Journal of Hazardous Materials 418 (2021) 126334

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A novel approach to the development of 1-hour threshold concentrations for exposure to particulate matter during episodic air pollution events

Michael E. Deary *, Simon D. Griffiths

Faculty of Engineering and Environment, Department of Geography and Environmental Sciences, University of Northumbria, Ellison Building, Newcastle upon Tyne NE1 8ST, UK

ARTICLE INFO

Editor: Dr. Zaher Hashisho

Keywords:
Particulate matter
Wildfires
Bushfires
Dust storms
AQI
California
Receiver Operating Characteristic
ROC

ABSTRACT

Episodic air pollution events that occur because of wildfires, dust storms and industrial incidents can expose populations to particulate matter (PM) concentrations in the thousands of μg m^{-3}. Such events have increased in frequency and duration over recent years, with this trend predicted to continue in the short to medium term because of climate warming. The human health cost of episodic PM events can be significant, and inflammatory responses are measurable even after only a few hours of exposure. Consequently, advice for the protection of public health should be available as quickly as possible, yet the shortest averaging period for which PM exposure guideline values (GVs) are available is 24-h. To address this problem, we have developed a novel approach, based on Receiver Operating Characteristic (ROC) statistical analysis, that derives 1-h threshold concentrations that have a probabilistic relationship with 24-h GVs. The ROC analysis was carried out on PM_{10} and $PM_{2.5}$ monitoring data from across the US for the period 2014–2019. Validation of the model against US Air Quality Index (AQI) 24-h breakpoint concentrations for PM showed that the maximum-observed 1-h PM concentration in any rolling 24-h averaging period is an excellent predictor of exceedances of 24-h GVs.

1. Introduction

In this paper we present a novel approach to the development of 1-h threshold concentrations (TCs) for exposure to particulate matter (PM) during episodic air pollution events, as might occur during wildfires (Rappold et al., 2017) dust storms (Milford et al., 2020, Zhang et al., 2016, Rublee et al., 2020) or incidents at industrial facilities (Griffiths et al., 2018). Populations exposed to episodic air pollution events can experience PM concentrations in the hundreds and even thousands of μg m^{-3} (Griffiths et al., 2018). Our approach uses a model that is developed using Receiver Operating Characteristic (ROC) statistical analysis of ambient monitoring data from the US over the period 2014–2019. The development of 1-h TCs is needed because health effects of elevated PM exposure are apparent at a timescale of hours, as evidenced by measurable inflammatory responses for volunteers exposed to PM in the 100–300 μg m^{-3} range, over short durations (Behndig et al., 2006, Tong et al., 2014, Ghio et al., 2000, Salvi et al., 1999, Stenfors et al., 2004). Short term (hours) health effects have also been noted in fire fighters (Greven et al., 2012, Swiston et al., 2008, Main et al., 2020). Nevertheless, unlike nitrogen dioxide and sulphur dioxide, which have health-based GVs for exposure durations of 1-h or less (US EPA, 2014, WHO, 2006b), and many chemical substances for which there are Acute Exposure Guideline Levels (AEGLs) for periods as short as 10 min (Stewart-Evans et al., 2016), no such values are available for PM_{10} and $PM_{2.5}$.

The need for the development of short-term PM exposure guidance has become more pressing in recent years because periods of highly elevated PM concentrations have increased in frequency, duration and extent, especially during wildfires (Balmes, 2018, Dodd et al., 2018, Ford et al., 2018, Howard et al., 2021). These events are responsible for causing significant ill health effects (Reid et al., 2016, Haikerwal et al., 2015, Faustini et al., 2015, Black et al., 2017, Cascio, 2018), particularly in the more vulnerable residents of an exposed area (Finlay et al., 2012, Liu et al., 2015, Holm et al., 2020, Wakefield, 2010). In addition, the toxicity of particulate emissions during combustion-related episodic pollution events has been found to be higher than for equivalent concentrations of ambient particulates (Wegesser et al., 2009). This enhanced toxicity is due to the wide range of chemical toxins present in PM that originate from combustion processes, including PAHs and benzene (Balmes, 2018, Wegesser et al., 2009). The incidence of wildfires, globally, is predicted to increase in the medium term as a result of a warming climate (Moritz et al., 2012) and there is also evidence that

* Corresponding author.
E-mail address: michael.deary@northumbria.ac.uk (M.E. Deary).

https://doi.org/10.1016/j.jhazmat.2021.126334
Received 17 March 2021; Received in revised form 22 May 2021; Accepted 3 June 2021
Available online 6 June 2021

fires at waste management sites are more frequent during warmer conditions (Griffiths et al., 2018). The health impacts of such changes may be considerable, with a recent study suggesting that premature deaths due to $PM_{2.5}$ exposure during wildfires in the US alone could increase from the current 17,000 per year to 42,000 per year by 2050 (Ford et al., 2018). The associated health related economic costs are also expected to be significant (Johnston et al., 2020, Kochi et al., 2016).

The absence of short-term GVs for PM_{10} and $PM_{2.5}$ has been acknowledged in the literature and there have been several studies that have derived surrogate exposure guidance for periods as short as one hour (Griffiths et al., 2018, Stieb et al., 2008, Mintz et al., 2013, European Union, 2020b, Connolly and Willis, 2013). A common theme to these approaches has been the relationship between the maximum hourly concentration within a 24-h period and the corresponding mean value. A notable example is the European Union's (EU) Common Air Quality Index (CAQI), which has five classes ranging from 'Very Low' to 'Very High', each with corresponding concentration thresholds for PM_{10} and $PM_{2.5}$, both for 1-h and 24-h measured concentrations (European Union, 2007). The 1-h thresholds between the CAQI categories for PM_{10} were derived from the 24-h limits by dividing the latter by a factor of 0.55, which is the ratio between the mean 24-h concentration and the maximum hourly concentration within the same period. The ratio of 0.55 is based on European ambient monitoring data from 52 urban monitoring stations for the period 2001–2004 (European Union, 2007). Thus, for the 24-h PM_{10} category boundary between 'medium' and 'high' (set at 50 μg m^{-3}, which is the same as the EU/WHO 24-h ambient guideline value) the calculated 1-h category boundary is set at 90 μg m^{-3} (after rounding). For $PM_{2.5}$, the class boundaries are based on those of PM_{10}, applying a factor of 0.6, which is the fraction of PM_{10} that is $PM_{2.5}$, again based on European monitoring data (European Union, 2020b). Stieb et al. (2008) employed a similar approach to CAQI in determining short term (3 h) TCs. The same ratio as for the CAQI data, 0.55, was observed between the 24-h mean concentration and the 3-h maximum concentration for PM_{10} or $PM_{2.5}$ (based on monitoring data collected in Canada over the period 1998–2000). Stieb et al. (2008) illustrated their approach using the US Air Quality Index (AQI) boundary between 'moderate' and 'unhealthy for sensitive groups' (AQI = 100), at which the corresponding 24-h PM_{10} sub-index guideline value of 150 μg m^{-3} has an equivalent 3-h TC of 275 μg m^{-3}.

In the UK, the Department of Environment, Food and Rural Affairs (DEFRA) have derived 1-h 'trigger' concentrations as a component of the UKs Daily Air Quality Index (DAQI) (Connolly and Willis, 2013, Holgate, 2011). The trigger concentrations establish a relationship between 1-h measurements and the 24-h mean concentration ranges that correspond to the DAQI air pollution categories of 'low', 'moderate', 'high' or 'very high'. Under the DAQI methodology, if two consecutive 1-h measurements breach a 'trigger' concentration, this is taken to indicate that current air quality falls within the relevant DAQI category, thus providing a 'real-time' element to public information about air pollution levels in the UK. The trigger concentrations were derived using a categorical model based on 270,000 days of PM_{10} data and 27,000 days of $PM_{2.5}$ data from automatic monitoring stations across the UK for the period 2004–2009 (Holgate, 2011).

Our own work in this field has been to develop a similar model to DAQI, though one that is based on the higher concentration ranges that are observed during major incident fires (Griffiths et al., 2018). The model uses 1-h PM_{10} and $PM_{2.5}$ measurements to predict exceedances of 24-h guideline and threshold concentrations that relate to public health advice during such incidents. The model development was based on monitoring data obtained from the UK's Air Quality in Major Incidents (AQiMI) programme (Griffiths et al., 2018), which coordinates field monitoring of a range of atmospheric pollutants arising from fires and loss of containment incidents at industrial facilities and waste disposal sites in the UK. Both authors of the present paper were involved in the AQiMI programme (Griffiths et al., 2018). The model demonstrated that there is a threshold concentration of 1-h measured PM that, when breached, gives a defined probability that a 24-h guideline value is also likely to be exceeded.

Other approaches to providing more responsive information on ambient PM exposure include: remote sensing (Krstic and Henderson, 2015), real-time dispersion modelling, e.g. the BlueSky wildfire smoke forecasting service used in Canada (Yao et al., 2013), predictive models based on autoregression neural networks (Videnova et al., 2006), non-linear models using 'big data' (Xu et al., 2020), textual analysis of social media postings (Sachdeva and Mccaffrey, 2018) and the US EPA Nowcast methodology which contributes to the US AirNow forecasting service (Mintz et al., 2013). The latter uses short term monitoring data (the previous 12 h) to predict an equivalent 'instantaneous' 24-h PM concentration, which can then be compared to various health criteria. This method is designed to be responsive at times of rapidly changing pollution conditions. It does this by giving greater weightings to the three most recent hours of air pollution data at times when the air quality is very variable but gives more equal weighting to the previous 12 h of air pollution data when pollution concentrations are more stable.

In the present paper we apply a novel approach based on Receiver Operating Characteristic (ROC) analysis (Fawcett, 2006) of US EPA data for the period 2014–2019 to develop a probabilistic model from which 1-h TCs can be derived for PM_{10} and $PM_{2.5}$. The US Air Quality Index (AQI), and associated breakpoint concentrations for the PM sub-indices, were used as the source of GVs, as they represent a wide range of pollutant concentrations (0–500 μg m^{-3} for $PM_{2.5}$ and 0–605 μg m^{-3} for PM_{10}) and are appropriate to the US monitoring data.

2. Method

2.1. Principles of ROC analysis, as applied to air quality data

The ROC model development work described in this paper was carried out as follows. Firstly, we built a model based on ROC analysis of PM_{10} and $PM_{2.5}$ ambient concentration measurements from across the US for the period 2014–2019. This allowed us to derive 1-h TCs that gave defined probabilities that selected 24-h GVs would be exceeded. We then evaluated the performance of the model using a cross-validation approach and also carried out a separate evaluation of an ROC model that was developed using monitoring data from California only.

ROC analysis is a classification metric that analyses the ability of a predictive parameter, or 'classifier', to discriminate between two outcomes. It has been widely used for health-related diagnostic analysis (Hajian-Tilaki, 2013, Phillips et al., 2010), for example to identify serum biomarkers for PM_{10} exposure (Lee et al., 2015). In the environmental field, ROC analysis has been applied to the development of a model to predict the quality of beach water for swimming, based on either the previous day's rainfall or bacterial counts (Morrison et al., 2003). We used a variant of this analysis when developing our original model for AQiMI data (Griffiths et al., 2018). For that work, and in the present study, the classifier is the maximum 1-h concentration in a rolling 24-h period. The outcome is whether (or not) the mean concentration of the corresponding 24-h period exceeds a selected 24-h GV. The value of the GV can be selected from a range of health-based values developed by the WHO, US EPA, EU and other bodies.

The utilisation of ROC analysis is illustrated in Fig. 1, which shows a typical output in which the true positive rate (TPR, or sensitivity) is plotted against the false positive rate (FPR, or 1 - specificity) (Fawcett, 2006). TPR and FPR are defined in Eqs. (1) and (2) respectively, where for a given set of analysed data, TP is the number of true positives (i.e. the model correctly predicts an exceedance of a 24-h guideline value), FN is the number of false negatives (the model incorrectly predicts that the 24-h guideline value is not exceeded), FP is the number of false positives (the model incorrectly predicts an exceedance of a 24-h guideline value) and TN is the number of true negatives (the model correctly predicts that the 24-h guideline value is not exceeded).

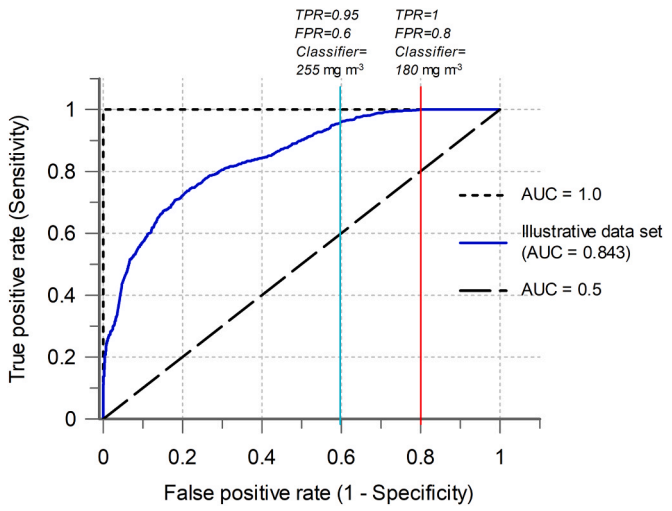

Fig. 1. Example ROC plot. The solid blue line shows an ROC curve for the analysis of an illustrative set of PM_{10} data where the 'outcome' is whether (or not) a defined 24-h guideline value has been exceeded. The 'classifier' is the maximum 1-h concentration within the same 24-h period. The dotted line shows an example of an ROC analysis where the 'classifier' correctly predicts 100% of outcomes (area under the curve, AUC = 1.0), whereas the dashed line shows an example curve where only 50% of outcomes are correctly predicted by the classifier (AUC = 0.5). The vertical lines indicate different selections of TPR, together with the corresponding FPR. The optimal situation is to have a TPR as close as possible to 1, whilst minimising the FPR. Also shown is the numerical value of the 'classifier' that corresponds to the selected TPR/FPR.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (1)$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \quad (2)$$

The area under the curve (AUC) for the solid line in Fig. 1 is an important parameter for ROC analysis, representing the overall probability that the chosen classifier parameter will rank a randomly chosen true positive instance above a randomly chosen true negative instance (Fawcett, 2006). For our illustrative dataset the AUC is 0.843, meaning that the value of the maximum 1-h concentration in a 24-h period correctly determines exceedances of the 24-h GV in 84.3% of cases.

Two other example curves are shown in Fig. 1: a dotted line with an AUC of 1.0 (100% probability of distinguishing between two outcomes) and a diagonal, dashed, line with an AUC of 0.5, which means there is only a 50% probability of correctly discriminating between two outcomes, i.e. no better than chance. Curves that appear below the diagonal represent situations where the model classifier is giving rise to a reciprocal classification.

Fig. 1, together with the AUC value, is useful for visualising the overall performance of the model. It can also be used to decide on the acceptable TPR that will at the same time minimise the FPR, as shown by the red and blue vertical lines, and their relationship with the solid curve. For example, we could decide it is essential that all true positive values are correctly identified (red line) and have to accept an 80% false positive rate, or we could compromise on a lower level of true positive identification, which has the advantage of a lower false positive rate, as shown by the blue line, where TPR = 0.95 and FPR = 0.60. For each selected TPR, there is an associated value of the classifier (the maximum 1-h concentration in any 24-h period). Thus, for the PM_{10} data shown in Fig. 1, the value of classifier concentration that will give a true positive rate of 0.95 (and a false positive rate of 0.60) is 255 μg m^{-3}, whereas to achieve a true positive rate of 1.0 (and a false positive rate of 0.80), we would need to lower the threshold to 180 μg m^{-3}. The value of the classifier at a given TPR is, de facto, a 1-h TC, that has a probabilistic link to the selected 24-h GV.

Regarding selection of an appropriate TPR/FPR value, the decision must be made on the basis of an acceptable scale of risk that balances public health protection against the resources that are available for incident response. The advantage of our ROC analysis approach to defining 1-h TCs is the ability to specify the probabilities on which these decisions are made.

It is important to emphasise that because we use the maximum 1-h concentration (in any 24-h period) as the classifier parameter in the ROC analysis, and that this value can occur at any position in a rolling 24-h period, the results obtained from the models on which this analysis is based must have an equivalent interpretation. In other words, a one-hour TC could be triggered at the beginning middle or end of the 24-h period that is predicted to exceed the relevant 24-h GV. A preliminary analysis using SPSS showed that the median position for the maximum 1-h concentration was at hour 12 for both PM_{10} and $PM_{2.5}$, as might be expected.

2.2. *Initial model development using ROC analysis of PM_{10} and $PM_{2.5}$ measurements from monitoring stations across the US.*

Pre-generated data files of hourly PM_{10} (US EPA parameter code 81102) and $PM_{2.5}$ (US EPA parameter code 88101) concentration data for the years 2014–2019, from ambient monitoring stations across the whole of the US, were downloaded in Comma Separate Values (CSV) file format from the US Environmental Protection Agency (US EPA) website (US EPA, 2021). The files were imported into Microsoft Access for further analysis. Table 1 summarises the number of monitoring stations in each state that were used in this study to provide 1-h measurement data for PM_{10} and $PM_{2.5}$ concentrations. The stations form part of a larger network of monitoring stations within these states. Additionally, Table S1 in the Supplementary material summarises the PM monitoring methods employed by the selected monitoring stations for the years 2014–2018. For these stations, beta attenuation monitoring was by far the most often used technique, but the other methods included Tapered Element Oscillating Microbalance (TEOM), Filter Dynamics Measurement System (FDMS) with TEOM, laser light scattering and broadband spectroscopy.

Pre-generated data files are available from the US EPA website in a format that has the null-data lines stripped out, whereas we required contiguous datasets at each monitoring station to calculate the rolling 24-h averages that are required for the ROC analysis. We reconstructed contiguous datasets for each year, with null values reinstated, using an SQL query in Microsoft Access based on a contiguous hourly time series for that year together with the pre-generated US EPA data file. Subsequently, these reconstructed hourly datasets of PM_{10} and $PM_{2.5}$ concentrations at each monitoring station were analysed in Excel after exporting as a CSV file from Microsoft Access.

For each year and each monitoring station, the data files were prepared for ROC analysis in SPSS using Microsoft Excel as follows. Firstly, we calculated the average concentration of PM_{10} and $PM_{2.5}$ for each rolling 24-h period for which there were at least 20-h of valid measurements (the average concentration is denoted $C_{24(i)}$, where i represents an individual rolling 24 h period). As a QC check on the data, Excel formulae were used to identify year-beginnings/ends, to ensure that there was no crossover analysis of data from different monitoring stations; and to flag up any non-contiguous data sequences. Secondly, we identified the maximum hourly concentration recorded in each rolling 24-h period, $C_{\mathrm{max24(i)}}$. The data was analysed in batches so that the maximum number of rows in Excel was not exceeded (ca. 950,000 for the 2016, 64-bit version of Excel used). It should be noted that there were no selection/exclusion criteria regarding the nature of any exceedances, i.e. whether they originated from wildfires, industrial emissions, dust storms, or as a result of unusual meteorological conditions.

The 24-h GVs used in the ROC modelling were the US AQI category

Table 1
Summary of the number of sites recording hourly PM concentrations across US states in 2019 (US EPA parameter codes 81102 and 88101 for PM_{10} and $PM_{2.5}$ respectively). Missing states had no monitoring records in the US EPA dataset for the specified parameter codes. The hourly measurement sites form part of a larger PM monitoring network in each state.

State	PM_{10} monitoring sites	$PM_{2.5}$ monitoring sites	State	PM_{10} monitoring sites	$PM_{2.5}$ monitoring sites
Alabama	4	3	Montana	11	15
Alaska	8	6	Nebraska	2	3
Arizona	43	18	Nevada	20	14
California	53	40	New Hampshire	1	7
Colorado	4	15	New Jersey	–	12
Connecticut	8	9	New Mexico	13	12
Delaware	–	6	New York	–	7
District Of Columbia	1	5	North Carolina	9	17
Florida	21	18	North Dakota	4	5
Georgia	2	8	Ohio	6	15
Hawaii	3	14	Oklahoma	5	13
Idaho	6	2	Oregon	–	3
Illinois	2	17	Pennsylvania	13	31
Indiana	3	16	Rhode Island	–	6
Iowa	1	5	South Carolina	4	5
Kansas	7	6	South Dakota	6	7
Kentucky	2	15	Tennessee	3	18
Louisiana	4	2	Texas	–	25
Maine	–	9	Utah	–	17
Maryland	–	11	Vermont	–	5
Massachusetts	–	14	Virginia	–	4
Michigan	5	9	Washington	5	20
Minnesota	8	23	West Virginia	2	2
Mississippi	1	8	Wisconsin	7	16
Missouri	9	13	Wyoming	16	11

boundary concentrations for: 'Unhealthy for sensitive groups' (35.5 μg m^{-3} and 155 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively; AQI = 100); 'Unhealthy' (55.5 μg m^{-3} and 255 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively; AQI = 150); and 'Very unhealthy' (150.5 μg m^{-3} and 355 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively; AQI = 200) (US EPA, 2014). There is a higher category boundary of 'hazardous' (250.5 μg m^{-3} and 425 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively; AQI = 300), however due to the limited number of measurements at these high concentrations, it was not used in the initial model development or validation. Nevertheless, 1-h TCs were calculated for this category in a separate analysis.

ROC analysis was performed using IBM SPSS Statistics Version 26. In total, across all monitoring stations for all six years, there were in excess of 16 million rolling 24-h periods for PM_{10} and 22 million rolling 24-h periods for $PM_{2.5}$. For each of the two PM size fractions, individual Excel files comprising the columns $C_{24(i)}$, $C_{max24(i)}$, 'State Code' and 'Year' were imported separately into SPSS and then combined into one SPSS data file using the 'Merge Files' feature. The final step in preparing the data for ROC analysis was to use the 'Recode to Different Variables' SPSS function to generate binary identity variables (denoted 'State Variable' in SPSS ROC terminology, though note that this is not related to the 'State Code' parameter) for each rolling 24-h period. The identity variables, $E_{GV(i)}$, indicated whether or not a rolling 24-h period exceeded a specific 24-h GV (1 = exceeded; 0 = not exceeded). Thus, for PM_{10} the following variables, based on US AQI breakpoint concentrations for 'Unhealthy for Sensitive Individuals', 'Unhealthy' and 'Very Unhealthy' were generated: $E_{155(i)}$, $E_{255(i)}$ and $E_{355(i)}$. For $PM_{2.5}$ the corresponding US AQI variables were $E_{35.5(i)}$, $E_{55.5(i)}$ and $E_{150(i)}$. Additional $E_{GV(i)}$ values were also generated for ROC analysis corresponding to other 24-h GVs.

ROC analysis was performed on $C_{max24(i)}$ (the 'classifier') against each of the $E_{GV(i)}$ values (as state variables) using the SPSS 'ROC Curve' analysis option. Outputs included the ROC curve, the AUC value, and the *p*-value for the model. Additionally, an option to display the coordinate points of the ROC curves (TPR and FPR) was selected, allowing the identification of specific TCs that would give true positive rates of 100% 99%, 95% and 90% for each of the selected 24-h GVs.

We were also interested in quantifying the probabilities associated with the EU (European Union, 2020a) and Stieb et al. (2008) approach to generating 1-h TCs, i.e. the division of the 24-h guideline by a factor of 0.55 (the ratio of $C_{24(i)}$: $C_{max24(i)}$). This was done by reading from the ROC output table, the corresponding TPR and FPR for the calculated 1-h TC. In addition, we repeated this analysis for $C_{24(i)}$: $C_{max24(i)}$ ratios calculated from the US EPA dataset used in the current study.

2.3. Validation

The ROC probability approach to deriving 1-h TCs was validated using a cross-validation design (Schaffer, 1993). The data was grouped according to year and, successively, each individual year's data was used as a test set for an ROC model based on the remaining five years of data (the training set). This involves running an ROC analysis independently on the test set and on the training set and comparing predicted probabilities (set at 100%, 99%, 95% and 90%) with actual TPRs. By repeating this process for each of the six years, a statistical assessment of the inter-year model variability can be made. Predicted and actual FPRs were similarly obtained.

We carried out an additional model development and validation in which data only from the US state of California was used as the training set (all years) and the remainder of data from the other US states (all years), used as the test set. California has the greatest number of monitoring stations of all US states (see Table 1) and also experiences a large number of wildfires (Rappold et al., 2017). Therefore, the rationale for this separate validation was to see whether a model based on this profile is transferable to other states, where sources of elevated PM may differ, for example the dust storms in Arizona and New Mexico (Hyde et al., 2018, Raman et al., 2014).

3. Results and discussion

3.1. Exceedance statistics for data used in the ROC analysis

ROC analysis was used to evaluate the effectiveness of the C_{max24} classifier to predict exceedances of selected 24-h health-based GVs for exposure to PM. The GVs used were the breakpoint 24-h concentrations for the PM sub-indices of the US AQI. These values have been developed

using epidemiological evidence (Federal Register, 2013). Specifically, an AQI value of 100 represents the boundary at which air quality might be considered unhealthy (for sensitive groups) and for which the PM sub-indices are set at the US 24-h ambient exposure GV (35.5 μg m^{-3} and 155 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively) (Federal Register, 2013, Perlmutt and Cromar, 2019). The top of the AQI scale is set at a value of 500, with the corresponding 24-h PM sub-index values based on concentrations estimated during historical wintertime pollution episodes in London, and their associated health effects (500 μg m^{-3} and 604 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively) (Federal Register, 2012). The other US AQI 24-h breakpoint PM concentrations are set as follows: the 'Unhealthy for sensitive groups'/ 'Unhealthy' breakpoint concentration (AQI = 150) is set proportionately to AQI = 100 (55.5 μg m^{-3} and 255 μg m^{-3} for $PM_{2.5}$ and PM_{10} respectively), whilst for AQI values above 150, the breakpoint concentrations are set at relatively uniform increments up to AQI = 500.

As part of the process of compiling monitoring data for the model development and validation stages of this paper, some overall summary data on exceedances of various health-based 24-h GVs were produced. Fig. 2 shows the relative proportion of exceedances, by state, of selected 24-h GVs, for both PM_{10} and $PM_{2.5}$. Specific numbers of exceedances, by state, are detailed in Tables S2 and S3 in the Supplementary material. In addition, Tables S4 and S5 show an overall statistical summary of PM_{10} and $PM_{2.5}$ concentrations (number of measurements, arithmetic average, standard deviation and maximum) for the US EPA datasets used in this study over the period 2014–2019.

For each state, the number of exceedances of any GV will be a function of both the number of monitoring sites that are in operation for PM_{10} and $PM_{2.5}$ and the propensity for acute air pollution events. For example, California has by far the greatest number of PM_{10} and $PM_{2.5}$ monitoring sites for the two US EPA parameter codes that we used (see Table 1) but also has a large number of wildfires each year (Rappold et al., 2017), hence a large number of exceedances across the range of GVs. Thus, for the AQI 'unhealthy' and AQI 'very unhealthy' 24-h GVs (the two highest 24-h GVs used in this analysis for each PM fraction), California contributes more exceedances than for all the other US states combined. For AQI 'unhealthy for sensitive groups' (150 μg m^{-3} and 35.5 μg m^{-3} for PM_{10} and $PM_{2.5}$ respectively), California contributes just under half of the exceedances.

For PM_{10}, the other states that have high numbers of exceedances of the 24-h exposure GVs include Arizona and New Mexico, which have frequent dust storms (Hyde et al., 2018, Raman et al., 2014) but also have a relatively high number of monitoring stations (see Table 1). For $PM_{2.5}$, Montana and Washington make the greatest contributions to exceedances after California, reflecting the relatively large number of $PM_{2.5}$ monitoring stations in each of these states, but also the relatively high incidence of wildfires (Fann et al., 2018, Rappold et al., 2017).

Finally, a clear trend from Fig. 2, particularly for $PM_{2.5}$, is that for the lower 24-h GVs there is a far greater diversity of states contributing exceedance data. In contrast, for AQI 'Unhealthy', for both PM_{10} and $PM_{2.5}$, only six states recorded exceedances during 2014–2019 (California, Arizona, New Mexico, Nevada, Washington and Alaska for PM_{10}; California, Washington, Montana, Colorado, Indiana and Arizona for $PM_{2.5}$). The relative contributions of different states to the total number of exceedances used in the ROC analysis, and the reasons for this, i.e. source and relative numbers of monitoring stations, is an important context for the overall model development and results.

3.2. ROC model performance

Fig. 3 shows the results of the ROC analysis for PM_{10} and $PM_{2.5}$ using data from all US states for the period 2014–2019, together with the respective AUC values. Published criteria for evaluating the performance of ROC, based on the value of AUC, indicates that values above 0.7 could be considered 'good', whereas values greater than 0.8 could be considered 'very good' and those higher than 0.9, 'excellent' (Bekkar et al., 2013). A more stringent interpretation of AUC is given by Zhu et al. (2010), who set the boundary for 'good' at 0.8. For our ROC analysis, the AUC values are all above 0.99, demonstrating that C_{max24} is an excellent classifier parameter for determining whether selected 24-h GVs will be exceeded.

Tables 2 and 3 show, for PM_{10} and $PM_{2.5}$ respectively, the values of

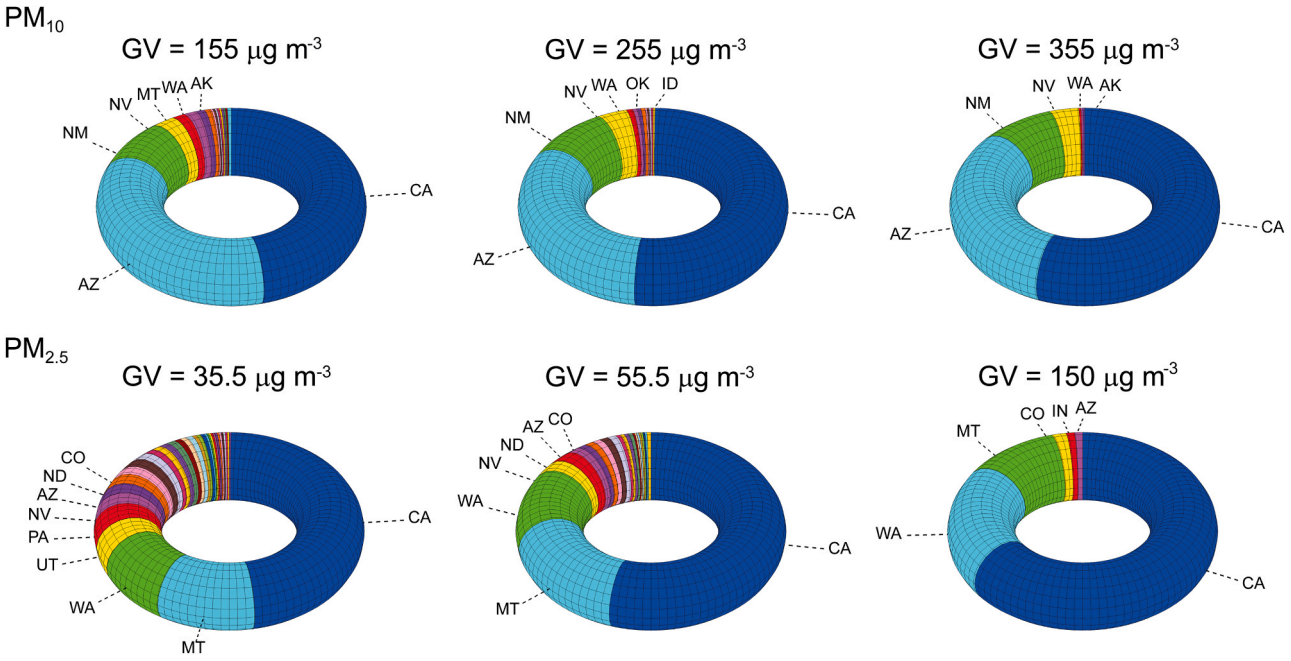

Fig. 2. Relative proportions, by state, of exceedances of specified 24-h GVs for PM_{10} and $PM_{2.5}$ over the period 2014–2019. From left to right the GV values represent the breakpoint concentrations for 'Unhealthy for Sensitive Groups', 'Unhealthy' and 'Very Unhealthy' respectively. Plot labels indicate the states that make the greatest contribution to the overall number of exceedances of each GV. State abbreviations are as follows: AK, Alaska; AZ, Arizona; CA, California; CO, Colorado; ID, Idaho; IN, Indiana; MT, Montana, NV, Nevada; NM, New Mexico; ND, North Dakota; OK, Oklahoma; PA, Pennsylvania; UT, Utah and WA, Washington.

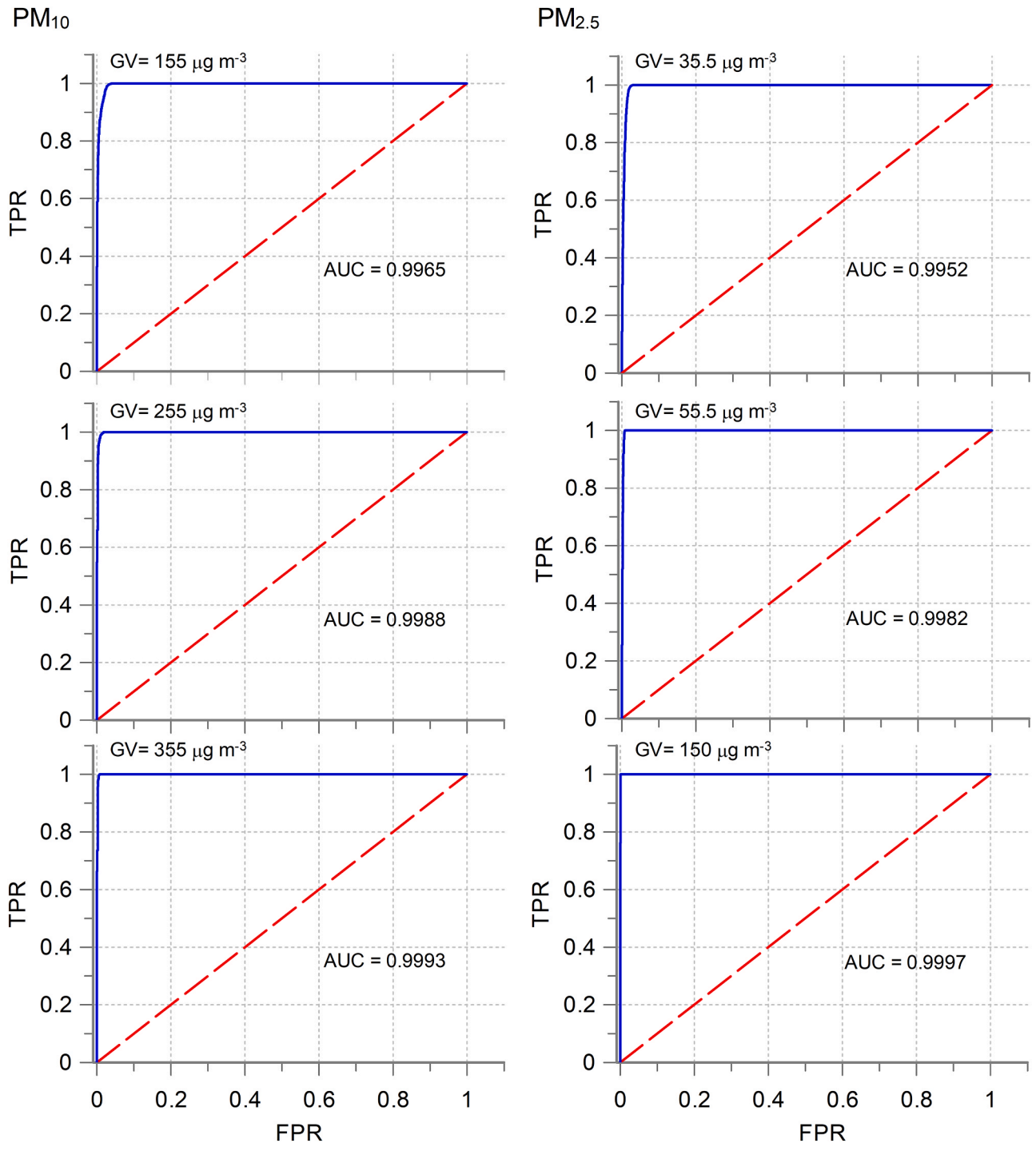

Fig. 3. ROC curves (blue line) at three different 24-h GVs for PM_{10} and $PM_{2.5}$. Also shown is the AUC value. The diagonal red, dashed, line corresponds to ROC curve with an AUC of 0.5.

C_{max24} that will achieve TPRs of 90%, 95%, 99% and 100% for the selected 24-h GVs. For clarity, we have denoted this value as $C_{max24(TPR)}$. The predicted FPR is also shown. The values of $C_{max24(TPR)}$ and FPR in Tables 2 and 3 were obtained from the ROC analysis output table for each corresponding TPR.

$C_{max24(TPR)}$ is, in effect, a 1-h TC that has a probabilistic link to a 24-h GV. Thus, from Table 2 we see that for PM_{10} at the AQI 'unhealthy' breakpoint concentration (255 μg m^{-3}), setting a 1-h TC of 305 μg m^{-3} is likely to result in the prediction of all exceedances of the 24-h GV. The associated FPR of 1.76% means that for each monitoring station, on average, there are likely to be ca. 154 rolling 24-h periods each year for which a breach of the 24-h GV would be falsely predicted. Accepting a lower TPR, e.g. 99% (1 h TC = 427 μg m^{-3}) or 95% (1 h TC = 708 μg m^{-3}), reduces the FPR considerably (to 1.00% and 0.43% respectively). The results from the ROC analysis for other PM_{10} 24-h guidelines in Table 2 can be similarly interpreted. Likewise for the $PM_{2.5}$ 24-h GVs in Table 3.

The nature of elevated air pollution events means that false positives are likely to come in clusters, as illustrated by Fig. 4, which shows an analysis, by date and site, of FPs associated with the $PM_{2.5}$ 24-h GV of 150 μg m^{-3} for 2018. Across 40 sites at which FPs were recorded, the FPs are clustered around dates corresponding to the Carr Fire in July/August (Lareau et al., 2018, Wong et al., 2020) and the Woolsey and Camp Fires

Table 2
Results for the overall ROC analysis for PM_{10}, based on over 16 million rolling 24-h periods from monitoring sites across the US. $C_{max24(TPR)}$ is the value of C_{max24} that will achieve true positive rates (TPRs) of 90%, 95%, 99% and 100% for the selected 24-h GVs. The predicted false positive rate (FPR) is also shown. $C_{max24(TPR)}$ can be considered to be a 1-h TC.

	GV = 155 μg m^{-3} (AQI >100: Unhealthy for sensitive groups)		GV = 255 μg m^{-3} (AQI >150: Unhealthy)		GV = 355 μg m^{-3} (AQI >200: Very unhealthy)	
Selected TPR	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%
100%	186	3.89	305	1.76	561	0.68
99%	227	2.72	427	1.00	707	0.46
95%	281	1.85	708	0.43	984	0.25
90%	370	1.11	890	0.27	–	–

Table 3
Results for the overall ROC analysis for $PM_{2.5}$, based on over 22 million rolling 24-h periods from monitoring sites across the US. $C_{max24(TPR)}$ is the value of C_{max24} that will achieve true positive rates (TPRs) of 90%, 95%, 99% and 100% for the selected 24-h GVs. The predicted false positive rate (FPR) is also shown. $C_{max24(TPR)}$ can be considered to be a 1-h TC.

	GV = 35.5 μg m^{-3} (AQI >100: Unhealthy for sensitive groups)		GV = 55.5 μg m^{-3} (AQI >150: Unhealthy)		GV = 150.5 μg m^{-3} (AQI >200: Very unhealthy)	
Selected TPR	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%	$C_{max24\ (TPR)}$ / μg m^{-3}	Predicted FPR/%
100%	41	2.89	65	0.83	178	0.09
99%	45	2.02	70	0.67	183	0.08
95%	50	1.48	77	0.50	189	0.07
90%	54	1.16	82	0.42	202	0.06

in November (Keeley and Syphard, 2019, Wong et al., 2020). These are also the dates at which most TPs were recorded. Nevertheless, for each site, there are long periods of the year where no FPs are recorded.

Decisions on the value of TPR that is acceptable depend on the proposed public health response to predicted exceedances of the 24-h value. For example, if the response is to send advisory health warnings to the affected population through media sources (Kochi et al., 2016, Mott et al., 2002), or to display the information through a mobile phone App, then the resource implications are clearly much less significant than if physical measures were taken to evacuate people (Stares et al., 2014, Wong et al., 2020). A higher FPR is likely to be tolerated in the former case, though there is a concern that too many false alarms will undermine confidence in the public health response system that is in place. Nevertheless, the results of the analysis in Fig. 4, show that FPs are associated with specific events, and so FPs might not necessarily be registered by the public as false alarms. An additional aspect of the statistics on FPs, as shown in Tables 2 and 3, is that lower FPRs are associated with higher 24-h GVs, i.e. where public health advice is most needed.

3.3. Model validation

Fig. 5 shows the results of the cross-validation study that was carried out on the overall dataset. There is excellent agreement between the predicted (red line) and observed (black line) TPRs, with the observed performance for several of the 24-h GVs exceeding that of the predicted. Nevertheless, we see that the standard deviation of the mean observed TPR values increases for lower values of predicted TPR, i.e. at 90% and 95%; this is because there are far fewer data points at the higher values of C_{24} and C_{max24} that are associated with these lower TPRs. Consequently, predictions made at TPRs of 95% and 90% are less reliable, though still likely to capture over 80% of exceedances for any particular year. The full set of cross-validation data for PM_{10} and $PM_{2.5}$ is available in Tables S6 and S7, respectively.

We also investigated the predictive performance of an ROC model based only on data from California. The results of this validation study are shown in Fig. 6, with the full dataset, including FPRs, available in Tables S8 and S9 for PM_{10} and $PM_{2.5}$ respectively. For PM_{10}, at all the 24-h GVs, the observed TPRs track above the predicted curves. The agreement between predicted and observed TPRs is particularly noteworthy because the model developed for California was based mainly on exceedances due to wildfires, whereas for the validation set, other sources such as dust storms were major reasons for exceedances, notably the large number of exceedances contributed by Arizona and New Mexico (see Fig. 2). The observed FPRs are also somewhat lower than the predicted FPRs. For $PM_{2.5}$, the agreement is less good, with the curve for the observed TPRs tracking below that of the predicted; this is especially notable for the AQI 'very unhealthy' 24-h GV (150 μg m^{-3}). The derived 1-h TCs at each probability level (Tables S6 and S7) are numerically very similar to those derived for the whole model, though for PM_{10}, there are some deviations at the 95% and 90% levels.

An additional piece of confirmation work for our ROC model would be to examine the statistical relationship between exceedances of 1-h TCs and selected health endpoints taken from population-based health

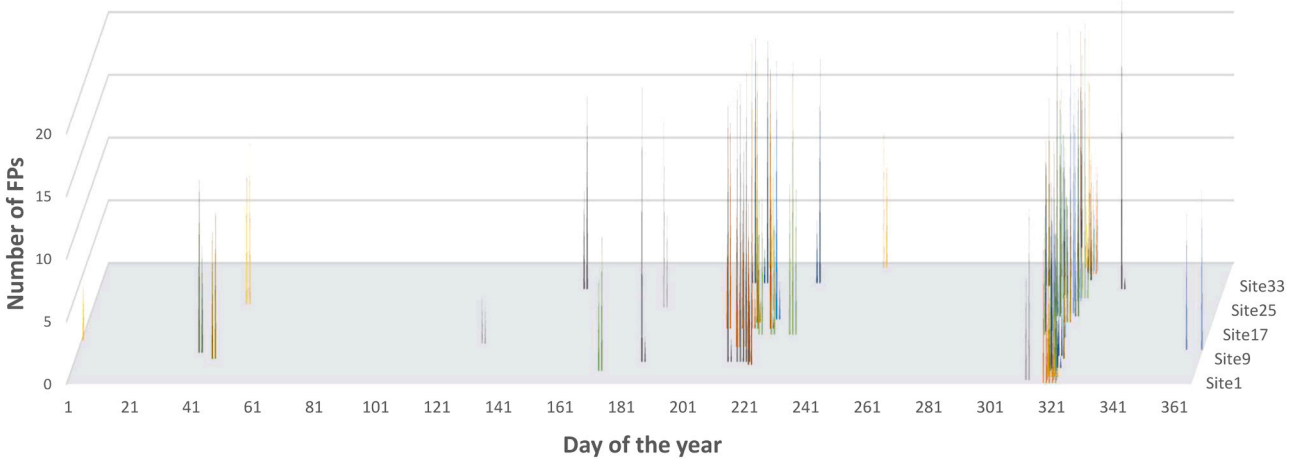

Fig. 4. Incidence of false positives, FP, for exceedance of the $PM_{2.5}$ 'Unhealthy' GV (150 μg m^{-3}) across sites in California in 2018. Only those sites recording one or more FPs are shown (40 in total). For each day, up to 24 FPs can be recorded at any particular site because the exceedances relate to rolling 24-h periods. In total, across all sites, there were 1567 FPs and 832 TPs.

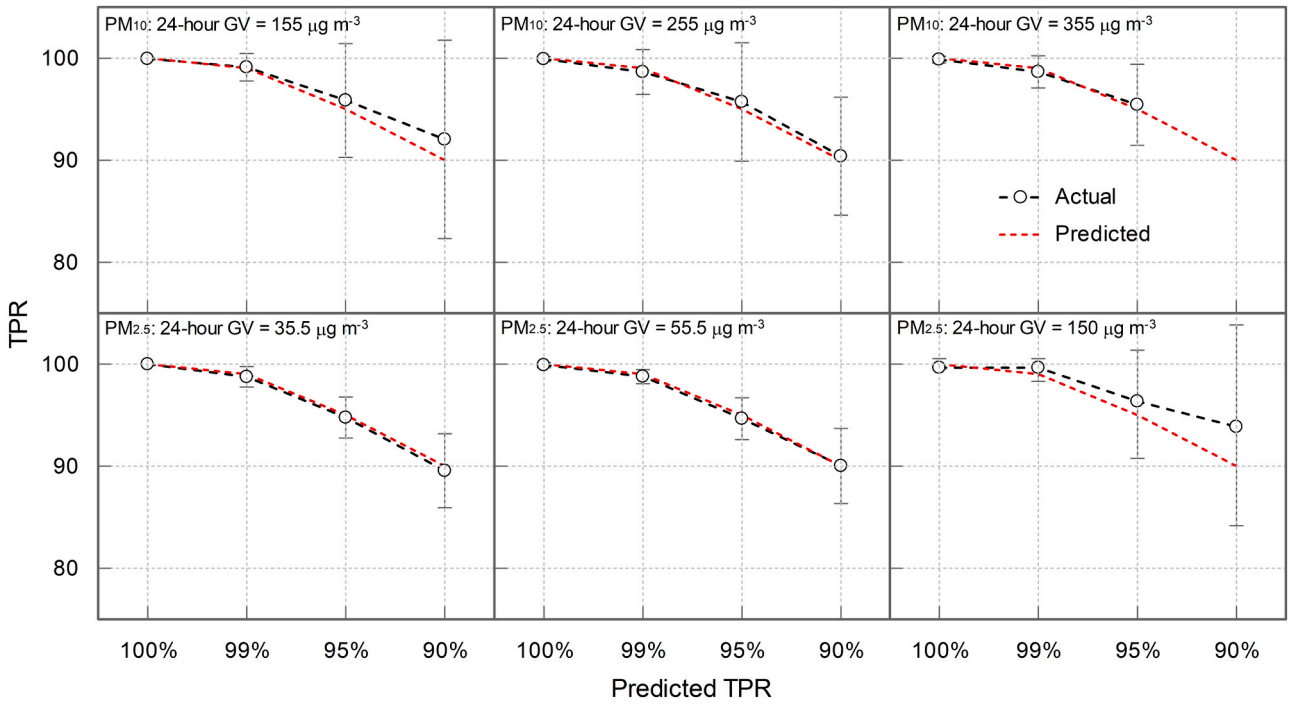

Fig. 5. Results of the cross-validation study for PM_{10} (upper set of panels) and $PM_{2.5}$ (lower set). From left to right the 24-h GV values represent the breakpoint concentrations for 'Unhealthy for Sensitive Groups', 'Unhealthy' and 'Very Unhealthy' respectively. Predicted TPRs (red dashed line) are at the set values of 90%, 95%, 99% and 100%. Black dashed lines show the average TPR obtained from six cross-validation runs in which, successively, each individual year's data was used as a test set for an ROC model based on the remaining five years of data. The error bars are ±1 sd. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

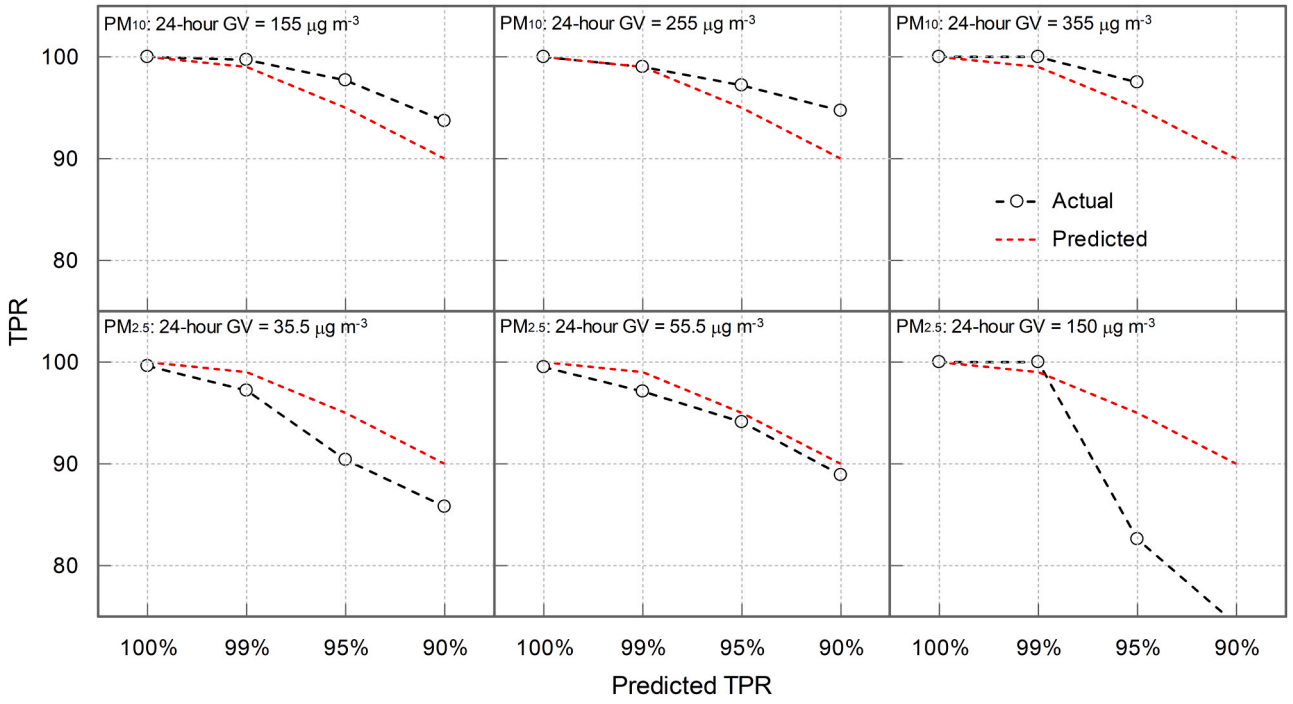

Fig. 6. PM_{10} (upper set of panels) and $PM_{2.5}$ (lower set) results for an ROC model based only on Californian data. From left to right the 24-h GV values represent the breakpoint concentrations for 'Unhealthy for Sensitive Groups', 'Unhealthy' and 'Very Unhealthy' respectively. Predicted TPRs (red dashed line) are at the set values of 90%, 95%, 99% and 100%. Black dashed lines show the TPR obtained from all US data, excluding California. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data for the affected areas. In so doing, the 1-h TCs could be assessed for their efficacy as health-relevant indicators of PM exposure in their own right.

3.4. Comparison with other approaches to calculating 1-h TCs

As discussed in the introduction, an alternative approach to the

development of 1-h TCs in the literature has been to divide the relevant 24-h GV by a fixed factor i.e., the ratio of the mean 24-h concentration to the maximum hourly concentration for the same period. This factor has been calculated to be 0.55 for PM_{10}, based on monitoring data from urban background and roadside monitoring stations in Europe (European Union, 2007). Separately, a numerically identical value has been derived for data from Canada (Stieb et al., 2008). For the European study, the purpose of producing a 1-h TC was to ensure that "the PM_{10} sub-index based on hourly values on a given day will be (on average) consistent with the daily value once it is calculated (the next day)" (European Union, 2007). Nevertheless, this approach, and that of Stieb et al. (2008), cannot provide any information on the proportion of exceedances of 24-h GVs that are predicted by their derived 1-h TCs. However, since our ROC approach also uses the maximum 1-h concentration (C_{max24}), we can calculate TPRs for 1-h TCs derived using the European Union (2007) and Stieb et al. (2008) approach. One caveat is that the CAQI index for which 1-h TCs were derived does not range as high as the 24-h GVs considered in the current paper: for PM_{10}, the boundary between CAQI High/Very High occurs at 100 μg m^{-3}.

We can also make comparisons with 1-h TCs calculated using ratios derived from the US data in the current study. Ratios of 0.46 and 0.48 were calculated for PM_{10} and $PM_{2.5}$ respectively, though these overall values mask considerable variation between states, as shown in Tables S10 and S11.

Table 4 summarises the predicted TPRs (and FPRs) for PM_{10}, based on 1-h TCs that were derived from 24-h guidelines using the C_{24}: C_{max24} approach. The 1-h TCs derived in this way do predict a high proportion of exceedances, with the accuracy increasing at the higher corresponding 24-h GVs, i.e. up to 99.7% for the 355 μg m^{-3} GV. The 1-h TC derived using US C_{24}: C_{max24} data gives a slightly lower accuracy, reflecting its higher value.

However, for $PM_{2.5}$, Table 5 shows that TPRs predicted using the C_{24}: C_{max24} ratio-derived 1-h TCs show a much lower accuracy, which decreases at the higher 24-h GVs. Thus, for the 150 μg m^{-3} 24-h GV, the corresponding 1-h TC (factor = 0.55) predicts only 47.9% of exceedances.

Taking these results together, whilst the use of a constant factor to derive 1-h TCs from the corresponding 24-h GVs has an agreeable simplicity, we have shown that it is not a suitable approach for $PM_{2.5}$, nor does it result in a consistent predictive accuracy across the range of 24-h GVs for either PM fraction. Nevertheless, it is helpful that such approaches can be quantified using the ROC analysis method that we have developed.

3.5. Using the ROC model to calculate 1-h TCs for 24-h GVs from the UK and WHO

Having demonstrated the excellent predictive capability of C_{max24} in the validation study, the model can be used with confidence to calculate 1-h TCs for a range of other 24-h health-based GVs. Table 6 shows additional 1-h TCs for PM_{10}, including for 24-h GVs corresponding to: WHO interim Target 2 and interim Target 3 (also equivalent to the UK's DAQI 24-h boundaries for 'high' and 'very' high); the UK Trigger to evacuate; and the US AQI boundary for 'hazardous' air pollution conditions (for the PM_{10} sub-index). Similarly, Table 7 shows 1-h TCs for $PM_{2.5}$ at the additional WHO interim targets, together with US AQI 'hazardous'. It should be noted that at the higher C_{max24} values, there are a limited number of data points and so it is not possible to accurately calculate 1-h TCs for the lowest TPRs (95% and 90%). Moreover, at these higher concentrations, there is a tendency for individual states, and even individual episodic events, to dominate the data. For example, across the US, over the period 2014-2019, there are only 180 exceedances of the $PM_{2.5}$ 24-h GV of 250 μg m^{-3} (AQI 'hazardous'), with the majority coming from one wildfire incident: the 'Thomas Fire' that affected Ventura and Santa Barbara Counties in December 2017 (Oakley et al., 2018, Zhou and Erdogan, 2019, Wong et al., 2020).

Finally, from Table 6, we note that at the UK DAQI PM_{10} boundary concentration for 'high' (24-h average = 75 μg m^{-3}), the corresponding DAQI 1-h trigger concentration of 107 μg m^{-3} (Connolly and Willis, 2013) will give a TPR of about 99% using our ROC approach (i.e. comparing against the closest $C_{max24(TPR)}$ entry in Table 6 of 109 μg m^{-3}). Similarly at the UK DAQI PM_{10} boundary concentration for 'very high' (24-h average = 100 μg m^{-3}), the corresponding DAQI 1-h trigger concentration of 177 μg m^{-3} (Connolly and Willis, 2013) will give a TPR of approximately 95%. The high corresponding TPRs for threshold (trigger) concentrations derived using an alternative methodology and dataset (DAQI uses UK ambient air quality data), gives us further confidence in the applicability of the ROC approach developed in this paper.

3.6. The public health response

It is acknowledged that decisions on public health interventions for episodic PM events need to be taken as quickly as possible (WHO, 2006b). Nevertheless, there is a problem that the shortest averaging period for which there are epidemiological-based PM GVs is 24-h (WHO, 2006a, 2006b, Federal Register, 2012). Therefore, the ability to robustly relate PM measurements gathered over shorter averaging periods to the epidemiologically based 24-h PM guidelines provides a way of addressing this shortcoming. Our ROC-based probability model allows precisely this type of relationship to be established and potentially speeds up the public health risk assessment and decision-making processes. The question then remains as to what public health response is required during episodic events, and at which 24-h GV (and associated 1-h TC) this should be triggered.

Several studies have found that communicating simple advisory measures about the need to restrict physical activity and to shelter indoors has beneficial effects (Kolbe and Gilchrist, 2009), including an association with reduced respiratory effects (Mott et al., 2002). In the US, states are already required to report 24-h AQI values for each metropolitan area with a population exceeding 350,000 (Code of Federal Regulations, 2016). For many states, this is done through the web-based NowCast system, which has the added advantage of predicting 24-h concentrations based on the previous 12 h of data (US EPA,

Table 4
Predicted TPRs (and FPRs) for 1-h PM_{10} TCs, that were derived by dividing 24-h guidelines by a fixed factor, based on the ratio of the mean 24-h concentration to the maximum hourly concentration for the same period. Two fixed factors are used: 0.55, derived from European data, and 0.46, derived from US monitoring data from the current study.

Basis for derivation of 1-h TC	24-h GV = 155 μg m^{-3} (AQI >100: Unhealthy for sensitive groups)			24-h GV = 255 μg m^{-3} (AQI >150: Unhealthy)			24-h GV = 355 μg m^{-3} (AQI >200: Very unhealthy)		
	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%
CAQI, European data: (24-h GV/0.55)	281	95.0	1.85	464	98.1	0.87	645	99.7	0.54
Current paper, US data: (24-h GV/0.46)	335	92.0	1.33	551	97.1	0.66	766	98.1	0.40

Table 5
Predicted TPRs (and FPRs) for 1-h $PM_{2.5}$ TCs, that were derived by dividing 24-h guidelines by a fixed factor, based on the ratio of the mean 24-h concentration to the maximum hourly concentration for the same period. Two fixed factors are used: 0.55, derived from European data, and 0.46, derived from US monitoring data from the current study.

Basis for derivation of 1-h TC	24-h GV = 35.5 μg m^{-3} (AQI >100: Unhealthy for sensitive groups)			24-h GV = 55.5 μg m^{-3} (AQI >150: Unhealthy)			24-h GV = 150 μg m^{-3} (AQI >200: Very unhealthy)		
	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%	1-h TC / μg m^{-3}	Predicted TPR/%	Predicted FPR/%
CAQI, European data: (24-h GV/0.55)	65	71	0.63	101	70.4	0.24	274	47.9	0.03
Current paper, US data: (24-h GV/0.46)	73	61.1	0.45	115	57.6	0.17	311	39.6	0.02

Table 6
Additional 1-h TCs for PM_{10}, where $C_{max24(TPR)}$ is the value of C_{max24} that will achieve true positive rates (TPRs) of 90%, 95%, 99% and 100% for the selected 24-h GVs. The predicted false positive rate (FPR) is also shown. $C_{max24(TPR)}$ can be considered to be a 1-h TC. AUCs of 0.9849, 0.9910, 0.9991 and 0.9993 were obtained for the 24-GVs of 75 μg m^{-3}, 100 μg m^{-3}, 320 μg m^{-3} and 425 μg m^{-3} respectively.

	GV = 75 μg m^{-3} (WHO Interim Target-3; UK DAQI 'high')		GV = 100 μg m^{-3} (WHO Interim Target-2, UK DAQI 'very high')		GV = 320 μg m^{-3} (UK Trigger to Evacuate)		GV = 425 μg m^{-3} (AQI >300: Hazardous)	
Selected TPR	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %
100%	91	12.4	125	7.52	439	0.99	658	0.53
99%	109	8.73	146	5.53	592	0.62	945	0.28
95%	129	6.14	171	4.01	932	0.27	–	–
90%	149	4.44	197	3.00	–	–	–	–

Table 7
Additional 1-h TCs for $PM_{2.5}$, where $C_{max24(TPR)}$ is the value of C_{max24} that will achieve true positive rates (TPRs) of 90%, 95%, 99% and 100% for the selected 24-h GVs. The predicted false positive rate (FPR) is also shown. $C_{max24(TPR)}$ can be considered to be a 1-h TC. AUCs of 0.9958, 0.9978, 0.9989 and 1.000 were obtained for the 24-GVs of 37.5 μg m^{-3}, 50 μg m^{-3}, 75 μg m^{-3} and 250 μg m^{-3} respectively.

	GV = 37.5 μg m^{-3} (WHO Interim Target-3)		GV = 50 μg m^{-3} (WHO Interim Target-2)		GV = 75 μg m^{-3} (WHO Interim Target-1)		GV = 250 μg m^{-3} (AQI >300: Hazardous)	
Selected TPR	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %	$C_{max24(TPR)}$ / μg m^{-3}	Predicted FPR/ %
100%	43	2.53	58	1.07	86	0.43	345	0.02
99%	48	1.76	63	0.86	93	0.34	–	–
95%	53	1.26	70	0.64	101	0.27	–	–
90%	57	1.02	75	0.51	108	0.23	–	–

2018, Mintz et al., 2013). Each AQI category has a specified cautionary statement, which depends on the sub-index that has been breached (i.e, those for PM_{10}, $PM_{2.5}$, O_3, CO, SO_2 or NO_2). These statements are particularly targeted at vulnerable people, urging them to stay indoors and to minimise physical activity. It is important that care is taken to reach the more marginalised members of society, who might also be the most affected (Santana et al., 2020, WHO, 2006b).

A more timely prediction that certain 24-h GVs will be exceeded allows for better planning, for example in ensuring that decisions are taken on identifying and safely evacuating the more vulnerable members of a population (Stares et al., 2014), in closing schools (Holm et al., 2020), advising on the wearing of masks (WHO, 2006b, Kolbe and Gilchrist, 2009), or on the use of air cleaning systems in homes, schools workplaces and smoke refuges (Stares et al., 2014, Mott et al., 2002, Holm et al., 2020, Barn et al., 2016).

Regarding the most extreme of these measures, evacuation, the highest AQI category, 'Hazardous' (PM_{10}, 425 μg m^{-3}; $PM_{2.5}$, 250.5 μg m^{-3}), does not contain a specific requirement to evacuate, rather the advice is that the vulnerable should stay indoors and keep physical activity levels low (US EPA, 2018). However, guidance given to public health officials in the event of wildfires in the US *does* specify the possible evacuation of 'at-risk populations' when the 24-h $PM_{2.5}$ concentration reaches the AQI 'Hazardous' category (Stone et al., 2019). The British Columbia Centre for Disease Control also recommends that evacuation is considered when US AQI 'Hazardous' is reached, though that the likely duration, plume toxicity, and presence of vulnerable subgroups, such as the elderly, children and those with underlying health conditions, are taken into account (Stares et al., 2014). From Tables 6 and 7, the 24-h AQI 'Hazardous' categories (PM_{10}, 425 μg m^{-3}; $PM_{2.5}$, 250.5 μg m^{-3}) would correspond to 1-h TCs of 658 μg m^{-3} and 345 μg m^{-3} for PM_{10} and $PM_{2.5}$ respectively at a TPR of 100% for our ROC analysis.

The practice of having threshold concentrations for evacuation is one that is used in other countries, for example the UK, where a 24-h PM_{10} 'trigger to evacuate' of 320 μg m^{-3} has been used during major incident fires (Brunt and Russell, 2012), and for which we have previously calculated a corresponding 1-h TC of 550 μg m^{-3} (100% TPR) (Griffiths et al., 2018). This 1-h TC is somewhat higher than the value of 439 μg m^{-3} obtained from the ROC model using US data in the current work (Table 6). Nevertheless, it is similar to the 99% TPR, 1-h TC of 590 μg m^{-3} in the current paper and also the now withdrawn 1–3-h average 'Recommended Action Level' for the closure of public buildings and possible evacuation, which was set of 526 μg m^{-3} for PM_{10}/$PM_{2.5}$ during wildfires (Lipsett et al., 2008).

Any decisions on evacuation should be carefully considered because there is evidence that such action can have a significant effect on mental and emotional health (Dodd et al., 2018, Krstic and Henderson, 2015) and may also involve other risks, such as exposure of vulnerable populations and responders to air pollutants during the evacuation process (Stewart-Evans et al., 2016). Other mitigating measures may be more

effective, such as sheltering in place (Stewart-Evans et al., 2016), combined with the use of portable air cleaning devices (Barn et al., 2016, Mott et al., 2002).

4. Conclusions

In this paper we have demonstrated the application of ROC analysis to derive 1-h TCs that have a probabilistic relationship with PM_{10} and $PM_{2.5}$ health-based 24-h exposure GVs. The analysis, based on 16 million and 22 million rolling 24-h periods for PM_{10} and $PM_{2.5}$ respectively, and involving a cross-validation design, shows that the maximum-observed PM concentration in any rolling 24-h averaging period is an excellent predictor of exceedances of 24-h GVs. An ROC analysis based on data only from California also provided a good basis for the prediction of exceedances from across the remaining states of the US.

The main advantages of our ROC method are as follows: (i) the high degree of accuracy that ROC-generated TCs can achieve in predicting exceedances of health-based 24-h guidelines; (ii) the consistency of year on year comparisons, as demonstrated by the validation analysis; (iii) the 'tunability' of the ROC method in generating TCs, i.e. the use of the ROC output table to select a TC that balances the need to achieve as high a TPR as possible, whilst also minimising the FPR; (iv) the transferability of this methodology to other datasets, e.g. in different countries, and to other pollutants for any 24-h health based GV for which the corresponding 1-h TC is required; and (v) the ease of use of the ROC model in generating TCs. The main disadvantage of the ROC approach is the high FPRs that are generated for the 24-h GVs at the lower end of the harmfulness scale, though we have also shown that false predictions tend to be clustered around specific episodic events, coincident with real exceedances, and thus might not registered as a false alarm by the affected population.

Elevated PM during episodic air pollution events is associated with significant short-term health impacts, including mortality, and so the ability to provide timely public health guidance on appropriate remedial measures for affected populations is vital. We hope that the straightforward approach to developing 1-h TCs that we have outlined in this paper might assist in this process.

CRediT authorship contribution statement

Michael Deary: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original draft, Writing – Review & Editing, Visualisation. **Simon Griffiths**: Conceptualisation, Investigation, Writing- Reviewing and 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.

Acknowledgements

This work uses data available from the United States Environmental Protection Agency. We are grateful to Northumbria University for funding this research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2021.126334.

References

Balmes, J.R., 2018. Where there's wildfire, there's smoke. N. Engl. J. Med. 378, 881–883.
Barn, P.K., Elliott, C.T., Allen, R.W., Kosatsky, T., Rideout, K., Henderson, S.B., 2016. Portable air cleaners should be at the forefront of the public health response to landscape fire smoke. Environ. Health 15, 1–8.
Behndig, A.F., Mudway, I.S., Brown, J.L., Stenfors, N., Helleday, R., Duggan, S.T., Wilson, S.J., Boman, C., Cassee, F.R., Frew, A.J., Kelly, F.J., Sandstrom, T., Blomberg, A., 2006. Airway antioxidant and inflammatory responses to diesel exhaust exposure in healthy humans. Eur. Respir. J. 27, 359–365.
Bekkar, M., Djemaa, H.K., Alitouche, T.A., 2013. Evaluation measures for models assessment over imbalanced data sets. J. Inf. Eng. Appl. 3.
Black, C., Tesfaigzi, Y., Bassein, J.A., Miller, L.A., 2017. Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. Environ. Toxicol. Pharmacol. 55, 186–195.
Brunt, H., Russell, D., 2012. Public health risk assessment and air quality cell for a tyre fire, Fforestfach, Swansea. Chem. Hazards Poisons Rep. 7–12.
Cascio, W.E., 2018. Wildland fire smoke and human health. Sci. Total Environ. 624, 586–595.
Code of Federal Regulations, 2016. Part 58 - Uniform Air Quality Index (AQI) and Daily Reporting. AE 2.106/3:40/.
Connolly, E., Willis, P., 2013. Update on Implementation of the Daily Air Quality Index. DEFRA, London, UK.
Dodd, W., Scott, P., Howard, C., Scott, C., Rose, C., Cunsolo, A., Orbinski, J., 2018. Lived experience of a record wildfire season in the Northwest Territories, Canada. Can. J. Public Health 109, 327–337.
European Union, 2007. Intererg IIIC Component 3, Comparing Urban Air Quality Across Borders: A review of existing air quality indices and the proposal of a common alternative [Online]. Available: ⟨http://citeair.rec.org/downloads/Products/ComparingUrbanAirQualityAcrossBorders.pdf⟩ [Accessed September 2020].
European Union, 2020a. Air Quality in Europe: Current Situation [Online]. Available: ⟨http://www.airqualitynow.eu/⟩ [Accessed September 2020].
European Union, 2020b. Air Quality in Europe: Indices Definition [Online]. Available: ⟨https://www.airqualitynow.eu/about_indices_definition.php#parag1⟩ [Accessed September 2020].
Fann, N., Alman, B., Broome, R.A., Morgan, G.G., Johnston, F.H., Pouliot, G., Rappold, A. G., 2018. The health impacts and economic value of wildland fire episodes in the US: 2008-2012. Sci. Total Environ. 610, 802–809.
Faustini, A., Alessandrini, E.R., Pey, J., Perez, N., Samoli, E., Querol, X., Cadum, E., Perrino, C., Ostro, B., Ranzi, A., Sunyer, J., Stafoggia, M., Forastiere, F., Group, M.-P. S., 2015. Short-term effects of particulate matter on mortality during forest fires in Southern Europe: results of the MED-PARTICLES project. Occup. Environ. Med. 72, 323–329.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.
Federal Register, 2012. National ambient air quality standards for particulate patter. Proposed rule.
Federal Register, 2013. National ambient air quality standards for particulate matter. Final Rule.
Finlay, S.E., Moffat, A., Gazzard, R., Baker, D., Murray, V., 2012. Health impacts of wildfires. PLoS Curr. 4, e4f959951cce2c.
Ford, B., Val Martin, M., Zelasky, S., Fischer, E., Anenberg, S., Heald, C.L., Pierce, J., 2018. Future fire impacts on smoke concentrations, visibility, and health in the contiguous United States. GeoHealth 2, 229–247.
Ghio, A.J., Kim, C., Devlin, R.B., 2000. Concentrated ambient air particles induce mild pulmonary inflammation in healthy human volunteers. Am. J. Respir. Crit. Care Med. 162, 981–988.
Greven, F.E., Krop, E.J., Spithoven, J.J., Burger, N., Rooyackers, J.M., Kerstjens, H.A., Van Der Heide, S., Heederik, D.J., 2012. Acute respiratory effects in firefighters. Am. J. Ind. Med. 55, 54–62.
Griffiths, S.D., Chappell, P., Entwistle, J.A., Kelly, F.J., Deary, M.E., 2018. A study of particulate emissions during 23 major industrial fires: Implications for human health. Environ. Int. 112, 310–323.
Haikerwal, A., Akram, M., Del Monaco, A., Smith, K., Sim, M.R., Meyer, M., Tonkin, A. M., Abramson, M.J., Dennekamp, M., 2015. Impact of fine particulate matter (PM 2.5) exposure during wildfires on cardiovascular health outcomes. J. Am. Heart Assoc. 4, e001653.
Hajian-Tilaki, K., 2013. Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp. J. Intern. Med. 4, 627–635.
Holgate, S., 2011. Review of the UK Air quality index. A Report by the Committee on the Medical Effects of Air Pollutants. Health Protection Agency, London.
Holm, S.M., Miller, M.D., Balmes, J.R., 2021. Health effects of wildfire smoke in children and public health tools: a narrative review. J. Expo. Sci. Environ. Epidemiol. 31, 1–20.
Howard, C., Rose, C., Dodd, W., Kohle, K., Scott, C., Scott, P., Cunsolo, A., Orbinski, J., 2021. SOS! Summer of Smoke: a retrospective cohort study examining the cardiorespiratory impacts of a severe and prolonged wildfire season in Canada's high subarctic. BMJ Open 11, e037029.
Hyde, P., Mahalov, A., Li, J.L., 2018. Simulating the meteorology and PM10 concentrations in Arizona dust storms using the Weather Research and Forecasting model with Chemistry (Wrf-Chem). J. Air Waste Manag. 68, 177–195.
Johnston, F.H., Borchers-Arriagada, N., Morgan, G.G., Jalaludin, B., Palmer, A.J., Williamson, G.J., Bowman, D.M., 2020. Unprecedented health costs of smoke-related PM 2.5 from the 2019–20 Australian megafires. Nat. Sustain. 1–6.
Keeley, J.E., Syphard, A.D., 2019. Twenty-first century California, USA, wildfires: fuel-dominated vs. wind-dominated fires. Fire Ecol. 15, 1–15.
Kochi, I., Champ, P.A., Loomis, J.B., Donovan, G.H., 2016. Valuing morbidity effects of wildfire smoke exposure from the 2007 Southern California wildfires. J. Econ. 25, 29–54.

Kolbe, A., Gilchrist, K.L., 2009. An extreme bushfire smoke pollution event: health impacts and public health challenges. NSW Public Health Bull. 20, 19–23.
Krstic, N., Henderson, S.B., 2015. Use of MODIS data to assess atmospheric aerosol before, during, and after community evacuations related to wildfire smoke. Remote Sens. Environ. 166, 1–7.
Lareau, N., Nauslar, N., Abatzoglou, J.T., 2018. The Carr Fire vortex: a case of pyrotornadogenesis? Geophys. Res. Lett. 45, 13,107–13,115.
Lee, K.Y., Feng, P.H., Ho, S.C., Chuang, K.J., Chen, T.T., Su, C.L., Liu, W.T., Chuang, H.C., 2015. Inter-alpha-trypsin inhibitor heavy chain 4: a novel biomarker for environmental exposure to particulate air pollution in patients with chronic obstructive pulmonary disease. Int J. Chronic Obstr. Pulm. Dis. 10, 831–841.
Lipsett, M., Materna, B., Stone, S.L., Therriault, S., Blaisdell, R., Cook, J., 2008. Wildfire Smoke: A Guide for Public Health Officials. Revised July 2008 (With 2012 AQI Values).
Liu, J.C., Pereira, G., Uhl, S.A., Bravo, M.A., Bell, M.L., 2015. A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. Environ. Res. 136, 120–132.
Main, L.C., Wolkow, A.P., Tait, J.L., Della Gatta, P., Raines, J., Snow, R., Aisbett, B., 2020. Firefighter's acute inflammatory response to wildfire suppression. J. Occup. Environ. Med. 62, 145–148.
Milford, C., Cuevas, E., Marrero, C.L., Bustos, J.J., Gallo, V., Rodriguez, S., Romero-Campos, P.M., Torres, C., 2020. Impacts of Desert Dust Outbreaks on Air Quality in Urban Areas. Atmosphere 11, 23.
Mintz, D., Stone, S., Dickerson, P., Davis, A., 2013. Transitioning to a new NowCast Method [Online]. US Environmental Protection Agency,. Available: ⟨https://www3.epa.gov/airnow/ani/pm25_aqi_reporting_nowcast_overview.pdf⟩ [Accessed 14/03/20 2020].
Moritz, M.A., Parisien, M.-A., Batllori, E., Krawchuk, M.A., Van Dorn, J., Ganz, D.J., Hayhoe, K., 2012. Climatechange and disruptions to global fire activity. Ecosphere 3, 1–22.
Morrison, A.M., Coughlin, K., Shine, J.P., Coull, B.A., Rex, A.C., 2003. Receiver operating characteristic curve analysis of beach water quality indicator variables. Appl. Environ. Microbiol 69, 6405–6411.
Mott, J.A., Meyer, P., Mannino, D., Redd, S.C., Smith, E.M., Gotway-Crawford, C., Chase, E., 2002. Wildland forest fire smoke: health effects and intervention evaluation, Hoopa, California, 1999. West J. Med. 176, 157–162.
Oakley, N.S., Cannon, F., Munroe, R., Lancaster, J.T., Gomberg, D., Ralph, F.M., 2018. Brief communication: Meteorological and climatological conditions associated with the 9 January 2018 post-fire debris flows in Montecito and Carpinteria, California, USA. Nat. Hazard Earth Syst. 18, 3037–3043.
Perlmutt, L.D., Cromar, K.R., 2019. Comparing associations of respiratory risk for the EPA Air Quality Index and health-based air quality indices. Atmos. Environ. 202, 1–7.
Phillips, M., Cataneo, R.N., Saunders, C., Hope, P., Schmitt, P., Wai, J., 2010. Volatile biomarkers in the breath of women with breast cancer. J. Breath. Res. 4, 026003.
Raman, A., Arellano, A.F., Brost, J.J., 2014. Revisiting haboobs in the southwestern United States: An observational case study of the 5 July 2011 Phoenix dust storm. Atmos. Environ. 89, 179–188.
Rappold, A.G., Reyes, J., Pouliot, G., Cascio, W.E., Diaz-Sanchez, D., 2017. Community vulnerability to health impacts of wildland fire smoke exposure. Environ. Sci. Technol. 51, 6674–6682.
Reid, C.E., Brauer, M., Johnston, F.H., Jerrett, M., Balmes, J.R., Elliott, C.T., 2016. Critical Review of Health Impacts of Wildfire Smoke Exposure. Environ. Health Perspect. 124, 1334–1343.
Rublee, C., Sorensen, C., Lemery, J., Wade, T., Sams, E., Hilborn, E., Crooks, J., 2020. Associations between dust storms and intensive care unit admissions in the United States, 2000–2015. GeoHealth 4, e2020GH000260.
Sachdeva, S., Mccaffrey, S., 2018. Using social media to predict air pollution during California wildfires. In: Proceedings of the 9th International Conference on Social Media and Society, pp. 365–369.
Salvi, S., Blomberg, A., Rudell, B., Kelly, F., Sandstrom, T., Holgate, S.T., Frew, A., 1999. Acute inflammatory responses in the airways and peripheral blood after short-term exposure to diesel exhaust in healthy human volunteers. Am. J. Respir. Crit. Care Med. 159, 702–709.
Santana, F.N., Fischer, S.L., Jaeger, M.O., Wong-Parodi, G., 2020. Responding to simultaneous crises: communications and social norms of mask behavior during wildfires and COVID-19. Environ. Res. Lett. 15, 111002.
Schaffer, C., 1993. Selecting a classification method by cross-validation. Mach. Learn 13, 135–143.
Stares, J., Erland, R., Mazey, P., Rideout, K., 2014. Evidence Review: Use of Evacuation to Protect Public Health During Wildfire Smoke Events. BC Centre for Disease Control, Vancouver BC.
Stenfors, N., Nordenhall, C., Salvi, S.S., Mudway, I., Soderberg, M., Blomberg, A., Helleday, R., Levin, J.O., Holgate, S.T., Kelly, F.J., Frew, A.J., Sandstrom, T., 2004. Different airway inflammatory responses in asthmatic and healthy humans exposed to diesel. Eur. Respir. J. 23, 82–86.
Stewart-Evans, J., Kibble, A., Mitchem, L., 2016. An evidence-based approach to protect public health during prolonged fires. Int. J. Emerg. Manag 12, 1–21.
Stieb, D.M., Burnett, R.T., Smith-Doiron, M., Brion, O., Shin, H.H., Economou, V., 2008. A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analyses. J. Air Waste Manag. 58, 435–450.
Stone, S.L., Sacks, J., Clune, A., Radonovich, L., D'alessandro, M., Wayland, M., Mirabelli, M. 2019. Wildfire Smoke: A Guide for Public Health Officials Revised 2019.
Swiston, J.R., Davidson, W., Attridge, S., Li, G.T., Brauer, M., Van Eeden, S.F., 2008. Wood smoke exposure induces a pulmonary and systemic inflammatory response in firefighters. Eur. Respir. J. 32, 129–138.
Tong, H., Rappold, A.G., Caughey, M., Hinderliter, A.L., Graff, D.W., Berntsen, J.H., Cascio, W.E., Devlin, R.B., Samet, J.M., 2014. Cardiovascular effects caused by increasing concentrations of diesel exhaust in middle-aged healthy GSTM1 null human volunteers. Inhal. Toxicol. 26, 319–326.
US EPA, 2014. Air Quality Index: A guide to air quality and your health [Online]. Available: ⟨https://www.airnow.gov/sites/default/files/2018–04/aqi_brochure_02_14_0.pdf⟩ [Accessed June 2020].
US EPA, 2018. Technical Assistance Document for the Reporting of Daily Air Quality - the Air Quality Index (AQI).
US EPA, 2021. Pre-Generated Data Files [Online]. Available: ⟨https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw⟩ [Accessed February 2021].
Videnova, I., Nedialkov, D., Dimitrova, M., Popova, S., 2006. Neural networks for air pollution nowcasting. Appl. Artif. Intell. 20, 493–506.
Wakefield, J., 2010. A Toxicological Review of the Products of Combustion. Health Protection Agency, Centre for Radiation, Chemical and Environmental.
Wegesser, T.C., Pinkerton, K.E., Last, J.A., 2009. California wildfires of 2008: coarse and fine particulate matter toxicity. Environ. Health Perspect. 117, 893–897.
WHO, 2006a. Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide. World Health Organization.
WHO, 2006b. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005: Summary of Risk Assessment. World Health Organization, Geneva.
Wong, S.D., Broader, J.C., Shaheen, S.A. 2020. Review of California Wildfire Evacuations from 2017 to 2019.
Xu, Y.A., Liu, H., Duan, Z., 2020. A novel hybrid model for multi-step daily AQI forecasting driven by air pollution big data. Air Qual. Atmos. Health 13, 197–207.
Yao, J., Brauer, M., Henderson, S.B., 2013. Evaluation of a wildfire smoke forecasting system as a tool for public health protection. Environ. Health Perspect. 121, 1142–1147.
Zhang, X.L., Zhao, L.J., Tong, D.Q., Wu, G.J., Dan, M., Teng, B., 2016. A systematic review of global desert dust and associated human health effects. Atmosphere 7, 158.
Zhou, S.Q., Erdogan, A., 2019. A spatial optimization model for resource allocation for wildfire suppression and resident evacuation. Comput. Ind. Eng. 138, 106101.
Zhu, W., Zeng, N., Wang, N., 2010. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In: NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland, vol. 19, p. 67.