Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Meteorological normalisation of PM$_{10}$ using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah

Marc Daniel Mallet$^{a,b}$

$^a$ School of Earth and Atmospheric Sciences, Queensland University of Technology, Brisbane, Australia
$^b$ Now at Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia

ARTICLE INFO

Keywords:
meteorological normalization
PM$_{10}$
Air quality
Mining activities
Machine learning

ABSTRACT

The impacts of poor air quality on human health are becoming more apparent. Businesses and governments are implementing technologies and policies in order to improve air quality. Despite this, the PM$_{10}$ air quality in the mining town of Moranbah, Australia, has worsened since measurements commenced in 2011. The annual average PM$_{10}$ concentrations during 2012, 2017, 2018 and 2019 have all exceeded the Australian National Environmental Protection Measure’s standard, and there has been an increase in the frequency of exceedances of the daily standard. The average annual increase in PM$_{10}$ was 1.2 ±0.5 $\mu g\ m^{-3}$ per year between 2011 and 2019 and has been 2.5 ±1.2 $\mu g\ m^{-3}$ per year since 2014. The cause of this has not previously been established. Here, two machine learning algorithms (gradient boosted regression and random forest) have been implemented to model and then meteorologically normalise PM$_{10}$ mass concentrations measured in Moranbah. The best performing model, using the random forest algorithm, was able to explain 59% of the variance in PM$_{10}$ using a range of meteorological, environmental and temporal variables as predictors. An increasing trend after normalising for these factors was found of 0.6 ± 0.5 $\mu g\ m^{-3}$ per year since 2011 and 1.7 ± 0.3 $\mu g\ m^{-3}$ per year since 2014. These results indicate that more than half of the rise in PM$_{10}$ is due to a rise in local emissions in the region. The remainder of the rise in PM$_{10}$ was found to be due to a decrease of soil water content in the surrounding region, which can facilitate higher dust emissions. Whether the presence of open-cut coal mines exacerbated the role of soil water content is unclear. Although fires can have drastic effects on the local air quality, changes in fire patterns are not responsible for the rising trend. PM$_{10}$ composition measurements or more detailed data relating to local sources is still needed to better isolate these emissions. Nonetheless, this study highlights the need and potential for action by industry and government to improve the air quality and reduce health risks for the nearby population.

1. Introduction

The adverse health effects of ambient air pollution, including particulate matter, has been thoroughly established (Rückerl et al., 2011; Brunekreef and Holgate, 2002; Liu et al., 2019a). Particulate matter below 10 $\mu m$ (PM$_{10}$) is respirable by humans and is one of the air pollutants regulated by the World Health Organization (WHO, 2006). Compared with other countries, the concentration of PM$_{10}$ in major Australian cities is relatively low (Liu et al., 2019a). The Australian National Pollutant Inventory (NPI) has identified open-cut coal mines as the most significant source of PM emissions, with approximately 25% of industrial emissions arising from coal mining activities (Richardson et al., 2018). Furthermore, a recent study, has shown the PM$_{10}$ concentrations in Australia are higher in coal mining towns than non-mining urban and rural regions (Hendryx et al., 2020). Previous studies have shown clear links between coal mining and respiratory diseases (Perret et al., 2017; Laney and Weissman, 2014). Populations living within close proximity to coal mines have been shown to be at an increased risk of mortality and/or morbidity across numerous disease classifications including neoplasms, diseases of the circulatory, respiratory and genitalourinary systems, metabolic diseases, diseases of the eye and skin, perinatal conditions, congenital and chromosomal abnormalities, and...
external causes of morbidity (Cortes-Ramirez et al., 2018).

There have been recent concerns about the air quality in the town of Moranbah in north Queensland, Australia (21.9995S, 148.0713E). Surrounding Moranbah are several large, operational, open-cut coal mines (see Fig. 1), which provide a potential source of dust to the local population (Alvarado et al., 2015; Richardson et al., 2018; Ghose and Majee, 2007; National Pollutant Inventory (NPI), 2020) (see Supplementary Fig. 1b). Although it has not been studied in this region, other potential sources of PM$_{10}$ identified around Australia include agricultural dust (Shattaran and D’odorico, 2014), the long-ranged transport of dust from central Australia (Chan et al., 2005; Radhi et al., 2010; Perron et al., 2020), smoke from fires (Morgan et al., 2010; Johnston et al., 2013; Milic et al., 2016), sea spray (Gras and Ayers, 1983; Radhi et al., 2010), and emissions from local traffic or industry (Keywood et al., 2011; Chan et al., 2008). Furthermore, PM$_{10}$ concentrations can also be affected by the surface meteorological conditions, atmospheric boundary layer height, air mass back trajectory and synoptic scale conditions (Wise and Comrie, 2005b, a; Grange et al., 2018). These effects can suppress or exacerbate the PM$_{10}$ concentrations, sometimes to a greater extent than the impacts from changes in source emissions (Fuller and Font, 2019). Alarmingly, PM$_{10}$ concentrations have increased in Moranbah, however the cause of this increase has not been established. It is therefore necessary to account for variation in meteorological conditions to examine whether this increase is because of changes in meteorology or because of increases in source emissions.

In Moranbah, and as is the case for most monitoring stations, chemical composition measurements have not been taken. Furthermore, while emissions inventories for surrounding coal mines are kept on an annual basis, these are not detailed enough to be able to draw detailed quantitative conclusions about how potential sources can influence the PM$_{10}$ concentration in the township. This problem can be overcome with long-term measurements with the aid of machine (or statistical) learning.

Machine learning methods are becoming increasingly popular tools in outdoor air quality modelling. Although other traditional statistical methods have been used to examine long term trends in air quality and

![Image of Moranbah and surrounding features](image.png)

Fig. 1. The location of Moranbah and surrounding features. a) Moranbah (white point) with the surrounding mine sites, b) the location of Moranbah in Queensland, Australia, c) the location of the air quality monitoring station within Moranbah, and d) the estimated populations of Moranbah and the Isaac region.
meteorology in Australia (de Jesus et al., 2020), the benefits of machine learning methods are yet to be realised in Australia and across the southern hemisphere (Rybarczyk and Zalakevičius, 2018). Ensemble machine learning methods, such as random forests (Breiman, 2001) or gradient boosted regression (Freund and Schapire, 1995; Friedman, 2001), use a range of predictor variables and an ensemble of decision trees to make predictions. They offer a considerable advantage over other machine learning methods such as neural networks because the relationships between the predictor variables and the predicted variable can be fully interpreted (Fuller and Font, 2019). Furthermore, both numeric and categorical predictor variables can be used, which allows complex systems such as regional synoptic conditions or air mass origins to be categorized and used to improve model explanatory power (Grange et al., 2018).

Meteorological normalisation involves using random forests or gradient boosted regression to predict PM$_{10}$ (or any air pollutant) and accounts for the influence of meteorological or other external factors such as wind speed and direction, temperature, boundary layer height and air mass backwards trajectory. The process of using random forests and meteorologically normalise on air pollutants is described in depth in Grange et al. (2018). They use random forests to predict and meteorologically normalise PM$_{10}$ concentrations at air quality monitoring sites across Switzerland. Their technique showed a significant decline in normalised PM$_{10}$ concentrations across the country and attribute this to reduced emissions due to vehicle and heating emissions control measures. Since then other studies have used this meteorological normalisation method to assess the impact of policy interventions on SO$_2$, NO$_x$ and NO$_y$ in England (Grange and Carslaw, 2019), the impact of clean air action on air quality in Bejing (Yu et al., 2019; Zhang et al., 2019; Liu et al., 2019b), and to assess the impacts of relocating air quality monitoring sites in the United Kingdom (Walker et al., 2019). This technique is therefore very useful for detecting changes in air quality in regions with potentially high emissions from industry.

Residents of Moranbah have reportedly been concerned with the high levels of dust appearing in households but, to date, a comprehensive investigation of trends and drivers of air quality in the townships has not been done. The objective of this study is to exploit the recent advances in machine learning to investigate the trends in PM$_{10}$ concentration in Moranbah and assess the impact of changes in local industrial actions on air quality using open-access datasets and techniques. The primary intent of this study is therefore to provide local and state governments, as well as industry, a starting point to assess how changes in industrial development, residential growth or modes of employment might influence the air quality to inform future policies or procedures. The secondary intent is to establish a methodology for this meteorological normalisation that accounts for the influence of nearby fires, which are an important source of particulate matter in the Australian dry season, as well as other environmental factors such as soil water content. This study will therefore provide an updated meteorological normalisation technique that can then be applied to the numerous datasets of long-term monitoring of air quality across Australia.

2. Methodology

All data loading, cleaning, processing, analyses, statistical modelling and visualisation was carried out using R (R Core Team, 2019), with many tools used from the tidyverse suite of packages (Wickham et al., 2019). Temporal trends in the PM$_{10}$ concentrations and corresponding meteorological and environmental predictor variables were estimated using the Theil-Sen estimator (Zeileis et al., 2003).

2.1. Study area

Moranbah (21.9995S, 148.0713E) is part of the Isaac region within the broader Bowen Basin, one of Australia’s most important source of export coal (Fig. 1). Moranbah grew substantially during a mining boom between 2002 and 2012 (Petkova et al., 2009; Warren et al., 2017), with the estimated residential population increasing from 6500 to 8900, but has remained relatively steady since then. The estimated residential population of the Isaac region grew from 19,000 to 23,000 between 2002 and 2012, and has decreased to 21,000 as of 2019. Moranbah and the Isaac region also have a substantial additional population of non-residents that either fly or drive in and out for employment in the region. The Isaac region non-residential population rose sharply from 2006 until it peaked in 2012 at 17,000, before dropping to 9400 in 2016 and slowly rising to 12,000 in 2019 (QGSO, 2020) (Fig. 1d).

2.2. Data

2.2.1. PM$_{10}$

Hourly PM$_{10}$ have been measured in Moranbah (21.9995S, 148.0713E) by the Queensland Government Department of Environment and Science since March 2011 (Queensland Government, 2020). PM$_{10}$ is measured using a Tapered Element Oscillating Balance (TEOM) at a height of 4 m above ground level. The TEOM is operated in accordance to AS 3580.98–2001 and adjusted in accordance with US EPA equivalence designation method EQPM-1090-079 requirements (i.e. scaled by 1.03 and offset by +3.0). After the adjustment, negative hourly concentrations less than –0.5 μg m$^{-3}$ are removed and all other negative concentrations are set to 0.0 μg m$^{-3}$.

2.2.2. Meteorological, environmental and temporal predictor variables

A range of meteorological, environmental and temporal variables were gathered to use as predictors for the PM$_{10}$ measured in Moranbah. Each of the meteorological and environmental variables are described in this section. Table 1 summarises each of these and the trends for each variable are shown in Figs. 2 and 3.

Wind speed and direction are measured using an ultrasonic sensor at a height of 10 m above ground level. Temperature is measured using a capacitive ceramic sensor, which can be prone to overestimating the actual temperature under still conditions. Wind speed and direction and temperature were all measured at the same monitoring station in Moranbah as the PM$_{10}$ measurements. To provide a comparison of the daily and weekly cycles of PM$_{10}$ air quality data from the Brisbane CBD (28.4774S, 153.0281E) was also retrieved. The time series of the meteorological variables is shown in Fig. 2.

Daily rainfall data was fetched using the bomrang R package (Sparks et al., 2017). The sites considered were Wentworth (22.07S; 147.72E), approximately 30 km west of Moranbah, from 2011 until 2019, and from the Moranbah Airport (22.06S; 148.08E) between 2012 and 2019. When any amount of rainfall was measured at both sites on any given day, the average from both sites was taken as the daily rainfall. On days when only one site recorded rainfall, that amount was taken for the daily rainfall.

Hourly atmospheric boundary layer heights, volumetric soil water

| Predictor variables used to model the PM$_{10}$ in Moranbah. | Predictor variable | Variable type | Data source |
|-------------------------------------------------------------|--------------------|---------------|-------------|
| Temperature (°C)                                            | Meteorological     | Queensland Government (2020) |
| Wind speed (ms$^{-1}$)                                      | Meteorological     | Queensland Government (2020) |
| Wind direction (°)                                          | Meteorological     | Queensland Government (2020) |
| Pressure (hPa)                                              | Meteorological     | CDS (2017)    |
| Boundary layer height (m)                                   | Meteorological     | CDS (2017)    |
| Rainfall (mm d$^{-1}$)                                       | Meteorological     | Sparks et al. (2017) |
| Air mass cluster                                            | Meteorological     | Knox et al. (1996) |
| Fire power (Wm$^{-2}$)                                       | Environmental      | Geoscience Australia (2020) |
| Soil water content (m$^{3}$m$^{-2}$)                         | Environmental      | CDS (2017)    |
| Hour of day                                                 | Temporal           | N/A           |
| Day of week                                                 | Temporal           | N/A           |
| Week of year                                                | Temporal           | N/A           |
| Unix date                                                   | Temporal           | N/A           |
content in the top 7 cm and ground-level pressure data were retrieved from the European Centre for Medium Weather Forecasting’s Reanalysis-5 (ERA5) reanalysis (CDS, 2017) from 2011 until 2019 between 22.50S and 21.75S, and 147.5E and 149.0E. The mean of these three variables over this spatial domain at each hour were then calculated and used as the representative values in the Moranbah region.

Historic fire hotspots were obtained from Geoscience Australia Sentinel Hotspots database (Geoscience Australia, 2020) that collates measurements from the Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) satellite instruments. Here, hotspots detected by MODIS on the Aqua satellite within 200 km of the Moranbah air quality monitoring site and within m 45° of the prevailing wind direction at each hour were selected. The hotspots are detected by MODIS on the polar orbiting satellite, Aqua, which has a flyover time just after noon local time. The occurrence, location and fire power was then assumed to be constant over the whole day.

120-h backwards trajectories were calculated every 6 h (00, 06, 12, 18 UTC) at a height of 100 m above ground level using the NCEP/NCAR Reanalysis meteorological data (Kalnay et al., 1996) and the HYSPLIT trajectory model (Stein et al., 2015) between 2011 and 2019 over the Moranbah measurement site. This was done using the splitR R package (Iannone, 2019). A K-means clustering analysis was then applied to each 6-hourly trajectories using the OpenAir R-package (Carslaw and Ropkins, 2012) (see Fig. 3). This was done using both the Euclidean distance and the angular distance between trajectories, and by selecting three up to eight clusters for each method. Seven clusters using the angular distance between trajectories was selected, because this resulted in the largest influence on PM$_{10}$ after using gradient boosted regression. The average trajectories of the seven clusters as well as the monthly frequency of each cluster is shown in Fig. 3.

The hour of day and day of week temporal variables are used as proxies for local traffic sources, while the week of year and unix date temporal variables are included to account for seasonal and long-term variability in PM$_{10}$ that are not accounted for in the meteorological and environmental variables (i.e. changes in local emissions).

Fig. 2. The time series of: (a) monthly temperature, (b) monthly wind speed, (c) monthly wind direction, (d) monthly pressure, (e) monthly boundary layer height, (f) the monthly mean of daily rainfall near Moranbah, (g) the monthly volumetric soil water content in the top 7 cm of soil, and (h) the monthly mean sum of MODIS-detected fire power from fires within 200 km of Moranbah and within 90° of the station wind direction. The trends were calculated using the Theil-Sen estimator with 95% confidence intervals. The +, *, and ** symbols next to the trend values indicate p < 0.1, p < 0.05 and p < 0.01, respectively. Horizontal bars represent the frequency distribution of daily averages.
2.3. Modelling PM$_{10}$ with machine learning algorithms

The meteorological, environmental and temporal variables were then used as predictors for modelling the PM$_{10}$ concentration in Moranbah using two machine learning algorithms.

Gradient boosted regression and random forest models are two types of ensemble learning methods that rely on building a strong prediction based on many so-called "weak" predictors based on decision trees. In random forest models the decision trees built independent of each other and are averaged to give a final prediction. In gradient boosted regression, the decision trees are built sequentially and iteratively predict the residual between the trained data and prediction from the previous decision tree. Each subsequent decision tree is multiplied by a learning rate (typically between 0.001 and 0.1) before being summed to give a final prediction.

The performance of these algorithms are sensitive to several hyperparameters, such as the number of trees, the interaction or tree depth and in the case of gradient boosting, the learning rate and the number of samples considered in each split within a decision tree. A trade-off for model performance is over-fitting. An over-fit model will be tuned to noise in the training data and therefore won’t be generalised and could fail when trying to make predictions on new data. Here, both the gradient boosted regression and random forest algorithms were applied to predict and meteorologically normalise the PM$_{10}$ concentrations in Moranbah.

2.3.1. Gradient boosted regression

Gradient boosted regression was performed using an adaption of the deweather R package (Carslaw, 2018), which uses the underlying gbm R package (Greenwell et al., 2019) to run the model. The functions to test and build the model were extended to allow for the interaction depth, learning rate and minimum number of samples considered for each node to be changed. These hyper parameters were tuned by using a train fraction of 0.8 (sampled randomly) and a bag fraction of 0.5 and running the model ten times for each combination of hyper parameters (see Supplementary Table 1 for the grid of tested hyper parameters). The average and standard deviation of the output statistics were then calculated for each combination of hyper parameters. The optimal combination of hyper parameters based on the coefficient of determination and degree of over-fitting were then chosen to build the model for meteorological normalisation and to explore the partial dependencies between each predictor variable and PM$_{10}$.

2.3.2. Random forest

The random forest modelling was performed using the rmweather R package (Grange et al., 2018; Grange and Carslaw, 2019), which uses the underlying ranger R package (Wright and Ziegler, 2017). The default training fraction of 0.8 (sampled randomly) from the rmweather package was used and a combination of different hyper parameters were also tested (see Supplementary Table 2). These included the number of trees, the number of variables per node and the minimum node size. The decision trees in random forests are independent from each other and therefore a learning rate hyper parameter is not considered as it is with gradient boosting. The best performing combination of hyper parameters were again selected for the rest of the analysis using the random forest algorithm.

2.3.3. Meteorological normalisation

The selected gradient boosted regression and random forest models were then used to meteorologically normalise the time series of PM$_{10}$. In this process, the PM$_{10}$ concentration for each hour over the sampling period is predicted using the partial dependence of a linear time variable and by randomly sampling the remaining meteorological and other predictor variables. For both the random forest and gradient boosted
regression meteorological normalisation, each hourly PM$_{10}$ concentration was predicted 1000 times using the random sampling and then averaged to give a normalised PM$_{10}$ concentration. This process of meteorological normalisation was applied to account for the influence of meteorology, fires and soil water content on the PM$_{10}$ concentrations in Moranbah. The meteorological normalisation procedure was repeated an additional three times, although excluding the temperature, soil water content and fire power variables, respectively. This allowed the impact of each of these on the normalised trend to be individually investigated.

A significant advantage of the gradient boosted regression and random forest machine learning techniques is that the partial dependencies between the predictor variables and predicant can be investigated. This is done by randomly sampling all but one of the predictor variables, one at a time. Exploiting this allows for the influence of each predictor variables on PM$_{10}$ to be isolated.

3. Results and discussion

3.1. Air quality in Moranbah

Measurements of PM$_{10}$ in Moranbah commenced in March 2011 (see Fig. 4). Since then there has been a statistically significant average annual increase in PM$_{10}$ of $1.2 \pm 0.5 \mu g \ m^{-3}$ per year ($p < 0.001$). Alarming, since 2014, this increase has been much higher, with an average annual increase of $2.5 \pm 1.2 \mu g \ m^{-3}$ per year ($p < 0.001$, not shown).

The Australian National Environmental Protection Measure (NEPM) have indicated safe levels of daily and annually average PM$_{10}$ concentrations of $50 \mu g \ m^{-3}$ and $20 \mu g \ m^{-3}$, respectively. In the case of the daily NEPM standard, an exceedance limit of five days per year has been recommended. Table 2 indicates the annually average PM$_{10}$ between 2011 and 2019 and the number of days that the daily limit was exceeded. Over these nine years, four of them (2012, 2017, 2018 and 2019) exceeded the annual limit and five (2011, 2012, 2017, 2018 and 2019) had at least five days per year that daily limit was exceeded.

These results clearly show PM$_{10}$ air quality in Moranbah often exceed recommended safe levels and that the air quality has worsened since measurements began in 2011. It is therefore important to identify the sources of PM$_{10}$ and the possible explanations for why concentrations are rising. Understanding whether the rising PM$_{10}$ is due to changes in local emissions or natural meteorological or environmental causes is therefore extremely important. Because composition measurements have not been taken, which would help identify sources of PM$_{10}$, other techniques are required. The remainder of this study will present the results of modelling and meteorological normalisation of the PM$_{10}$ concentrations in Moranbah using machine learning algorithms and a range of meteorological and environmental predictor variables.

3.2. Model results

The models built using the random forest algorithm outperformed those built using gradient boosted regression. For the random forest models, the coefficient of determination ($R^2$) ranged between 0.49 and 0.59 across the grid of hyper parameters (Supplementary Fig. 2), while for the gradient boosted regression models, $R^2$ ranged between 0.25 and 0.49 (Supplementary Fig. 3). It is difficult to identify the reason that the random forest algorithm outperformed the gradient boosted regression in this study. Even though a wide range of hyper parameters were tested with both models, the random forest models were able to explain more than 10% more of the variance in PM$_{10}$ than the gradient boosted regression. One possible reason is that the random forest models were less prone to over-fitting than the gradient boosted regression for this data set.

The optimal random forest model ($R^2 = 0.59$, RMSE $= 19.5$) was given by using 600 trees, six variables per node and a minimum node size of 10, although overall performance did not appear to be overly sensitive to the selection of hyper parameters. For the gradient boosted regression models, the performance was sensitive only to the number of trees selected, with a higher coefficient of determination given a higher number of trees. Although selecting 4000 trees gave the highest $R^2$ values for the training datasets, the average $R^2$ values for the training datasets were significantly higher ($R^2 > 0.8$), which indicates that these models were over-fit. The optimal gradient boosted regression model ($R^2 = 0.46$; RMSE $= 21.1$) was therefore selected with 1000 trees, an interaction depth of 4, a learning rate of 0.005 and a minimum of 10 observations in each node. Although this combination still resulted in a training $R^2$ of 0.67, indicating a smaller degree of over-fitting, it was also computationally much faster than a much higher number of trees.

The optimal random forest model and gradient boosted regression model explained 59% and 46% of the variance in PM$_{10}$ concentrations in Moranbah, respectively, using the selected meteorological, environmental and temporal predictor variables. This is sufficient to then perform the meteorological normalisation of PM$_{10}$ and to explore the partial dependencies between PM$_{10}$ and each predictor variable.

3.3. Meteorological normalisation of PM$_{10}$

The meteorologically normalised trend using both the gradient boosted regression and random forest techniques are shown in Fig. 5. The normalised trends from both models agree with each other well ($R^2 = 0.66$ and a linear regression coefficient of 1.2, Supplementary Fig. 5). Using the Theil-sen estimator, statistically significant trends ($p < 0.001$) were observed in both normalised trends of $0.62 \pm 0.5 \mu g \ m^{-3}$ per year for the random forest model and $0.5 \pm 0.4 \mu g \ m^{-3}$ per year for the gradient boosted regression model. Since 2014, the normalised trends have been $1.7 \pm 0.3 \mu g \ m^{-3}$ per year and $1.9 \pm 0.4 \mu g \ m^{-3}$ per year for the random forest and gradient boosted regression models, respectively.

---

Table 2

| Year | Annual PM$_{10}$ ($\mu g \ m^{-3}$) | Daily exceedences |
|------|---------------------------------|------------------|
| 2011 | 20.6                            | 5                |
| 2012 | 28.8*                           | 36               |
| 2013 | 22.7                            | 1                |
| 2014 | 20.5                            | 0                |
| 2015 | 21.5                            | 4                |
| 2016 | 22.3                            | 0                |
| 2017 | 25.9*                           | 7                |
| 2018 | 30.1*                           | 19               |
| 2019 | 31.0*                           | 31               |

Fig. 4. The time series of daily PM$_{10}$ where the black and red horizontal dashed lines represent the Australian NEPM standards for the annual average and the 24-h average, respectively. The connected black dots represent the annually average PM$_{10}$ concentrations. Horizontal bars represent the frequency distribution of daily concentrations.
Both reveal several consistent changes in the normalised PM$_{10}$ concentration. These changes (i.e. breakpoints) in the normalised PM$_{10}$ are due to changes in emissions of PM$_{10}$ sources, rather than meteorological or environmental effects. The dates that these breakpoints were observed are detailed in the supplementary material (Supplementary Table 3). Most of the observed breakpoints are relatively abrupt, unlike changes observed in other studies that observed a transition in emissions due to policy intervention (Grange et al., 2018; Grange and Carslaw, 2019; Vu et al., 2019). These breakpoints are unlikely to be associated with natural causes or variability (e.g. fires, long range transport of dust, increased emissions of biogenic vapors) (see Section 3.4).

The trends in the meteorologically normalised PM$_{10}$ are approximately half that of the trend in the measured PM$_{10}$ concentrations. This indicates that, while meteorological or environmental variability is a partial cause for the rising PM$_{10}$ in Moranbah, changes in local source emissions are also responsible. In order to disentangle these contributions, it is useful to explore the partial dependencies between PM$_{10}$ and each predictor variable as well as the overall trends in the predictor variables themselves.

![Fig. 5. The meteorologically normalised PM$_{10}$ concentrations at Moranbah calculated using gradient boosted regression and random forest. The vertical dashed lines represent identified break-points in the time series.](image)

![Fig. 6. The partial dependencies of each PM$_{10}$ predictor variable from gradient boosting. Error shadings represent the standard deviation of the 1000 predictions using the gradient boosting model. The violet dots represent the partial dependencies from the random forest model. Horizontal bars represent the frequency distribution of daily concentrations.](image)
3.3.1. Partial dependencies

The partial dependence plots from both the gradient boosted regression (orange lines) and random forest (purple dots) models are shown in Fig. 6. The partial dependencies of each predictor variable and PM$_{10}$ are broadly consistent between the two models. Some differences are present (e.g. for northwesterly wind directions, high wind speeds and high atmospheric boundary layers) which are likely due to fewer samples observed for those conditions. Despite these differences the trends between the predictor variables and predicted PM$_{10}$ appear robust across both models.

The sum of fire powers within 200 km of the Moranbah site and prevailing wind direction was the predictor variable with the largest influence on the predicted PM$_{10}$ concentration from gradient boosted regression. Although the impact of fires on particulate matter and air quality has not been studied in this region, they have been shown to be an important source in north Australia during the dry season (Meyer et al., 2008; Mallet et al., 2017), in Brisbane (Mille et al., 2016; He et al., 2016; Chan et al., 1999), as well as the other regions of Australia (Keywood et al., 2011; Reisen et al., 2011). Here, the relationship between the PM$_{10}$ concentration and the fire power is not linear, which highlights the complexity and difficulty of predicting PM$_{10}$ from satellite retrievals of fire location and power. A large spike in PM$_{10}$ concentrations is seen in the partial dependence plot for a fire power of $1750~\text{W}$, which is due to a large fire event in late November of 2012 within 50 km east of Moranbah. Despite this, the PM$_{10}$ concentration is, on average, higher when the fire power is above $1700~\text{W}$ and much more variable. A limitation in the methodology used to retrieve fires is that they are based on retrievals from polar orbiting satellites. Therefore, only fires that burn for a considerable time to be detected at the time of the satellite fly-over are counted. Furthermore, only fires within 200 km of the prevailing wind direction are considered, but smoke from large fires can be carried over distances of hundreds or thousands of kilometres. Lastly, the fire power is only a weak proxy for emissions of particulate matter, since smouldering fires with a lower power can emit enormous amounts compared with larger open flame fires (Desservetiaz et al., 2017). Despite these caveats, by not including the fire power as an explanatory variable reduced the performance of the gradient boosted regression model, decreasing the R value between the predicted and observed PM$_{10}$ by 0.1. Interestingly, although the fire power was an important predictor variable of PM$_{10}$, the meteorologically normalised PM$_{10}$ trend was rather insensitive to its inclusion or exclusion.

The variable with the second largest influence on the PM$_{10}$ concentration was the trend variable, which is the date expressed as the Unix time of each observation (i.e. number of seconds since January 1, 1970) and can be used as a proxy for changes in emissions or other variables that are unaccounted for. This partial dependence gives an indication of the meteorologically normalised trend which will be discussed later.

Temperature was an important predictor variable on the predicted PM$_{10}$. There are many ways that temperature can influence air quality from changes in various source emissions as well as atmospheric processing during transport. Cold temperatures can lead to residents using wood fire heating to warm households which can lead to severe increases in PM concentrations (Johnston et al., 2013). Warm temperatures associated with increased emission of biogenic vapors which can lead to increased PM concentrations secondary aerosol formation (Griffin et al., 1999). Higher temperatures could also be linked to fires, and any influence by those not considered in the sum of fire power predictor variable would be included in the partial dependence of temperature and PM$_{10}$. The partial dependence of PM$_{10}$ and week of the year shows a clear annual cycle, with higher concentrations in the Austral spring. This could be due to annual cycles in the activities of the township of Moranbah and surrounding industries, but, like temperature, is also likely to be partially representative of the annual fire season in this region.

The volumetric soil water content in the top 7 cm in the region surrounding Moranbah was an influential variable on PM$_{10}$. Because PM$_{10}$ emissions are lower for high soil water content, this suggests that wind blown dust is an important source of PM$_{10}$, as soil moisture has been shown to inhibit dust emissions (Funk et al., 2008; Csavina et al., 2012).

The partial dependencies of wind speed, wind direction, weekday and hour of the day highlight the elevated PM$_{10}$ concentrations when the wind speed is large ($>5~\text{m/s}$), the elevated PM$_{10}$ concentrations that throughout the evening, as well as the relatively similar concentrations observed on weekdays and weekends. The partial dependence of PM$_{10}$ and wind direction indicates that winds from the north-east and south are responsible for the highest PM$_{10}$ concentrations. This will discussed further in the next section.

The air mass backwards trajectory was not an influential variable on the predicted PM$_{10}$, giving strong evidence that local, rather than regional or continental, emissions are responsible for variability in PM$_{10}$. Nonetheless, there are still distinct differences in PM$_{10}$ for different air mass backwards trajectory clusters. The two clusters, C5 and C7, that resulted in the highest PM$_{10}$ concentration were those that spent the most time over continental Australia, suggesting that long-range transport of terrestrial sources such as dust or biomass burning aerosol do have a small impact on the air quality at Moranbah. In contrast the two clusters, C1 and C2, resulted in the smallest concentrations of PM$_{10}$ and were the two clusters that spent the least amount of time over continental Australia. Emissions from long-ranged sources, such as dust storms and large-scale distant fires are somewhat accounted for within the air mass back trajectory cluster predictor variable.

The atmospheric boundary layer height partial dependence showed a minimum PM$_{10}$ for boundary layers $<1000$m, and elevated concentrations below this as well as above 2000m, although these cases were rare. This phenomena was also observed in PM$_{10}$ measurements across Switzerland (Grange et al., 2018). They suggest that the PM$_{10}$ concentrations during low boundary layers and low temperatures and lead to high rates of surface-based emissions in winter, and that regimes consisting of high boundary layers are associated with deep convection that allows for the transport of PM$_{10}$ from other sources. Whether this explanation is valid for this site or not is uncertain. It is also plausible that the periods when the boundary layer was high corresponded with a period when smoke from fires impacted the air quality. This was the case for an event in early December 2018 when the boundary layer was greater than 3000 m when extensive smoke can easily be observed over the entire region from MODIS satellite imagery.

Lastly, rainfall, pressure and week of year were not important predictor variables for PM$_{10}$. Nonetheless, higher amounts of rainfall lead to lower PM$_{10}$. There are two possible explanations for this. The first is that after or during rainfall, soil can be more moist and therefore lower dust emissions as discussed earlier. The second is that mining activities halt when there is rainfall, thereby reducing potential sources of PM$_{10}$. Although there is still a small seasonal cycle shown in the partial dependence plot for the week of year, most of the seasonal effects are taken into account in the other predictor variables (e.g. soil water content, fire power and temperature).

3.4. Explaining the rising PM$_{10}$

The partial dependencies are useful for describing how each meteorological or environment factor influences the PM$_{10}$ concentration. The combined changes in these over time, however, is what potentially drives the trend in PM$_{10}$. The results from the meteorological normalisation indicate that approximately half of the rise in PM$_{10}$ in Moranbah since 2011 has been due to meteorological and environmental changes. Excluding the soil water content from the meteorological normalisation process resulted in a normalised trend of $1.1\pm0.3~\mu\text{g m}^{-3}\text{per year}$ ($p<0.001$), which is nearly the same as the unnormalised trend of $1.2\pm0.5~\mu\text{g m}^{-3}\text{per year}$. As shown in Fig. 2, the soil water content has decreased by approximately 1% per year since 2011 ($p<0.001$), and these results

\[PM_{10} = \sum_{i=1}^{n} \text{Fire Power}_i \times \text{Temperature}_i \times \text{Time}_i \times \text{Soil Water Content}_i\]

where $n$ is the number of observations, Fire Power is the fire power variable, Temperature is the temperature variable, Time is the Unix time variable, and Soil Water Content is the volumetric soil water content variable.
The remainder of the rise in PM$_{10}$ is likely due to direct increases in local emissions. Potential major PM$_{10}$ sources are likely to include traffic from the town of Moranbah, nearby highways, and emissions from mining vehicles, the Moranbah airport to the south, local and distant fires, wind-blown dust from natural or agricultural land as well as from the open-cut mines that surround Moranbah. Other mining-related activities such as blasting or mechanical disturbances are also potential sources. By exploring the meteorologically normalised trend in PM$_{10}$, as well as the supporting meteorological and environmental data, the rest of this section will highlight why a change in mining related activities are the most plausible explanation for the increase in PM$_{10}$.

In addition to the partial dependence plots, the bivariate polar plot of mean PM$_{10}$ concentrations with wind speed and wind direction (Fig. 7) shows that high concentrations of PM$_{10}$ are present from multiple wind directions and are more common during periods of high wind speed. This indicates that mechanically activated sources of PM$_{10}$ such as dust are a likely contributor to the high concentrations of PM$_{10}$ at the Moranbah site. It should be noted that the polar plot gives an indication of the mean PM$_{10}$ for different wind speeds and directions but does not indicate the frequency of these conditions. Although in Moranbah the wind can come from any direction, the prevailing winds are easterly and southeasterly (see Fig. 2 and Supplementary Fig. 3).

The PM$_{10}$ concentrations in Moranbah also exhibit a wind direction-dependent daily cycle that has also changed between 2011 and 2019. This is shown in the yearly polar annulus plots in Fig. 8. As stated earlier, 2012, 2017, 2018 and 2019 had high annual PM$_{10}$ concentrations, but the polar annulus plots indicate differences in the sources across these years. Interestingly, 2012 showed high mean PM$_{10}$ concentrations from all wind directions, albeit at different times of the day. Prominent peaks around midday from easterly winds and evening hours from northerly winds can be observed. In contrast to all other years except 2019, the hours between midnight and sunrise (~0600h) show the lowest mean PM$_{10}$ concentrations through the day. In urban areas, it is typical to observe a diurnal cycle in air pollutants with a peak during morning hours and another peak during late afternoon hours that each represent emissions from morning and afternoon traffic as people drive and commute to work (Srimuruganandam and Nagendra, 2010; Morawska et al., 2007). This is not the case for Moranbah. It can be seen in the polar annulus plots that mean PM$_{10}$ concentrations remain high throughout the night.

To explore this further, it is useful to compare the average diurnal and weekday cycle in PM$_{10}$ from Moranbah and the capital city of Queensland, Brisbane (population of 2.3 million). This is summarised in Fig. 9. The first thing to note is that PM$_{10}$ concentrations in Brisbane are only roughly 60% of those in Moranbah, which is remarkable considering the population difference. The second is that the diurnal cycles between the two sites are different. Both exhibit a morning peak just after 0600h local time. The afternoon peak in Brisbane, however, is at approximately 1400h local time, while it does arrive until after 18h in Moranbah. Furthermore, the difference between the PM$_{10}$ concentrations in the morning and afternoon peaks and the trough at midday is considerably smaller in Brisbane than in Moranbah. This suggests a more continual source of PM$_{10}$ throughout the day in Brisbane, but a larger and more continual source of PM$_{10}$ throughout the night in Moranbah. As established earlier, there is a increased trend in PM$_{10}$ in Moranbah between 2011 and 2019. A further trend analysis for different periods of the day shows a much more pronounced, statistically significant (p $<$ 0.001), trend in PM$_{10}$ during night hours (2100h–0900h) than the rest of the day (see Supplementary Fig. 4), while no statistically significant trend was observed for any period of day in Brisbane.

Furthermore, the diurnal cycle in Brisbane is different on weekends, with less pronounced morning and afternoon peaks on Saturday, no morning peak on Sunday, later afternoon peaks, and lower PM$_{10}$ concentrations on average. This is in contrast with Moranbah, which still shows similar concentrations and a prominent morning and afternoon peak on both Saturday and Sunday. This strong diurnal pattern that continues over the weekend strongly suggests the impact that the typical working week and day in Moranbah has on the local air quality. Nearly 40% of the towns population are full-time employed in the surrounding coal mines (Australian Bureau of Statistics, 2016) and work 12 h shifts. It is therefore not surprising that the differences in the typical working week between Moranbah and Brisbane is reflected in the weekly and diurnal cycles of PM$_{10}$ concentrations. Furthermore, the diurnal cycles of PM$_{10}$ in Moranbah strongly indicate the importance of local mining, or mining-related, activity on the air quality in the town. Although some of this contribution could be due to traffic to and from mine sites, the significantly elevated concentrations compared to Brisbane, which has a much higher population and traffic volume, suggests that other non-traffic sources are very important.

Although fires are an important source of PM$_{10}$, there is strong evidence against these being responsible for the increasing trend of PM$_{10}$ in Moranbah. The first is that the meteorological normalisation accounts for PM$_{10}$ due to nearby fires and there has been no increasing trend in fire power since 2011 (see Fig. 2) This is further supported by the fact that, the fire power and temperature predictor variables from the meteorological normalisation did not significantly change the normalised PM$_{10}$ as it did for soil water content. The second is that the increasing trend in PM$_{10}$ is present across all seasons, even when fires are not occurring. Third, the increasing trend in PM$_{10}$ is stronger at night than during the day, which is consistent with mining activity that is operational throughout the night. There have been some suggestions that the poor air quality in Moranbah is due to wood fire heating, however the positive trend is consistent across all wind directions and seasons (Supplementary Figs. 6 and 7) so this is unlikely to be a large contributor.

The Australian National Pollution Inventory records annually reported PM$_{10}$ emissions for the major mine sites across Australia, including those in the Bowen Basin around Moranbah. The numerous mine sites around Moranbah are known to be significant emitters of PM$_{10}$ with annual emissions from the nearest 19 sites on the order of 100 million kilograms (see Fig. 10). With fires and local traffic sources unlikely to be the reason for the rising PM$_{10}$ concentrations in Moranbah, there is no obvious explanation for the rise in the meteorologically normalised PM$_{10}$ other than activity related to these mines.
Fig. 8. The average annual polar annulus of PM$_{10}$ concentration at Moranbah. For each yearly polar annulus, the angle represents the wind direction and the radius represents the local hour of the day (Australian Eastern Standard Time).

Fig. 9. The time variation of PM$_{10}$ in Moranbah and the Brisbane CBD where a) is the average hourly PM$_{10}$ concentrations for each week day, b) is the average hourly PM$_{10}$ concentrations for all days, c) is the average daily concentrations for each week day, and d) is the average monthly PM$_{10}$ concentrations.
Interestingly, the total reported emissions from all of the mine sites were negatively correlated with the annually average normalised PM$_{10}$ trend ($R = -0.79$). Negative correlations were observed for most individual mine sites, with the exception of the Broadlea North Coal Project ($R = 0.75$), the Isaac Plains Coal Mine ($R = 0.16$), Coppabella Coal Mine ($R = 0.17$), Moranbah North ($R = 0.28$), the South Walker Creek Mine Operations ($R = 0.10$) and Grosvenor ($R = 0.43$). It should be noted that the magnitude of mining emissions reported may be a conservative estimate (Hendryx et al., 2020). Although the reported emissions from each mine are only estimates and are only on an annual basis, this might suggest that the mining activity from one individual mine is not responsible for the entirety of increase in the normalised PM$_{10}$ observed in the town of Moranbah, or that the reported annual emissions are not accurate.

Additionally, the estimated non-residential workers in the Isaac region can be considered a proxy for mining-related activity around Moranbah. The correlation was weak between the estimated non-residential population and the measured and normalised PM$_{10}$ ($R^2 = 0.075$ and $R^2 = 0.094$, respectively) between 2011 and 2019. Interestingly, by excluding 2019 which had anonymously high PM$_{10}$ concentrations, the correlation between the estimated non-residential population and the measured PM$_{10}$ was somewhat stronger ($R^2 = 0.13$), while it was significantly stronger for the normalised PM$_{10}$ ($R^2 = 0.37$). Although only a crude comparison due to the limited number of years, this does provide additional, albeit weak, evidence that mining related activity has been responsible for the rise in the normalised PM$_{10}$.

Although the evidence suggests that the sharp changes in PM$_{10}$ emissions and the overall increasing trend are related to mining activity, this analysis is not able to reveal what mining related activities are contributing most to the PM$_{10}$ concentrations in the town. Possible sources are emissions from mining vehicles, blasts to disturb and break up soil, or dust stirred up from exposed soil and/or driving. Compositional measurements will help identify the source type and detailed dispersion modelling would be useful in identifying where the specific sources are.

### 3.5. Future work

There is significant scope to extend on this work. Although the purpose of this study was to use freely available, open-access datasets to investigate the trends and influences of PM$_{10}$ in Moranbah, there is room to improve on the model by considering more predictor variables. With mine sites being an active source of PM$_{10}$, other datasets that can be used as a proxy for emissions from these mines would be valuable. These data could include the frequency of blasting and the number of operational vehicles. The inclusion of this data would could drastically improve the capability to predict PM$_{10}$, while also better resolving the activities that lead to high concentrations. Although emissions inventories exist for each of the mine sites around Moranbah, these are only provided on a yearly basis and therefore not suitable as predictor variables in the meteorological normalisation method.

As of October 2019, the monitoring of PM$_{2.5}$ in Moranbah has commenced. This will further help in the attribution of sources of coarse (e.g. dust) and fine (e.g. vehicle emissions) particles. Once enough data has been gathered (i.e. at least a full year to capture the seasonal cycle), the classical and machine learning techniques applied in this study should be extended to both the PM$_{2.5}$ and PM$_{10}$ data.

The scope of this study was to explore the alarming increase in PM$_{10}$ to above safe levels in Moranbah. Beyond the local and regional area surrounding Moranbah, this meteorological normalisation should be applied to the air quality monitoring datasets across Australia. This could potentially reveal how effective changes in vehicle emissions standards are in major cities or how prescribed burning practices or uncontrolled fires influence the air quality across Australia. 2019 and 2020 have been anomalous years for Australia, starting with the unprecedented fires across southern and eastern Australia, followed by the impact of COVID-19. These fires had severe impacts on the air quality in major cities and bushfire smoke is estimated to have resulted in more than 400 excess deaths (Borchers Arriagada et al., 2020). Further, the impacts of COVID-19 on industry and traffic are likely to cause changes in air quality (Wang et al., 2020). The meteorological normalisation technique applied in this study should be used to isolate the influences of these fires and COVID-19 on air quality in Australia and elsewhere.

### 4. Conclusions

The air quality in Moranbah worsened between 2011 and 2019. PM$_{10}$ concentrations increased by an average 1.2 ±0.5 μg m$^{-3}$ per year ($p < 0.001$). This study used gradient boosted regression and random forest models to investigate the changes in PM$_{10}$ in Moranbah. By accounting for meteorological and environmental effects, the results show that there has been an increase of 0.6 ± 0.3 μg m$^{-3}$ per year due to changes in local emissions, with the remainder mostly due to a decrease.
in soil water content in the region which can facilitate dust emissions. Furthermore, 10 distinct breakpoints in the normalised PM$_{10}$ time series were identified over the study period. Environmental agencies and local industries can use these breakpoints to relate changes in industrial or urban activities to explicitly assess the impacts of these changes on the air quality in Moranbah. The meteorological normalisation using machine learning is proving to be a powerful technique in determining and investigating changes in emissions that lead to improved, or deteriorated, air quality. This study extended on the meteorological normalisation applied in other continents by including a proxy for fire emissions, which have significant impacts on air quality across Australia, as well as soil water content which was shown to be important. The methodology outlined in this study should now be applied to long-term air quality monitoring data sets from urban and rural sites across Australia.

CRedit author contribution statement

Marc Daniel Mallet: Mallet is the sole author. He did all of the CRedit roles.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The author has family members that live in the town that is the focus of this study and employed in local industry. These personal relationships did not influence the analyses or discussion presented in this study in any way.

Acknowledgements

The author would like to acknowledge the Queensland Government Statistician’s Office, the Queensland Government Department of Environment and Science, Geoscience Australia, the Australian Bureau of Meteorology, the European Centre for Medium-Range Weather Forecasts, the National Centers for Environmental Prediction and the National Center for Atmospheric Research for providing all of the datasets used in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apr.2020.08.001.

Data and code availability

All raw data are available from free, publicly available, sources that are outlined in the methods. All code that is used to load, clean, analyse and visualise data, as well as the prepared dataset used for gradient boosted regression and random forest modelling, is available on the public github repository, https://github.com/marc-mallet/moranbah

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alvarado, M., Gonzalez, F., Fletcher, A., Doshi, A., 2015. Towards the development of a low cost airborne sensing system to monitor dust particles after blasting at open-pit mine sites. Sensors 15, 19667–19687.
Bhattachan, A., D odorico, P., 2014. Can land use intensification in the mallee, Australia increase the supply of soluble iron to the southern ocean? Sci. Rep. 4, 1–5.
Borchers Arriagada, N., Palmer, A.J., Bowman, D.M., Morgan, G.G., Jalaludin, B.B., Johnston, F.H., 2020. Unprecedented smoke-related health burden associated with the 2019–20 bushfires in eastern Australia. Med. J. Aust. https://doi.org/10.5969/mja2.50545.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brunskeved, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 1233–1242.
Carslaw, D.C., 2018. Deweather: Remove the Influence of Weather on Air Quality Data. http://github.com/davidcarslaw/deweather. r package version 0.5.
Carslaw, D.C., Hopkins, K., 2012. Openair — a r package for air quality data analysis. Environ. Model. Software 27–28, 32–61. https://doi.org/10.1016/j.envsoft.2011.09.008.
CDS, 2017. Copernicus Climate Change Service (C3S): ERA5: Fifth Generation of ECMWF Atmospheric Reanalysis of the Global Climate. https://cds.climate.copernicus.
Eugridapp/home. (Accessed 23 March 2020).
Chan, Y., Simpson, R., Denison, L., Wong, N., Hodge, M., Comino, E., Carwell, S., 2008. Apportionment of sources of fine and coarse particles in four major Australian cities by positive matrix factorisation. Atmos. Environ. 42, 374–389.
Chan, Y.C., Cohen, D.D., Hawau, O., Stelzer, E., Simpson, R., Denison, L., Wong, N., Hodge, M., Comino, E., Carwell, S., 2008. Apportionment of sources of fine and coarse particles in four major Australian cities by positive matrix factorisation. Atmos. Environ. 42, 374–389.
Cortes-Ramirez, J., Naish, S., Sly, P.D., Jagah, P., 2018. Mortality and morbidity in populations in the vicinity of coal mining: a systematic review. BMC Publ. Health 18, 721.
Cavina, J., Field, J., Taylor, M.P., Gao, S., Landázuri, A., Betterton, E.A., Saez, A.E., 2012. A review on the importance of metals and metalloids in atmospheric dust and aerosol from mining operations. Sci. Total Environ. 433, 58–73.
Dessertvazett, M., Paton-Walsh, C., Griffith, D.W., Kettlewell, G., Keywood, M.D., Vanderschoot, M.V., Ward, J., Mallet, M.D., Milic, A., Miljevic, B., et al., 2017. Emission factors of trace gases and particles from tropical savanna fires in Australia. J. Geophys. Res. B: Atmosphere 122, 6059–6074.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
He, C., Miljevic, B., Crilley, L.R., Surawski, N.C., Bartsch, J., Salimi, F., Uhde, E., 2010. Effect of moisture on fine dust emission from tillage operations on agricultural soils. Earth Surf. Process. Landforms: The Journal of the British Geomorphological Research Group 33, 1851–1863.
Geoscience Australia, 2020. Sentinel Hotspot. https://sentinel.ga.gov.au/#/. (Accessed 22 March 2020).
Ghose, M.K., Majee, S., 2007. Characteristics of hazardous airborne dust around an Indian surface coal mining area. Environ. Monit. Assess. 130, 17–25.
Grange, S.K., Carslaw, D.C., 2019. Using meteorological normalisation to detect interventions in air quality time series. Sci. Total Environ. 653, 578–588.
Grange, S.K., Carslaw, D.C., Lewis, A., Boilei, E., Heuling, C., 2018. Random forest meteorological normalisation models for swiss pm10 trend analysis. Atmos. Chem. Phys. Discuss. 18, 6223–6239.
Graz, J., Ayers, G., 1983. Marine aerosol at southern mid-latitudes. J. Geophys. Res.: Oceans 88, 10661–10666.
Greenwell, B., Boehnke, B., Cunningham, J., Developers, G., 2019. gbm: generalized boosted regression models. https://CRAN.R-project.org/package=gbm. r package version 2.1.5.
Griffin, R.J., Cocker III, D.R., Flagan, R.C., Seinfeld, J.H., 1999. Organic aerosol formation from the oxidation of biogenic hydrocarbons. J. Geophys. Res.: Atmosphere 104, 3555–3567.
He, C., Miljevic, B., Grilley, L.R., SURAWSKI, N.C., BARTCH, J., Salimi, F., Uhde, E., SCHNELL-KREIS, J., ORASCHE, J., RISTOVSKI, Z., et al., 2016. Characterisation of the impact of open biomass burning on urban air quality in Brisbane, Australia. Environ. Int. 91, 230–239.
Hendryx, M., Islam, M.S., Dong, G.H., Paul, G., 2020. Air pollution emissions 2008–2018 from australian coal mining: implications for public and occupational health. Int. J. Environ. Res. Public Health 17, 1570.
Iannone, R., 2019. Split: Use the Hysplit Model from inside R. R package version 0.4.0.9000.
de Jesus, A.L., Thompson, H., Knibbs, L.D., Kowalski, M., Cyrys, J., Niemi, J.V., Konu, A., Timonen, H., Laoma, K., Petaja, T., et al., 2020. Long-term Trends in PM2.5 Mass and Particle Number Concentrations in Urban Air: The Impacts of Mitigation Measures and Extreme Events Due to Changing Climates. Environmental Pollution, p. 114500.
Johnston, F.H., Hanigan, I.C., Henderson, S.B., Morgan, G.G., 2013. Evaluation of interventions to reduce air pollution from biomass smoke on mortality in launceston, Australia: retrospective analysis of daily mortality, 1994–2007. BMJ 346, e4846.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al., 1996. The ncep/nmar 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–472.
Keywood, M., Cope, M., Meye, C.M., Ilinna, Y., Emmerson, K., 2015. When smoke comes to town: the impact of biomass burning smoke on air quality. Atmos. Environ. 121, 13–21.
Perret, J.L., Plush, B., Lachapelle, P., Hinks, T.S., Walter, C., Clarke, P., Irving, L., Liu, X., Tan, W., Tang, S., 2019b. A bagging-gbdt ensemble learning model for city air pollution concentration prediction. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing., 022027.

Mallet, M.D., Dessvertta, M.J., Miljevic, B., Milic, A., Ristovski, Z.D., Alore, J., Cravigan, I.T., Jayaratne, E.R., Paton-Walshe, C., Griffith, D.W.T., Wilson, S.R., Kettlewell, G., van der Schot, M.V., Selleck, P., Reisen, F., Lawson, S.J., Ward, J., Harnwell, J., Cheng, M., Gillett, R.W., Molloy, S.B., Howard, D., Nelson, P.I., Morrison, A.L., Edwards, G.C., Williams, A.G., Chambers, S.D., Werczynski, S., Williams, L.R., Winton, V.H.L., Atkinson, B., Wang, X., Keywood, M.D., 2017. Biomass burning emissions in northern Australia constrained by aerosol measurements: modelling the distribution of hourly emissions. Atmos. Environ. 42, 1629–1646.

Meyer, C.M., Luhar, A.K., Mitchell, R.M., 2008. Biomass burning emissions over northern Australia constrained by aerosol measurements: ‘modelling the distribution of hourly emissions. Atmos. Environ. 42, 1629–1646.

Richardson, C., Rutherford, S., Agranovski, I., 2018. Characterization of particulate emissions from Australian open-cut coal mines: toward improved emission estimates. J. Air Waste Manag. Assoc. 68, 596–607.

Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., Peters, A., 2011. Health effects of particulate air pollution: a review of epidemiological evidence. Inhal. Toxicol. 23, 555–592.

Wang, P., Chen, K., Zhu, S., Wang, Z., Hwang, S., 2020. Impact of smoke from biomass burning on air quality in rural communities in Queensland queensland mining communities. Parity 30, 41.

WHO, 2006. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005: Summary of Risk Assessment. Technical Report. World Health Organization, Geneva.

Wickham, H., Averick, M., Bryan, J., Chang, B., Christensen, T.W., et al., 2019. Welcome to the tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686.

Wise, E.K., Comrie, A.C., 2005a. Extending the Kolmogorov–zurbenko filter: application to ozone, particulate matter, and meteorological trends. J. Air Waste Manag. Assoc. 55, 1208–1216.

Wise, E.K., Comrie, A.C., 2005b. Meteorologically adjusted urban air quality trends in the southwestern United States. Atmos. Environ. 39, 2969–2980.

Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of random forests for high dimensional data. R package version 0.12.1.

Richardson, C., Rutherford, S., Agranovski, I., 2018. Characterization of particulate emissions from Australian open-cut coal mines: toward improved emission estimates. J. Air Waste Manag. Assoc. 68, 596–607.

Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., Peters, A., 2011. Health effects of particulate air pollution: a review of epidemiological evidence. Inhal. Toxicol. 23, 555–592.

Rybacka, Y., Zalakeviciute, R., 2018. Machine learning approaches for outdoor air quality modelling: a systematic review. Appl. Sci. 8, 2570.

Sparks, A.H., Padgham, M., Parsons, H., Pemberton, K., 2017. 2017. National Assessment of Particulate matters-pm10, pms and pm emissions from the heterogeneous traffic near an urban roadway. Atmospheric Pollution Research 1, 184–194.

Stein, A., Draxler, R.R., Rolph, G.D., Stunder, B.J., Cohen, M., Ngn, F., 2015. NaaS/NOAA’s hydaspit atmospheric transport and dispersion modeling system. Bull. Am. Meteorol. Soc. 96, 2059–2077.

Vu, T.V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., Harrison, R.M., 2019. Assessing the impact of clean air action on air quality trends in beijing using a machine learning technique. Atmos. Chem. Phys. 19, 11303–11314.

Walker, H., Heal, M., Braban, C., Ritchie, S., Conolly, C., Sanocka, A., Dragosits, U., Twigg, M., 2019. Changing supersites: assessing the impact of the southern UK emep super site relocation on measured atmospheric composition. Environmental Research Communications 1, 041001.

Wang, P., Chen, K., Zhu, S., Wang, Z., Hwang, S., 2020. Severe air pollution events not avoided by reduced anthropogenic activities during covid-19 outbreak. Resour. Conserv. Recycl. 158, 104814.

Warren, S., McDonald, D., McAuliffe, D., et al., 2017. Homelessness in rural and regional Queensland mining communities. Parity 30, 41.

WHO, 2006. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005: Summary of Risk Assessment. Technical Report. World Health Organization, Geneva.

Wickham, H., Averick, M., Bryan, J., Chang, B., McGowan, L.D., Frana-o, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., MAiller, K., Moons, J., Robinson, D., Seidel, D.P., Spinnu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686.

Wise, E.K., Comrie, A.C., 2005a. Extending the Kolmogorov–zurbenko filter: application to ozone, particulate matter, and meteorological trends. J. Air Waste Manag. Assoc. 55, 1208–1216.

Wise, E.K., Comrie, A.C., 2005b. Meteorologically adjusted urban air quality trends in the southwestern United States. Atmos. Environ. 39, 2969–2980.

Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of random forests for high dimensional data. R package version 0.12.1.