work was done in a recent study (Colangeli et al., 2022), which classify PM emissions as anthropogenic or natural by means of deep-machine learning technique on PM grain size.

In this study, we use a machine learning approach to generate an algorithm with labeled data (i.e., magnetic data from known pollutant sources) aiming to establish from a pattern recognition a set of rules that composes a model. Unlabeled data (i.e., magnetic data from unknown pollutant sources) are then passed into the model to infer the predicted source. In a first part, we will present the labeled and unlabeled samples used for the training and the prediction phases, respectively. Then, the magnetic parameters given to the algorithm are presented with details on how they were obtained from the reduction of

the raw data. Finally, the machine learning algorithm is described and the results and predictions are discussed.

2. MATERIALS

2.1 Labeled samples for the model training phase. Recent studies highlight that PM emitted through combustion engines (exhaust emissions) is not the only source of the PM from vehicular origin, but the particles generated from abrasive and fatigue wear processes also play an important role and must be taken into account (Charron et al., 2019; Grange et al., 2021). For this purpose, a total of 16 samples related to traffic sources were collected and classified according to one of the five identified categories of source (brake pads, exhaust emissions, tire-fatigue wears, resuspension and mineral dust). The exhaust fumes category comprises three diesel, and one gasoline combustion residues, all collected by removing the powder inside the tailpipes with a plastic spatula. The wear source category comprises abrasive, and surface-fatigue wear products. The abrasive wear products include six worn-out brake pads obtained from a mechanic shop, two brake powder samples (BrD1 and BrD2) collected from two different cars by removing and scraping the brake powder around the wheel rim, brake disks and pads, and a fresh asphalt concrete sample recovered from a construction site in Montpellier, France. The surface-fatigue wear product is represented by one tire thread collected by removing the surface layer of a worn-out tire. Lastly, resuspension category comprises a mineral dust sample collected after a Sahara wet-deposition episode by wiping car windshields and garden tables, and one street dust sample collected with a portable vacuum cleaner in a street canyon located in the city center of Montpellier, France, where the traffic is 7000 vehicles/day.

2.2 Unlabeled samples for the model prediction phase. To test the model after the training period of the machine learning algorithm, we used a variety of accumulative surface

samples from different urban areas to estimate the main category source. First, plant leaves from *Cistus monspeliensis* trees were collected in January 2020, around 1m-height on vegetated earth-berms along a dual 6-lanes motorway located in south-east of Montpellier, France (Letaïef et al., 2020). Then, five passive filters (B-010, B-013, B-015, B-027 and B-028) were deployed on the first and second floor balconies in a busy street canyon (around 7000 vehicles/day) during a period of 3 months (from December 2020 to March 2021). Lastly, we used two certified reference material dedicated to chemical analysis in isotopic, trace and major elements contents: BCR-723 (Institute for Reference Materials and Measurements; IRMM) is a certified and homogenized road dust material with a particle size of less than 90 μm; ERM-CZ120 (IRMM) is a certified fine dust sample which was processed in a way to resemble PM_{10} as close as possible.

3. METHODS AND BACKGROUND

The choice of data type is the central issue that must be addressed in a well-thought-out manner for a machine learning approach. Basically, in the environmental magnetism approach applied to monitor air quality, we aim at quantifying Fe-bearing PM. To do this, we measure a set of magnetic parameters derived from environmental magnetism techniques in order to provide information about the magnetic concentration, nature, and size of the sample. In our study, these magnetic parameters, summarized in Table 1, were also complemented by concentrations of chemical elements (Table 1).

Table 1: Summary of magnetic and chemical experiments performed to obtain the parameters which are included in the machine learning algorithm.

Magnetic experiments	*Magnetic parameters*
Hysteresis loops	• M_{rs}/M_s • B_{cr}/B_c
First order reversal curves (FORCs)	• Marginal distribution ($\rho_c B_c$) • Coercive field ($B_{c,max}$) • Vertical distribution (ρ_u B_u) • Bias field ($B_{u,max}$)
Unmixing of IRM acquisition curves	• Number of components • Mean field ($B_{1/2}$) of the two main components • Variance (DP) of the two main components • Contribution (height) of the two main components
Temperature dependency of susceptibility (kT curves)	• A_{max} • A_m • A_{cr} • Presence of the Verwey transition • Maximal Curie temperature (for samples which have more than one Curie temperatures)

Chemical experiments	*Elemental concentrations*
XRF	• **[Zn]** • **[Fe]** • **[Ba]**

3.1 Magnetic parameters.

3.1.1 Hysteresis parameters. Hysteresis loops were performed on samples tightly packed in gelatin capsules with a Vibrating Sample Magnetometer (VSM) from Princeton Measurements Corporation, which is part of the Mineral Magnetism Analysis Platform of Institut de Physique du Globe de Paris and Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie. Magnetic hysteresis curves, i.e., the variation of the magnetization with an applied field, were measured in a maximum applied field up to 1.8 T (Figure 2). The data obtained from this technique are routinely reduced to four main parameters. The saturation magnetization M_s, the remanent magnetization M_{rs}, and the coercive field B_c are obtained

directly from the main hysteresis loop. The coercivity of remanence B_{cr} is obtained with the backfield curve, which monitors the decrease of the remanent magnetization after saturation with an increasingly negative laboratory field. The values of the hysteresis parameters derived from the major hysteresis loop were linearly corrected from diamagnetism and paramagnetism effects. Finally, these parameters are commonly used as ratios: M_{rs}/M_s and B_{cr}/B_c, and plotted on a two-variable scatter (Day) plot (Day et al., 1977). Data trends on Day plot are often interpreted in terms of grain size or domain state variations (Dunlop, 2002). However, these ratios measure an average of the magnetic properties of all the particles, which is especially true for dust samples containing a complex mixture of magnetic minerals with different grain sizes (Heslop and Roberts, 2012a, 2012b; Roberts et al., 2014).

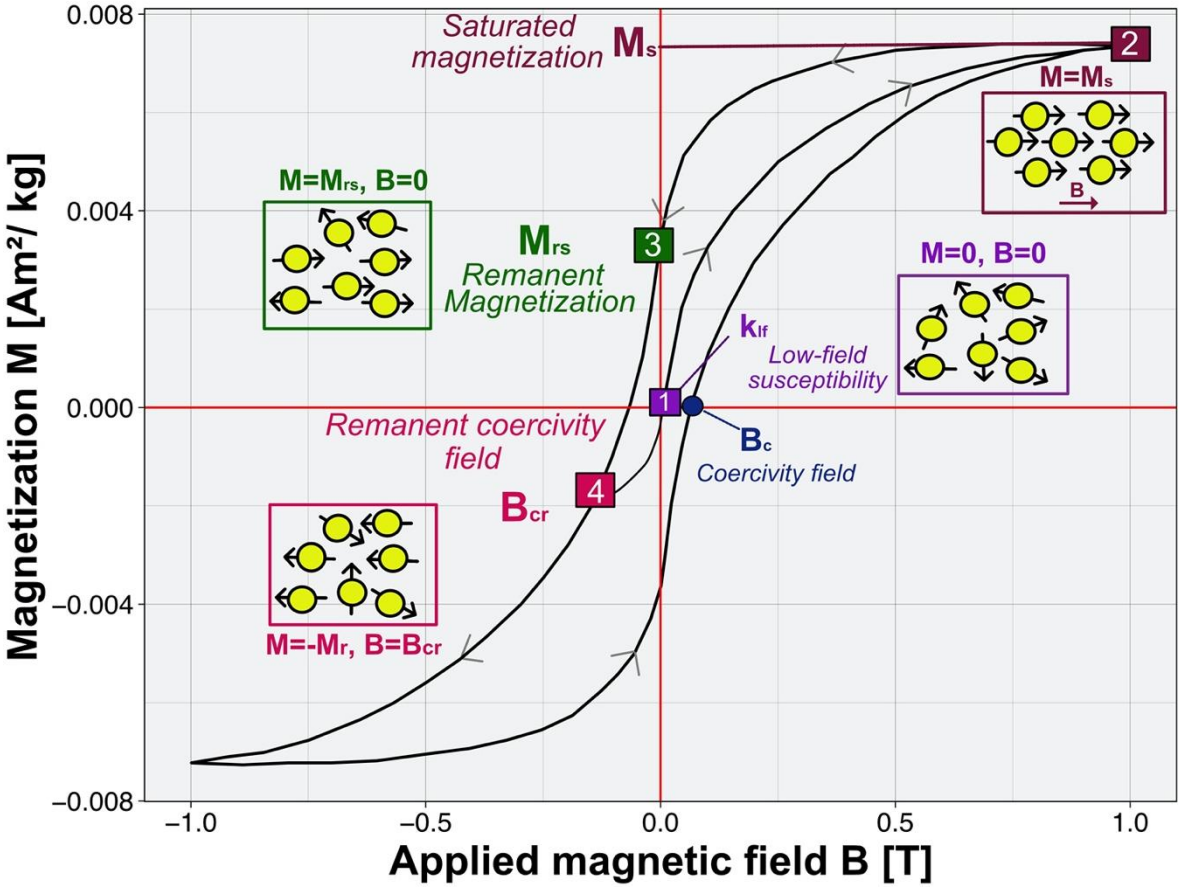

Figure 2. Schematic example of a hysteresis cycle illustrating the magnetization stages of a random assembly of single domain magnetic grains and its resulting parameters (M_s, M_{rs}, B_c and B_{cr}).

3.1.2 FORC diagrams. First Order Reversal Curves (FORC(Pike et al., 1999; Roberts et al., 2000) diagrams provide a map of the magnetic response of all particles in a sample with irreversible magnetizations. FORC diagrams have become a useful tool in the rock magnetism field because under some circumstances, they allow to estimate the coercivity field distribution and the magnetic interactions between particles within a sample. A FORC function $\rho(B_c, B_u)$ from a weakly interacting and nonthermally activated ensemble of particles has a finite vertical spreading $\rho_u(B_u)$ and can be physically interpreted by its marginal distribution $\rho_c(B_c)$ (Egli, 2006) defined by:

$$\rho_c(B_c) = \int_{-\infty}^{\infty} \rho(B_c, B_u) dB_u$$

This equation describes the total magnetic contribution of all particles following a switching field B_c.(Winklhofer and Zimanyi, 2006) It also corresponds to the effective coercivity distribution of the particles. Therefore, we reduce the FORC distributions to their vertical and marginal distributions (Figure 3), all calculated with the same smoothing factor SF (vertical and central ridge SF = 10, horizontal and vertical SF = 15) with FORCinel software (Egli, 2006) The vertical distribution ρ_u (B_u) is simply defined as the vertical profile of the FORC distribution that passes through the FORC peak.

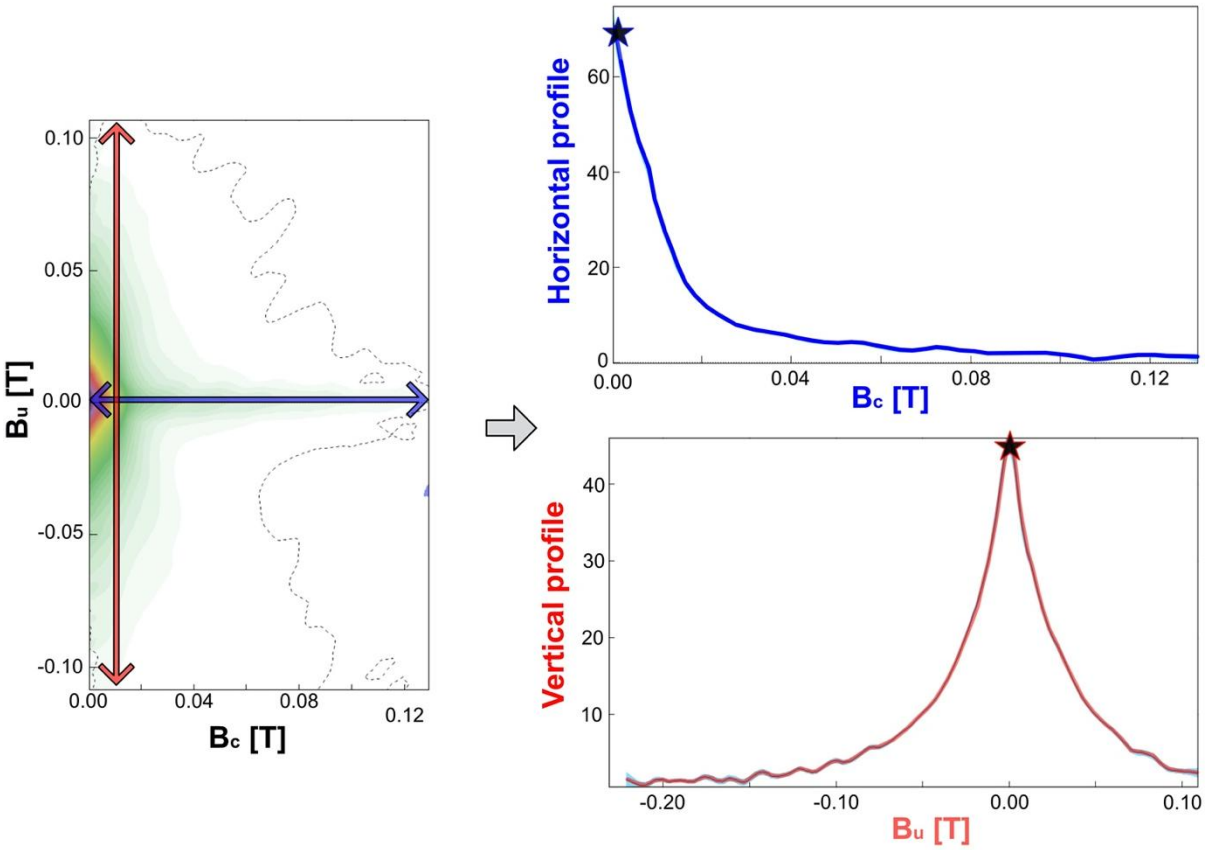

Figure 3. Visualization of marginal and vertical distributions calculated from a FORC diagrams of several source samples. The red and the blue stars illustrate points from coordinates: $B_{u,max}$; ρ_u B_u and $B_{c,max}$; ρ_c B_c, respectively, which were incorporated in the machine learning algorithm.

3.1.3 Distribution of coercive fields. Another way of discriminating between different magnetic components within a sample is deconvoluting the cumulative Isothermal Remanent Magnetization (IRM) curves by means of unmixing techniques (Figure 4a; Egli, 2004; Heslop et al., 2002; Kruiver et al., 2001; Robertson and France, 1994) Before IRM is acquired, all the samples were demagnetized with a 170 mT alternating field and then measured in zero field. This measurement was used to set the baseline. For brake pad samples, stepwise IRMs were manually imparted to the sample with an impulse magnetizer ASC Model IM-10-30 in 30 increments ranging from 1 up to 2800 mT. The magnetization steps were specifically chosen to be approximately equidistant in a log-scale. After waiting for 24h to prevent any viscous effect, their magnetic moments were measured with an AGICO

JR5A spinner magnetometer. For these samples, IRM acquisition and measurement were carried out at the Géosciences Montpellier laboratory. The IRM acquisition curves for the other samples were directly imparted and measured with the VSM at IPGP. Subsequently, unmixing analyses performed on IRM acquisition curves allow to describe the distribution of the coercive fields within a sample, by assuming that the distribution of the coercive fields for each magnetic component present in a mixture sample is a log-normal distribution. Here, the cumulative log-normal (CLN) functions are parameterized(Robertson and France, 1994) with: (1) the height of the CLN function corresponding to the saturation of the IRM (SIRM); (2) the mean of the applied field at which half the SIRM is acquired ($B_{1/2}$), corresponding also to the average remanence coercivity (B_{cr}) and, (3) the standard deviation of the log-distribution, also named the dispersion parameter (DP). Finally, inference about the parameters for each component was performed by means of a Bayesian approach. The prior specifications in our modeling are the following: (i) the components of the coercivity distribution follow normal distributions on the means (μ_k) after logarithmic transformation of the applied field, (ii) a Dirichlet distribution is assigned for the heights of the CLN, (iii) an inverse gamma function is assigned for the standard deviations, and (iv) a normal distribution is assigned for the conjugate prior ($\mu_k \mid \sigma_k$). Initial values such as the number of components, their means, proportions and variances are fixed arbitrarily before the modeling. We calculated the k normal distributions sum and their parameters (Figure 4b) by randomly selecting the log-normal distribution values in 3 Monte-Carlo Markov chains (MCMC) by means of the Gibbs sampler. The confrontation of these a posteriori laws with the data allows a readjustment of initial parameters until the chains converge.

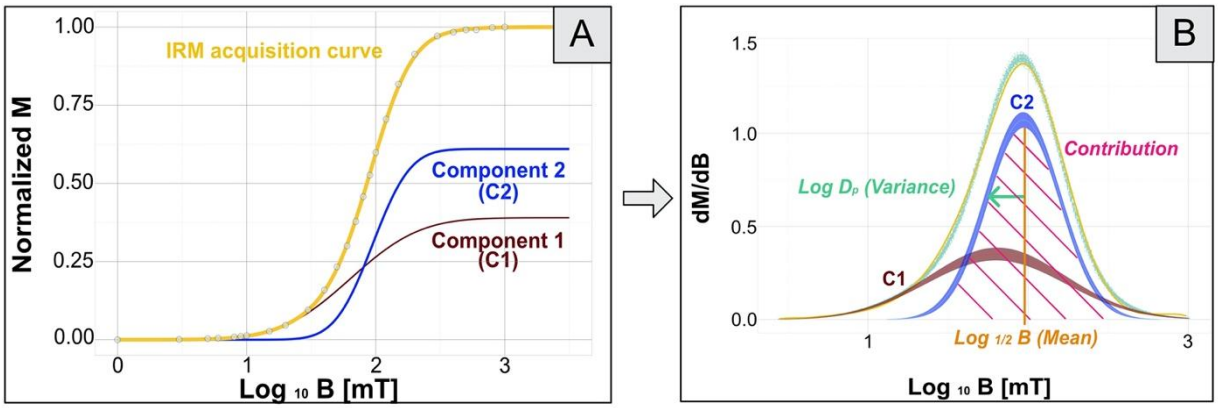

Figure 4. Example of IRM curve unmixing performed on one source sample: [A] linear acquisition plot (LAP) illustrating the two components needed to explain the IRM acquisition curve; [B] posterior densities observed from the last 200 Gibbs iterations when MCMC algorithm has converged.

3.1.4 Temperature dependency of initial susceptibility. The temperature variations of bulk magnetic susceptibility (k_0-T curves) depend on the magnetic mineralogy, and it is therefore a useful characterization tool (Dunlop, 2014). To perform the measurements, samples were crushed as powder and placed in a non-magnetic silicon glass measuring tube. This technique provides a continuous measurement of susceptibility, which is measured with the Agico KLY-3S Kappabridge, while the sample is heated from -194°C to 700°C and subsequently cooled down to room temperature. This measurement allows to identify specific Curie temperatures of magnetic minerals but also their possible alteration during heating and cooling processes. When a large set of specimens are processed and need to be compared, it is useful to characterize the alteration of the magnetic fraction with numerical indices calculated from the curves (Hrouda, 2003). In our case, five parameters derived from the k_0-T curves (Figure 5) will be used for the machine learning:

(1) The highest Curie temperature (T_C), determined when the second derivative of k_0-T curve is zero (Prévot et al., 1983). We arbitrarily chose to take the highest Curie temperature since not all the samples display several Curie temperatures.

(2) The Verwey transition, a drastic increase of magnetic susceptibility found around -152°C that is distinctive of the magnetite presence.

(3) The mean alteration index (A_m; Hrouda, 2003) which quantifies the susceptibility differences between the heating and the cooling phases. This index is calculated for *i* temperatures as:

$$A_m = \frac{100}{NK_{40°C}} \sum_{i=1}^{N} (k_i - K_i)$$

Where $K_{40°C}$ is the susceptibility at 40°C during the heating, the k_i and K_i pairs are obtained from susceptibilities measured in the cooling and the heating phases, respectively, at a regular step of 1°C by means of a linear interpolation of the cubic splines smoothing curves. N is the number of pairs considered. The positive index indicates higher cooling than heating susceptibilities while the negative index indicates the opposite.

(4) The maximum alteration index (A_{max}; Hrouda, 2003) which corresponds to the maximum difference between the heating and the cooling curves referred to $(k\text{-}K)_{max.}$ Thus, A_{max} is calculated as:

$$A_{max} = 100 * \frac{(k - K)_{max}}{K_{40°C}}$$

Where *k* and *K* are the susceptibilities measured at the same temperature during the cooling and the heating phases, respectively, and $K_{40°C}$ is the susceptibility at 40°C obtained from the heating curve.

(5) The crossed index (A_{cr}; Hrouda, 2003) was established to indicate when the heating and cooling curves cross each other:

$$A_{cr} = \frac{\sum_{i=1}^{N}(k_i - K_i)}{\sum_{i=1}^{N}(|k_i| - |K_i|)}$$

Where $|K_i|$ and $|k_i|$ are the absolute values of K_i and k_i. A_{cr} =1 when the heating and cooling curves do not cross each other, $|A_{cr}| < 1$ when they cross each other while a prevailing part of the heating curve shows higher susceptibility than the cooling curve, and $|A_{cr}| > 1$ when the opposite is true.

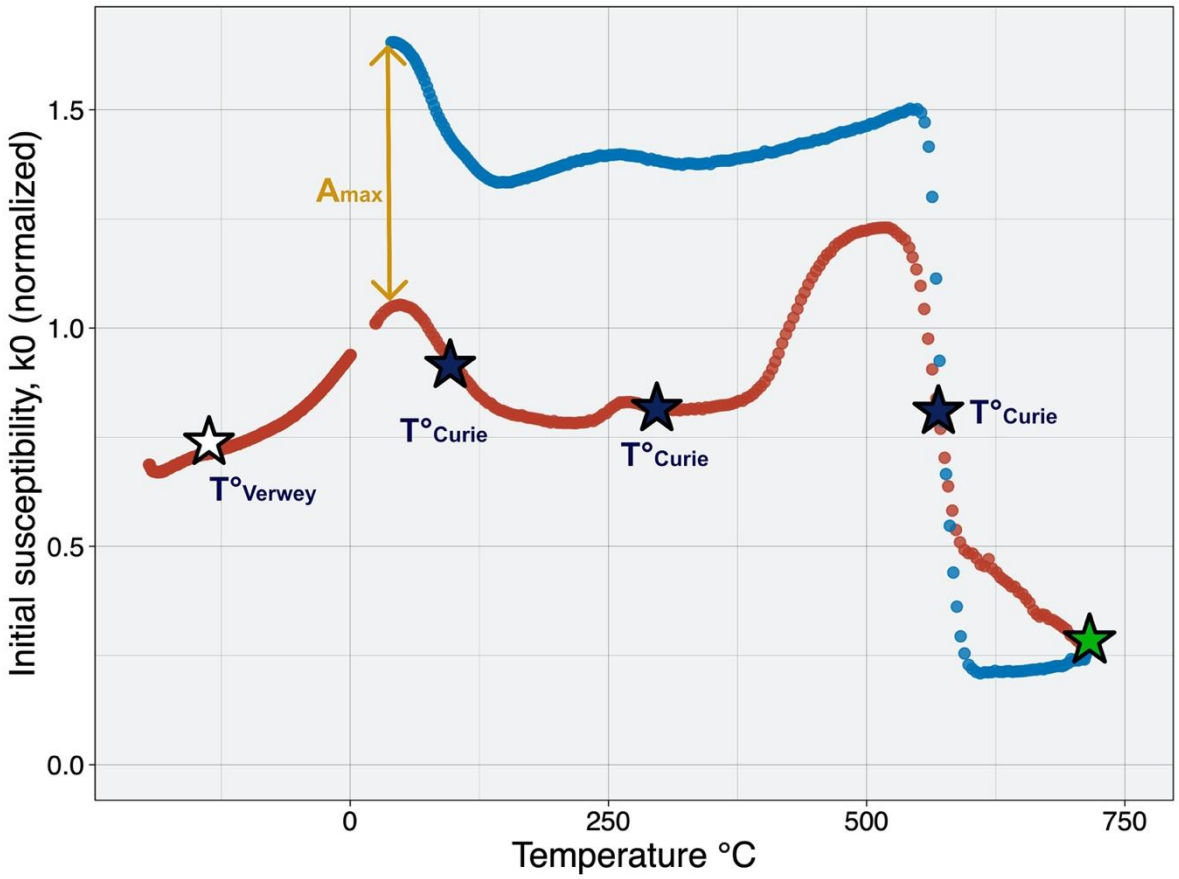

Figure 5. Illustration of alteration parameters (Hrouda, 2003) from a street dust sample. The green star represents a non-reached Curie Temperature above 700°C. The susceptibility is normalized with the value of the susceptibility measured at room temperature ($K_{0,20°C}$).

3.2 Elemental analysis. Chemical elemental analyses were performed with a Niton XL3t-900 Goldd X-ray fluorescence (XRF) portable analyzer (Thermo Scientific®, Waltham, MA, USA) on ex-situ samples reduced to powder after drying. Before analysis, sample holders

were measured empty to discard the ones that showed a strong signal. The interpretation of XRF spectra was processed in two steps: first, we assessed the presence of an element from a qualitative visual inspection of spectrum and then calculated its concentration with the NITON Data Transfer software provided by Thermo Electron CorporationTM, Waltham, MA, USA. Results were obtained in μg/g (ppm). For the purpose of this study, we used concentrations in zinc (Zn), iron (Fe), and barium (Ba), as these elements are known to discriminate of traffic-related PM sources.

3.3 Supervised-Classification Machine Learning: the model algorithm, tuning and accuracy. A machine learning algorithm is described as supervised when the user guides the learner. In our case, the aim is to classify data in several categories based on their similarities. We performed this classification by means of a k-nearest neighbor (kNN) algorithm included in the *mlr* R library.(Rhys, 2020) This is a two-phase process: (1) The training phase, during which the discriminating magnetic parameters for the different categories of sources are inserted. We first tell the algorithm which source belongs to which category (see section 2.1) and call them labeled data. Then, the model trains by progressively learning the data trends and rules. (2) The prediction phase, where the unlabeled data (the accumulative surface samples) are put into the algorithm. The algorithm will classify the unlabeled data between their nearest and their least similar neighbors by calculating Euclidian distances between the labeled and the new unlabeled data.

To see how the model performs, we used the performance () function to compare the predicted classes by the model with the true ones and returns the performance metrics. The latter indicating how well the predicted and true classes match to each other. This function returns a list of arguments called *measures* composed mainly of (1) the *mean*

misclassification error or the *mmce* and, (2) the *acc*, or *accuracy*, both a range comprises between 0 and 1. MMCE represents the proportion of samples which are misclassified in another class than their true class. On the contrary, accuracy is the proportion samples well classified.

Because a model will perform better when tested with data from which it was already trained on, we performed *k-fold cross validation* to verify the accuracy of the model and avoid overfitting or underfitting. During this process, the dataset is randomly split into approximately equal-sized sub datasets called folds. One of the folds is put aside as a test set and the remaining data are used as the training set. Then, we iteratively pass the test set through the model with different folds until all the folds have been used once as the test set and we make a record of the relevant performance metrics.

4. RESULTS AND DISCUSSION

4.1 Model optimization. In the kNN algorithm, *k* (or the number of near neighbors) is not considered as a parameter because it is not directly estimated from data by the algorithm itself. However, *k* is known as a hyperparameter: it controls how the model will make predictions and it is up to the user to choose its best value. This can be done either manually or automatically by a procedure called hyperparameter tuning. During this process, the algorithm tries different values of *k* between 1 and 10 and finds the best-performing value of *k* (Figure 6). In our case, the best estimated value of *k* is always 1, which implies that the model will take the nearest neighbor category as the sample predicted category. Finally, after including the hyperparameter tuning in the cross-validation process, the model accuracy is estimated by 99.62 %, which basically represents the proportion of cases that were correctly classified by the model.

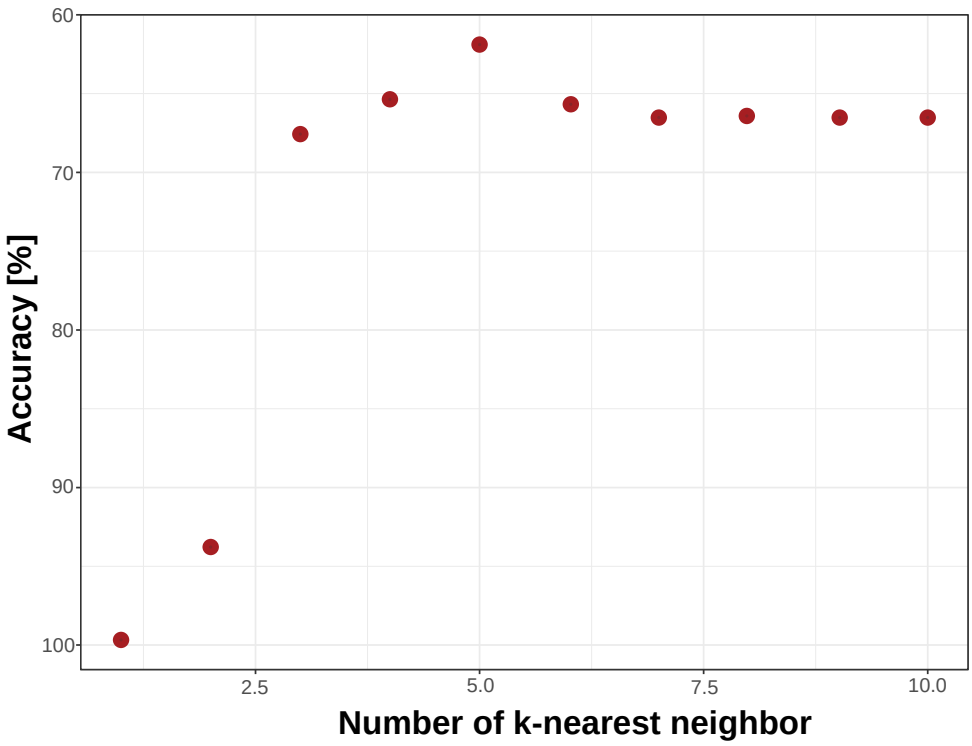

Figure 6. Model accuracy estimated with different values of k-nearest neighbor. For this study, the best k value implying the best model performance is 1.

4.2 Testing the model with unlabeled source data. Some source samples (one sample of gasoline, diesel, BrD1, BrD2 and, asphalt concrete) could not be measured with all techniques for various reasons such as the lack of sample amount availability, technical difficulties, or a magnetic signal that was too low. Therefore, we used them as "fake source samples" by mixing their complementary magnetic properties between two samples known to be part of the same source category. Five "fake" samples were made: Mix1 and Mix2 are composed of a mix between BrD1 and BrD2; Mix 3 and Mix 4 are made of a mixture with one diesel and one gasoline exhaust pipe sample; and Mix 5 is a mixture of street dust and asphalt concrete samples. Since we know the category of each mixture sample, this offers the opportunity to test the model on data that was unknown to it, even during the training phase. Results are presented on Figure 7. Ultimately, the model predicted categories that are the expected ones

for Mix 1, Mix2, Mix3 and Mix 4, which is very encouraging. However, the classification for Mix 5 came out as “mineral dust” instead of resuspension.

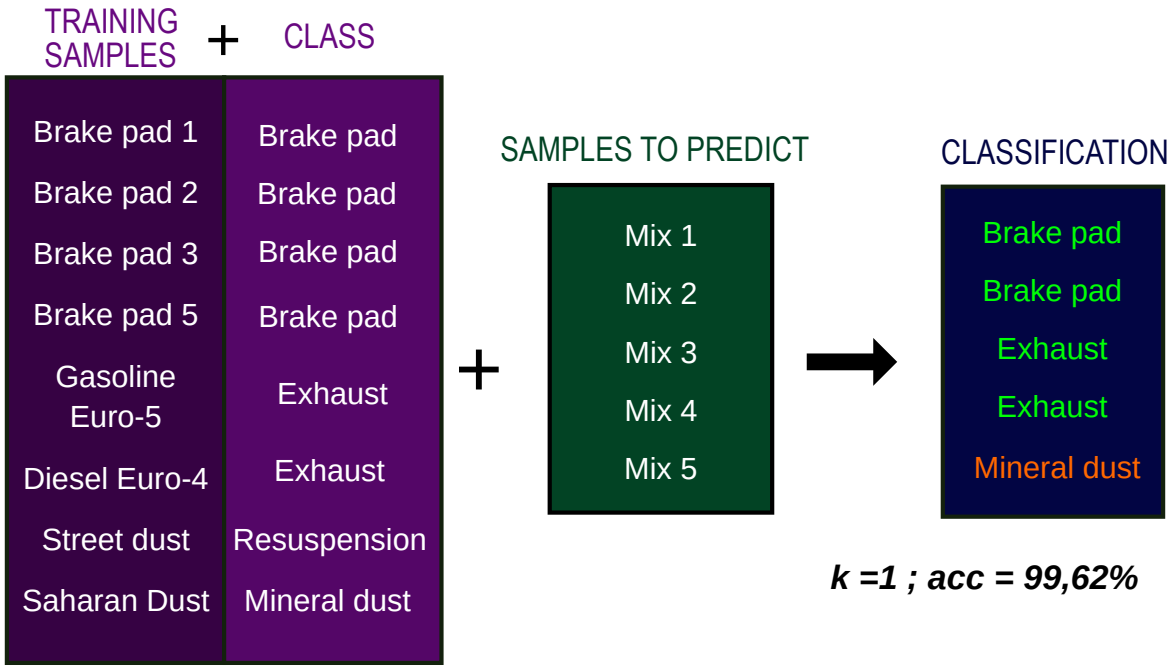

Figure 7. Model testing performed on mixed samples that the model did not train with and for which the classifications are known. Classifications in green represent a good prediction while the one in orange illustrates an inaccurate prediction.

4.3 Estimation of the main source on accumulative surface samples. In this application of the model, we used all the sources and the Mix samples, which were predicted before to train the model. We asked the model to predict two types of accumulative surfaces (plant leaves and passive filters) and the two certified standards (ERM-CZ120 and BCR-723) usually used for isotopic chemical analysis in urban PM pollution. The tree leaf samples collected from the motorway are identified as having a magnetic signal with a mineral dust origin (Figure 8). This result is consistent, considering that the leaves were sampled along a motorway on a flat-top earth berm(Letaïef et al., 2020). Indeed, a magnetic mapping was performed in the study area and showed very low magnetic signals due to the high dispersion potential for traffic pollutants of these earth berms (Letaïef et al., 2020)

ERM-CZ120 and BCR-723 are identified as coming from the resuspension category. Giving the nature of these certified standards (PM_{10} and road dust) and the fact that the resuspension class is composed of street dust sample and a mixture between street dust and asphalt samples (Mix 4), the results from the model are also consistent.

The results for the five passive filters deployed in the street canyon located in the city center of Montpellier show different source categories. B-010 and B-015 are predicted to come from to the exhaust class. This is consistent with the fact that these filters were both deployed at the beginning of Saint-Louis Street, which is 245 m long, and consists of one single traffic lane traveled by about 7,000 vehicles daily. Filters B-013 and B-028 are predicted to have a resuspension origin. This is consistent for sample B-013, which was placed in the street adjacent to Saint-Louis street. However, the prediction for filter B-028, which was deployed at the ground floor inside the street canyon, is more surprising. Finally, B-027 is the only filter which was predicted as a brake pad source. This is also consistent considering that the filter measure station was placed at the end of the street, which ends with a traffic light, and where vehicles are therefore likely to brake and stop.

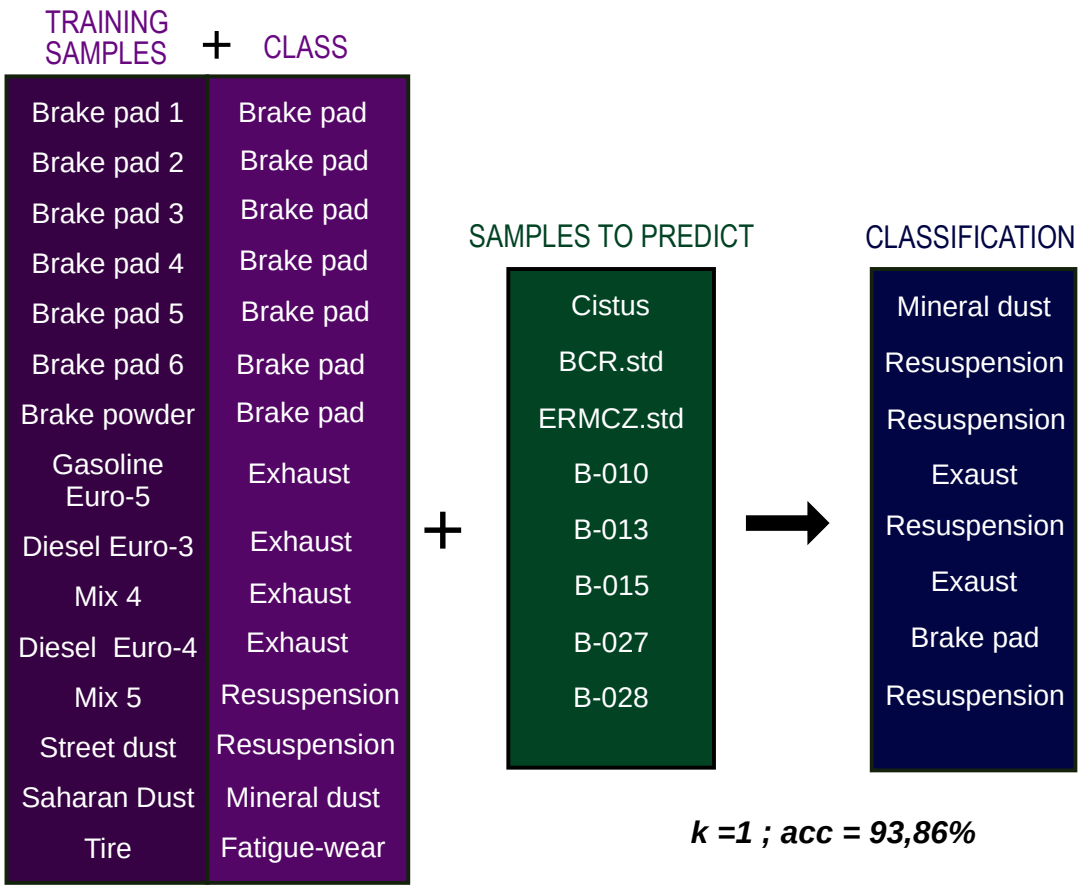

Figure 8. Model results for the main source predicted on accumulative surfaces and two certified chemical standards.

4.4 Conclusion and Perspectives. Magnetic mapping has been widely used during the last decades to characterize air pollution (Hofman et al., 2017) and has proven to be very accurate at high resolution (Letaïef et al., 2020). This is an important improvement compared to numerical modeling of dispersion and prediction of pollutants usually performed with low resolution data coming from monitoring stations that are too few in comparison with the scale of the city. However, achieving a real representativeness in the source of magnetic signal in urban PM remains challenging since there are many types of traffic-related PM sources and since samples from the same kind of source might display a large heterogeneity.

This is why we chose the machine learning approach, which is an innovative approach, and had never been tested before in the environmental magnetism field. Thanks to a classification algorithm fully parameterized with magnetic data, we were able to characterize the different PM traffic-related sources and PM accumulative surfaces.

The results turned out to be very promising and consistent given the locations of the samples for which the PM source is predicted by the model. The machine learning algorithm trained with the magnetic parameters was able to differentiate the origin of the magnetic signal of accumulative surfaces, as it was also the case for the passive filters deployed in the street canyon and can be used for a source-to-sink purpose.

However, the magnetic parameters classification by using a kNN algorithm also shows some limitations. For further developments, it appears necessary to:

- Extend the dataset since the samples are usually heterogenous. Indeed, it would be desirable to integrate more samples to obtain a better representativeness. In addition, the dataset seems not large enough to perform other classification algorithms as the logistic-regression one, which would be useful to compare both algorithms and to conclude on the best method to use.
- Define more precisely the discriminant power of each of the magnetic parameters because the prediction accuracy is highly impacted by noisy data or outliers (Rhys, 2020).

Finally, it appears that we need to test more sophisticated algorithms which provide a way to include non-available values for some parameters (*NA*). For now, the model is not able to deal with missing data. This is unfortunate because it is sometimes not possible to perform all the desired magnetic measurements on a particular sample for various reasons. In these cases, the

whole set of measurements on that sample must be removed from the data set. In addition, we need a possibility to predict more than one category per sample to quantify the contribution from each category and to refine the result. This would allow to not only predict the main source, but also estimate the contribution of all the defined sources involved in the measured magnetic signal. Finally, to truly evaluate the efficiency of machine learning technique, it would be very interesting to compare our machine learning classification based on magnetic properties with isotopic chemistry analysis or other machine learning algorithms.

To conclude, our study demonstrates the great potential of artificial intelligence applied to the study of air quality and parameterized with magnetic properties. It helps to better understand the physical provenance of PM detected with magnetic methods, which will, in a first step, be used as a complement of the conventional measurements for the parametrization of the numerical models, but could also, on a longer term, be useful to public deciders involved in urban PM mitigation.

AUTHOR INFORMATIONS

Corresponding authors

Sarah Letaïef
Géosciences Montpellier, University Montpellier and CNRS, Montpellier, F-34095, France.
Email : letaief-sarah@laposte.net
ORCID : https://orcid.org/0000-0002-8057-1414
Authors

Pierre Camps
Géosciences Montpellier, University Montpellier and CNRS, Montpellier, F-34095, France.
Phone : +33 (0) 4 67 14 39 38
Email: pierre.camps@cnrs.fr
ORCID: https://orcid.org/0000-0002-6637-4342

Claire Carvallo,

Sorbonne Université, UMR 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, F-75005, Paris, France.
Email : claire.carvallo@sorbonne-universite.fr
ORCID : https://orcid.org/0000-0002-0986-3203

CRediT authorship contribution statement

S. Letaïef: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing.

P. Camps: Conceptualization, Supervision, Resources, Project administration, Funding acquisition, Writing – review & editing.

C. Carvallo: Formal analysis, Investigation, Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available on Zenodo repository: https://doi.org/10.5281/zenodo.7198329

Acknowledgments

This research was funded by the Agence Nationale de la Recherche (ANR) grant numbers ANR-19-CE04-0008. We thank Mathilda de Villiers-Best for help with the English.

REFERENCES

Cao, L., Appel, E., Hu, S., Ma, M., 2015. An economic passive sampling method to detect particulate pollutants using magnetic measurements. Environ. Pollut. 205, 97–102. https://doi.org/10.1016/j.envpol.2015.05.019

Charron, A., Polo-Rehn, L., Besombes, J.-L., Golly, B., Buisson, C., Chanut, H., Marchand, N., Guillaud, G., Jaffrezo, J.-L., 2019. Identification and quantification of particulate tracers of exhaust and non-exhaust vehicle emissions. Atmos. Chem. Phys. 19, 5187–5207. https://doi.org/10.5194/acp-19-5187-2019

Colangeli, C., Palermi, S., Bianco, S., Aruffo, E., Chiacchiaretta, P., Di Carlo, P., 2022. The Relationship between PM2.5 and PM10 in Central Italy: Application of Machine Learning Model to Segregate Anthropogenic from Natural Sources. Atmosphere (Basel). 13, 484. https://doi.org/10.3390/atmos13030484

Day, R., Fuller, M., Schmidt, V.A., 1977. Hysteresis properties of titanomagnetites: Grain-size and compositional dependence. Phys. Earth Planet. Inter. 13, 260–267. https://doi.org/10.1016/0031-9201(77)90108-X

Department of Economic and Social Affairs, 2019. World Urbanization Prospects: The 2018 Revision, World Urbanization Prospects: The 2018 Revision. UN, New York. https://doi.org/10.18356/b9e995fe-en

Dunlop, D.J., 2014. High-temperature susceptibility of magnetite: a new pseudo-single-domain effect. Geophys. J. Int. 199, 707–716. https://doi.org/10.1093/gji/ggu247

Dunlop, D.J., 2002. Theory and application of the Day plot (M rs / M s versus H cr / H c) 2. Application to data for rocks, sediments, and soils. J. Geophys. Res. 107, 2057. https://doi.org/10.1029/2001JB000487

Egli, R., 2006. Theoretical aspects of dipolar interactions and their appearance in first-order reversal curves of thermally activated single-domain particles. J. Geophys. Res. Solid Earth 111, n/a-n/a. https://doi.org/10.1029/2006JB004567

Egli, R., 2004. Characterization of Individual Rock Magnetic Components by Analysis of Remanence Curves, 1. Unmixing Natural Sediments. Stud. Geophys. Geod. 48, 391–446.

https://doi.org/10.1023/B:SGEG.0000020839.45304.6d

Gonet, T., Maher, B.A., 2019. Airborne, Vehicle-Derived Fe-Bearing Nanoparticles in the Urban Environment: A Review. Environ. Sci. Technol. 53, 9970–9991. https://doi.org/10.1021/acs.est.9b01505

Grange, S.K., Fischer, A., Zellweger, C., Alastuey, A., Querol, X., Jaffrezo, J.L., Weber, S., Uzu, G., Hueglin, C., 2021. Switzerland's PM10 and PM2.5 environmental increments show the importance of non-exhaust emissions. Atmos. Environ. X 12. https://doi.org/10.1016/j.aeaoa.2021.100145

Hama, S.M.L., Cordell, R.L., Monks, P.S., 2017. Quantifying primary and secondary source contributions to ultrafine particles in the UK urban background. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2017.07.013

Heslop, D., Dekkers, M.J., Kruiver, P.P., Van Oorschot, I.H.M., 2002. Analysis of isothermal remanent magnetization acquisition curves using the expectation-maximization algorithm. Geophys. J. Int. 148, 58–64. https://doi.org/10.1046/j.0956-540x.2001.01558.x

Heslop, D., Roberts, A.P., 2012a. Estimating best fit binary mixing lines in the Day plot. J. Geophys. Res. Solid Earth 117, n/a-n/a. https://doi.org/10.1029/2011JB008787

Heslop, D., Roberts, A.P., 2012b. A method for unmixing magnetic hysteresis loops. J. Geophys. Res. Solid Earth 117. https://doi.org/10.1029/2011JB008859

Hofman, J., Maher, B.A., Muxworthy, A.R., Wuyts, K., Castanheiro, A., Samson, R., 2017. Biomagnetic Monitoring of Atmospheric Pollution: A Review of Magnetic Signatures from Biological Sensors. Environ. Sci. Technol. 51, 6648–6664. https://doi.org/10.1021/acs.est.7b00832

Hofman, J., Wuyts, K., Van Wittenberghe, S., Brackx, M., Samson, R., 2014. On the link

between biomagnetic monitoring and leaf-deposited dust load of urban trees: Relationships and spatial variability of different particle size fractions. Environ. Pollut. 189, 63–72. https://doi.org/10.1016/j.envpol.2014.02.020

Hrouda, F., 2003. Indices for numerical characterization of the alteration processes of magnetic minerals taking place during investigation of temperature variation of magnetic susceptibility. Stud. Geophys. Geod. 47, 847–861. https://doi.org/10.1023/A:1026398920172

Hunt, C.P., Moskowitz, B.M., Banerjee, S.K., 2013. Magnetic Properties of Rocks and Minerals. pp. 189–204. https://doi.org/10.1029/RF003p0189

Ingo, G.M., Riccucci, C., Pisani, G., Pascucci, M., D'Ercole, D., Guerriero, E., Boccaccini, F., Falso, G., Zambonini, G., Paolini, V., Di Carlo, G., 2022. The vehicle braking systems as main source of inhalable airborne magnetite particles in trafficked areas. Environ. Int. 158, 106991. https://doi.org/10.1016/j.envint.2021.106991

Kakosimos, K.E., Hertel, O., Ketzel, M., Berkowicz, R., 2010. Operational Street Pollution Model (OSPM) - a review of performed application and validation studies, and future prospects. Environ. Chem. 7, 485. https://doi.org/10.1071/EN10070

Kruiver, P.P., Dekkers, M.J., Heslop, D., 2001. Quantification of magnetic coercivity components by the analysis of acquisition curves of isothermal remanent magnetisation. Earth Planet. Sci. Lett. 189, 269–276. https://doi.org/10.1016/S0012-821X(01)00367-3

Kukutschová, J., Moravec, P., Tomášek, V., Matějka, V., Smolík, J., Schwarz, J., Seidlerová, J., Šafářová, K., Filip, P., 2011. On airborne nano/micro-sized wear particles released from low-metallic automotive brakes. Environ. Pollut. 159, 998–1006. https://doi.org/10.1016/j.envpol.2010.11.036

Letaïef, S., Camps, P., Poidras, T., Nicol, P., Bosch, D., Pradeau, R., 2020. Biomagnetic

Monitoring vs. CFD Modeling: A Real Case Study of Near-Source Depositions of Traffic-Related Particulate Matter along a Motorway. Atmosphere (Basel). 11, 1285. https://doi.org/10.3390/atmos11121285

Liati, A., Spiteri, A., Dimopoulos Eggenschwiler, P., Vogel-Schäuble, N., 2012. Microscopic investigation of soot and ash particulate matter derived from biofuel and diesel: implications for the reactivity of soot. J. Nanoparticle Res. 14, 1224. https://doi.org/10.1007/s11051-012-1224-7

Maher, B.A., Ahmed, I.A.M., Davison, B., Karloukovski, V., Clarke, R., 2013. Impact of roadside tree lines on indoor concentrations of traffic-derived particulate matter. Environ. Sci. Technol. https://doi.org/10.1021/es404363m

Maher, B.A., Moore, C., Matzka, J., 2008. Spatial variation in vehicle-derived metal pollution identified by magnetic and elemental analysis of roadside tree leaves. Atmos. Environ. 42, 364–373. https://doi.org/10.1016/j.atmosenv.2007.09.013

Mathissen, M., Scheer, V., Vogt, R., Benter, T., 2011. Investigation on the potential generation of ultrafine particles from the tire–road interface. Atmos. Environ. 45, 6172–6179. https://doi.org/10.1016/j.atmosenv.2011.08.032

Matzka, J., Maher, B.., 1999. Magnetic biomonitoring of roadside tree leaves: identification of spatial and temporal variations in vehicle-derived particulates. Atmos. Environ. 33, 4565–4569. https://doi.org/10.1016/S1352-2310(99)00229-0

Muxworthy, A.R., Lam, C., Green, D., Cowan, A., Maher, B.A., Gonet, T., 2022. Magnetic characterisation of London's airborne nanoparticulate matter. Atmos. Environ. 287, 119292. https://doi.org/10.1016/j.atmosenv.2022.119292

Pike, C.R., Roberts, A.P., Verosub, K.L., 1999. Characterizing interactions in fine magnetic particle systems using first order reversal curves. J. Appl. Phys. 85, 6660–6667.

https://doi.org/10.1063/1.370176

Prévot, M., Mankinen, E.A., Grommé, S., Lecaille, A., 1983. High paleointensities of the geomagnetic field from thermomagnetic studies on Rift Valley pillow basalts from the Mid-Atlantic Ridge. J. Geophys. Res. 88, 2316. https://doi.org/10.1029/JB088iB03p02316

Rahimi, M., Bortoluzzi, D., Wahlström, J., 2021. Input Parameters for Airborne Brake Wear Emission Simulations: A Comprehensive Review. Atmosphere (Basel). 12, 871. https://doi.org/10.3390/atmos12070871

Rhys, H.I., 2020. Machine Learning with R, the tidyverse, and mlr. Manning.

Roberts, A.P., Heslop, D., Zhao, X., Pike, C.R., 2014. Understanding fine magnetic particle systems through use of first-order reversal curve diagrams. Rev. Geophys. 52, 557–602. https://doi.org/10.1002/2014RG000462

Roberts, A.P., Pike, C.R., Verosub, K.L., 2000. First-order reversal curve diagrams: A new tool for characterizing the magnetic properties of natural samples. J. Geophys. Res. Solid Earth. https://doi.org/10.1029/2000JB900326

Robertson, D.J., France, D.E., 1994. Discrimination of remanence-carrying minerals in mixtures, using isothermal remanent magnetisation acquisition curves. Phys. Earth Planet. Inter. 82, 223–234. https://doi.org/10.1016/0031-9201(94)90074-4

Sagnotti, L., Taddeucci, J., Winkler, A., Cavallo, A., 2009. Compositional, morphological, and hysteresis characterization of magnetic airborne particulate matter in Rome, Italy. Geochemistry, Geophys. Geosystems. https://doi.org/10.1029/2009GC002563

Vlasov, D., Ramírez, O., Luhar, A., 2022. Road Dust in Urban and Industrial Environments: Sources, Pollutants, Impacts, and Management. Atmosphere (Basel). 13, 607. https://doi.org/10.3390/atmos13040607

Winkler, A., Contardo, T., Lapenta, V., Sgamellotti, A., Loppi, S., 2022. Assessing the impact of vehicular particulate matter on cultural heritage by magnetic biomonitoring at Villa Farnesina in Rome, Italy. Sci. Total Environ. 823, 153729. https://doi.org/10.1016/j.scitotenv.2022.153729

Winkler, A., Contardo, T., Vannini, A., Sorbo, S., Basile, A., Loppi, S., 2020. Magnetic Emissions from Brake Wear are the Major Source of Airborne Particulate Matter Bioaccumulated by Lichens Exposed in Milan (Italy). Appl. Sci. 10, 2073. https://doi.org/10.3390/app10062073

Winklhofer, M., Zimanyi, G.T., 2006. Extracting the intrinsic switching field distribution in perpendicular media: A comparative analysis. J. Appl. Phys. 99, 08E710. https://doi.org/10.1063/1.2176598

World Health Organization, 2022. WHO Air quality Database 2022 [WWW Document]. Consult. Novemb. 10th. URL https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database

Yang, Y., Vance, M., Tou, F., Tiwari, A., Liu, M., Hochella, M.F., 2016. Nanoparticles in road dust from impervious urban surfaces: distribution, identification, and environmental implications. Environ. Sci. Nano 3, 534–544. https://doi.org/10.1039/C6EN00056H