INDENV 3 (2026) 100191

Contents lists available at ScienceDirect

Indoor Environments

journal homepage: www.sciencedirect.com/journal/indoor-environments

Influence of ventilation rate on the potential for indoor particulate matter exposure

Samuel G.A. Wood *, Alice E.E. Handy, Henry C. Burridge

Imperial College London, Department of Civil and Environmental Engineering, London, SW7 2AZ, UK

ARTICLE INFO

Keywords:
Indoor air quality
Ventilation
Particulate matter
Indoor exposures
Exposure modelling

ABSTRACT

Increased ventilation is widely recommended for improving indoor air quality, yet ventilation can also act as a pathway for outdoor particulate matter to enter indoor spaces. We investigate the mechanisms and extent by which ventilation may increase the potential for particulate matter exposure using both physical and statistical modelling. The physical model explores the dynamics of an idealised room over differing exposure-durations, exploring the duration-averaged particulate matter concentration as an indicator of the potential for exposure. This indicates that whilst duration-averaged indoor concentrations exhibit a simple linear increase with increasing outdoor concentrations, increasing the ventilation rate can either increase or decrease the potential for exposure indoors. Moreover, the physical model identifies two conditions under which increased ventilation can reduce duration-averaged indoor $PM_{2.5}$: transient removal from initially elevated indoor concentrations, and cases where indoor sources dominate over ingress from outdoors. We further exploit the large-scale 'SAMHE' dataset of air quality measurements made in UK schools. Using generalised additive modelling regression, we show that outdoor particulate matter concentration is the strongest predictor of the potential for particulate matter exposure, although its interaction with ventilation reveals a more complex non-linear relationship. Over the two school years of data examined, the statistical modelling suggests that increasing ventilation by $5\,L\,s^{-1}\,p^{-1}$ would, on average, increase daily mean indoor $PM_{2.5}$ by approximately $0.74\,\mu g\,m^{-3}$, with around 99% of schools predicted to experience increases, particularly those with higher outdoor pollution and lower baseline ventilation. However, these predictions were sensitive to the ventilation rates being achieved and the outdoor particulate matter concentrations. While increases in the ventilation rates associated with these schools would likely result in relatively small increases in the particulate matter concentrations therein, the reverse would be expected for spaces with larger indoor sources such as homes, dependent upon outdoor conditions.

1. Introduction

Indoor air quality research broadly promotes increased ventilation as a key strategy for reducing carbon dioxide concentration, lowering the risk of airborne infection, and diluting a range of indoor pollutants. Nevertheless, this advice is accompanied by a recurring qualification: ventilation can also act as a pathway for bringing in particulate matter, and other pollutants, from outdoors. which risks increasing indoor exposures to these outdoor pollutants. This trade-off is highlighted explicitly in recent school-focused work [1], where it is noted that "Ventilation is the process of exchanging indoor with outdoor air, which may lower indoor CO_2 levels but may introduce indoors high levels of certain pollutants (NOx, PM e.g. from traffic)". In the context of $PM_{2.5}$ exposure, Handy et al. [2] suggest that "some consideration could be devoted to periods of elevated outdoor PM2.5 concentrations ... changes in the operation of school ventilation or air cleaners/filters ... could be considered as a response". Relevant evidence from other building types tends to focus on naturally ventilated spaces and leads to similar conclusions; for example, Liu et al. [3] observed that "When outdoor $PM_{2.5}$ pollution is severe, natural ventilation may bring large quantities of outdoor particulate matter into indoor spaces, thus worsening the indoor air quality" within their study of residential buildings. These conclusions can be expected to hold for indoor spaces with any ventilation provision that provides no filtration of the incoming air flow. Together, these findings within the literature emphasise that while ventilation remains central to the management of indoor air quality and comfort, its potential as a driver for particulate matter exposure indoors warrants deeper consideration.

Some studies have explicitly linked an increase in ventilation to an increase in particulate matter concentration when outdoor particulate

* Corresponding author.
E-mail address: s.wood@imperial.ac.uk (S.G.A. Wood).

https://doi.org/10.1016/j.indenv.2026.100191
Received 20 February 2026; Received in revised form 18 August 2026; Accepted 21 August 2026
Available online 31 August 2026
2950-3620/© 2026 The Authors. Published by Elsevier Inc. on behalf of International Society of Indoor Air Quality and Climate. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

matter concentrations are high. For example, a study of fourteen residential buildings in Korea found that in urban areas with high outdoor pollution an increase in window opening time was linked to higher indoor particulate matter concentrations [4]. Other works have made similar findings across a variety of building types — Zhang et al. [5] monitored indoor and outdoor $PM_{2.5}$ in London nurseries and found that window opening during periods of elevated outdoor pollution led to measurable increases in indoor concentrations, highlighting the direct role of natural ventilation in transporting outdoor particles to indoors. Ścibor et al. [6] found the relationship between indoor particulate matter concentration, outdoor particulate matter concentration, and window opening to be complex and heavily influenced by the weather and the outdoor conditions.

Perhaps the widest review of attempts to quantify the relationship between indoor and outdoor particulate matter was carried out by Chen and Zhao [7]. It was found that across a number of studies, naturally ventilated spaces without substantial indoor particulate matter sources saw indoor/outdoor (I/O) ratios typically of less than unity. Furthermore, it was evident that while I/O ratios might provide a general indication of the relationship between indoor and outdoor air, the I/O ratios were shown to vary widely and be influenced by a wide variety of external factors. Elsewhere within the literature, the relationship between indoor and outdoor particulate matter concentration is often characterised more crudely using the infiltration factor, defined as the steady-state fraction of outdoor particles that appear suspended indoors in the absence of indoor sources [8]. It is derived from a steady-state mass balance, with particles entering the building envelope with an effective penetration factor via ventilation, and removed via deposition. This formulation assumes constant outdoor concentration, constant ventilation rate, and sufficient time for the indoor concentration to reach equilibrium. Such conditions are rarely met in naturally ventilated rooms, where wind- and buoyancy-driven flows and occupant behaviour (including window opening) lead to continuous variations in ventilation rate. As such, even though a number of studies exist which estimate infiltration factors for specific groups of buildings (with the resulting estimates typically spanning a wide range, e.g. [8–10]), characterisation of the relationship between indoor and outdoor particulate matter concentration, and ventilation rate in operational indoor spaces is absent.

To address this, we carry out physical modelling of the potential for particulate matter exposure within idealised rooms. We use this model to highlight scenarios for which increasing the ventilation rate can be expected to lead to a decrease, or increase, in the potential for exposure, as inferred from the duration-averaged concentration (Section 2). In the latter part of this study, we look at data from 388 operational schools fitted with air quality monitors, alongside data from their closest outdoor monitoring station (Section 3). This data enables construction of regression models, see Section 4, providing indoor particulate matter concentrations as a function of both outdoor particulate matter concentrations and inferred per-person ventilation rate. This provides predictions of the effect of increasing ventilation rate on indoor particulate matter concentrations and the potential for exposure. We then draw conclusions (Section 6).

Throughout, we use duration-averaged indoor $PM_{2.5}$ concentration as an indicator of the potential for exposure. In the physical model this average is taken over a prescribed exposure-duration, while in the statistical analysis it is taken over the school day (taken to be 09:00–16:00, see [11]). We note that the actual exposures that would be realised depend additionally on a number of factors, including individuals' breathing rate and occupancy patterns.

2. Physical modelling of the duration-averaged particulate matter concentrations in idealised rooms

To demonstrate some of the subtleties that changing the ventilation rate might have on the realised indoor particulate matter concentrations, p_{in}, we consider the evolution of these concentrations within a simplified model of a room. We take the room, of volume V, to be well-mixed (i.e. spatially uniform concentrations within), and subject to a ventilation flow of constant rate Q which dilutes indoor concentrations but also brings in particulate matter from the outdoor air. Finally, we assume the room is subject to a series of indoor sources (human activity, resuspension, etc.), providing a source term S. In addition to ventilation losses, we allow for deposition at rate $\mathcal{R}_d$, and treat $\mathcal{R}_d$ as independent of ventilation rate, as supported by evidence showing that turbulent deposition velocities for $PM_{2.5}$ are relatively low and insensitive to friction velocity and hence ventilation rate [12]. As such, the effective removal rate is $Q^* = Q + \mathcal{R}_d V$; however, we note that practically altering the deposition rate is challenging and we therefore consider changing the effective removal rate to be synonymous with changing the ventilation rate.

In this deliberately idealised physical model, we retain only the parameters required to isolate the most important governing mechanisms. During each exposure-duration the system evolves transiently from a specified initial condition while all parameters are held constant in time.

As such the evolution of the indoor particulate matter concentration is described by

$$\frac{\mathrm{d}p_{\text{in}}}{\mathrm{d}t} = \frac{S}{V} + \frac{Q}{V}p_{\text{out}} - \frac{Q^*}{V}p_{\text{in}}\,, \tag{1}$$

where the $PM_{2.5}$ concentration of the air entering the room is $p_{\text{out}} = \eta\, p_{\text{amb}}$ where p_{amb} is the ambient outdoor concentration and $\eta \leq 1$ is a penetration factor accounting for any reduction in particle transmission through the building envelope. We note that penetration factors in naturally ventilated buildings tend to lie within the range 0.7–1.0, often approaching unity when windows are open [7]; as such, for simplicity we hereafter incorporate the penetration factor into p_{out} and hereafter refer to p_{out} as the outdoor concentration. Such a room approaches a steady-state at large time for which the indoor particulate matter concentration reaches

$$p_\infty = \frac{S}{Q^*} + \frac{Q}{Q^*}\,p_{\text{out}}\,, \tag{2}$$

and we note that $p_\infty > p_{\text{out}}$ can only be satisfied if $S > p_{\text{in}}\mathcal{R}_d V\,(p_{\text{out}}/p_{\text{in}})$, i.e. if the indoor source term is greater than the deposition in the space divided by the I/O ratio. In addition, (2) shows that when there is no ventilation, $Q = 0$, then the steady-state concentration p_∞ has no dependence on the outdoor concentration, as one would naturally expect.

Taking the initial indoor particulate matter concentration to be p_0, with no loss of generality, the concentration at time t is given by

$$p_{\text{in}}(t) = p_\infty + (p_0 - p_\infty)e^{-Q^* t/V}\,. \tag{3}$$

We are interested in the potential for exposures, indicated by the average particulate matter concentrations over some exposure-duration T_{exp}, and so we define the duration-averaged concentration as

$$\overline{p}_{\text{in}} = \frac{1}{T_{\text{exp}}}\int_0^{T_{\text{exp}}} p_{\text{in}}(t)\,\mathrm{d}t = p_\infty + (p_0 - p_\infty)\frac{1 - e^{-Q^* T_{\text{exp}}/V}}{Q^* T_{\text{exp}}/V}\,. \tag{4}$$

Furthermore, our interest is to explore the effects of changing ventilation rate on the potential for exposure indoors, through the duration-averaged indoor concentration, $\overline{p}_{\text{in}}$. We note that (4) highlights that $\overline{p}_{\text{in}}$ is only affected by the outdoor concentration via changes in the steady-state concentration, see (2). Moreover, it can be seen that the duration-averaged indoor concentration depends on ventilation rate through two distinct mechanisms; transient dynamics relating to the initial concentration, and through changes to the steady-state concentration, see (2).

Fig. 1 presents the duration-averaged indoor $PM_{2.5}$ concentration, $\overline{p}_{\text{in}}$, calculated according to (4), for a room of volume $V = 140\,\mathrm{m}^3$, with an initial concentration of $p_0 = 2\,\mu\mathrm{g\,m}^{-3}$ and an indoor source term of $S = 400\,\mu\mathrm{g\,h}^{-1}$ and deposition rate of $\mathcal{R}_d = 0.4\,\mathrm{h}^{-1}$ (see [13]), with an exposure-duration of 6 h. These values are selected to be

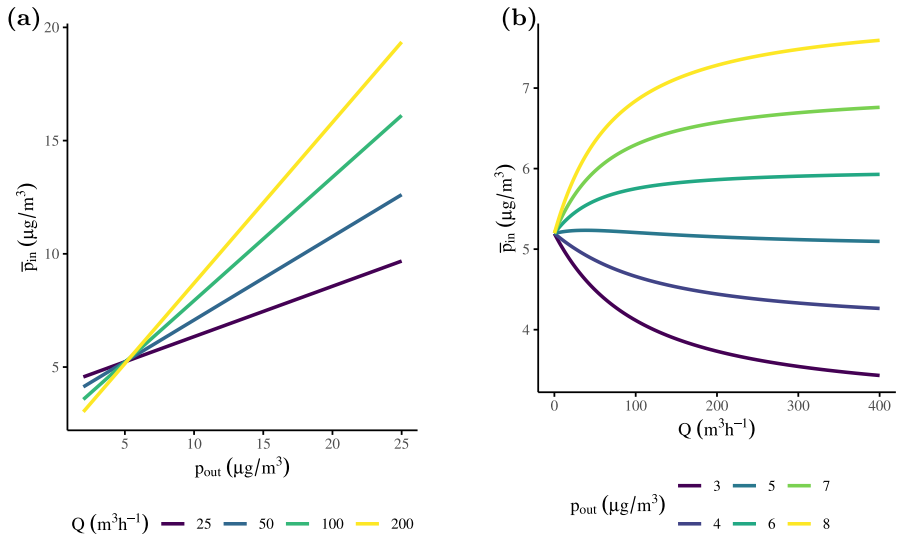

Fig. 1. Illustrative example of the duration-averaged indoor $PM_{2.5}$ concentration, $\overline{p}_{in}$, calculated according to (4) for an idealised room of volume 140 m^3, with an initial concentration of $p_0 = 2\,\mu g\,m^{-3}$ and an indoor source term of $S = 400\ \mu g\,h^{-1}$, a deposition rate of $\mathcal{R}_d = 0.4\ h^{-1}$ with an exposure-duration of 6 h. The figure shows the variation of $\overline{p}_{in}$ with: (a) the outdoor $PM_{2.5}$ concentration, illustrated for a number of fixed values of the ventilation rate, Q; and (b) the ventilation rate, illustrated for a number of fixed values of the outdoor $PM_{2.5}$ concentration, p_{out}.

illustrative of a typical naturally ventilated classroom [2,14]. Fig. 1(a) illustrates that, for a constant ventilation rate Q, the duration-averaged indoor particulate matter concentration $\overline{p}_{in}$ exhibits a relatively simple linear dependence on p_{out}; with the gradient set by the removal rate Q^* (for this illustrative example, the intersection points happen to fall at $p_{out} \approx 6\,\mu g\,m^{-3}$).

Fig. 1(b) illustrates more complex responses in $\overline{p}_{in}$ for increasing ventilation rates. This figure shows a single illustrative example with varied outdoor concentrations, for which all curves share the same vertical-axis intercept, $\overline{p}_{Q=0} = 5.2\ \mu g\,m^{-3}$. The exact value of $\overline{p}_{Q=0}$ depends on several parameters; namely, $\overline{p}_{Q=0} = f(p_0, S, \mathcal{R}_d, V, T_{exp})$, see (4) for the functional form. As discussed above, when the ventilation is set to zero, i.e. $Q = 0$, then the indoor concentration, $\overline{p}_{in}|_{(Q=0)} = \overline{p}_{Q=0}$, has no dependence on the outdoor particulate matter concentration. Whilst, in the scenarios for which ventilation is small (but non-zero), such that $Q \ll V/T_{exp}$, the sign of the change in $\overline{p}_{in}$ between different ventilation rates depends predominantly on whether the outdoor concentration is less than the indoor concentration in the limit of zero ventilation, or not. When $p_{out} < \overline{p}_{Q=0}$, increasing ventilation dilutes particulate matter within the room, with the reverse true when $p_{out} > \overline{p}_{Q=0}$, as illustrated in Fig. 1(b) by the changing sign of the gradients for curves with p_{out} below, or above, $\overline{p}_{Q=0} = 5.2\,\mu g\,m^{-3}$. For scenarios of intermediate ventilation rates, i.e. Q of the order of V/T_{exp}, the concentration $\overline{p}_{in}$ can be sensitive to the transient dynamics relating to the initial concentration, and also sensitive to changes that varying Q induces in the steady-state concentration. This can lead to complex, non-monotonic behaviours in $\overline{p}_{in}$ with increasing Q, as is subtly illustrated by the curve relating to the scenario $p_{out} = 5\,\mu g\,m^{-3}$, in Fig. 1(b). Finally, in all scenarios at higher ventilation rates, $\overline{p}_{in}$ becomes less sensitive to changes in ventilation rate. To explore these observations more rigorously, we examine a dimensionless form of idealised rooms that adhere to (4).

We write the dimensionless duration $\beta = Q^* T_{exp}/V$ so that we can express the amount by which this duration-averaged particulate matter concentration deviates from its steady state as

$$\frac{\overline{p}_{in} - p_\infty}{p_\infty} = \left(\frac{p_0 - p_\infty}{p_\infty}\right)\left[\frac{1 - e^{-\beta}}{\beta}\right]. \tag{5}$$

The relative difference between the period mean and steady state condition can therefore be characterised as the product of two terms: the first of these is the relative deviation from the initial concentration, and the second is a dynamic term which characterises the extent to which the initial concentration remains significant. The dynamic term, $(1 - e^{-\beta})/\beta$, takes a maximum value of unity at $\beta = 0$ and decreases monotonically towards zero at long durations (large β) with a half-life of $\beta \approx 1.6$, a duration of about 1.6 V/Q^*. Finally, it is worth noting that the potential for indoor $PM_{2.5}$ exposures in any scenario, within rooms described by (5), depends only on the initial and steady concentrations (p_0 and p_∞) and the dimensionless duration (β).

Our interest is in how the duration-averaged indoor concentration $\overline{p}_{in}$ responds to changes in the removal rate Q^* — that is, whether increasing ventilation tends to increase or decrease the potential for exposure. To quantify this, we define the dimensionless concentration gradient based on the partial derivative of $\overline{p}_{in}$ with respect to the (time-invariant) removal rate Q^*, with all other parameters remaining constant.[1] We therefore define the dimensionless concentration gradient as $d_{p_{Q^*}} = (Q^*/p_\infty)\partial\overline{p}_{in}/\partial Q^*$, and (by exploiting (2), including that $\partial p_\infty/\partial Q^* = -S\,Q^{*-2} + \mathcal{R}_d V\,Q^{*-2} p_{out}$) we express this as

$$d_{p_{Q^*}} = \left(\frac{p_0 - p_\infty}{p_\infty}\right)\left[\frac{\beta e^{-\beta} + e^{-\beta} - 1}{\beta}\right] + \left(\frac{p_{out} - p_\infty}{p_\infty}\right)\left[1 - \frac{1 - e^{-\beta}}{\beta}\right]. \tag{6}$$

The sign of $d_{p_{Q^*}}$ determines whether increasing the removal (or ventilation) rate will increase or decrease the average particulate matter concentration, $\overline{p}_{in}$, over the duration T_{exp} — a negative sign indicating that increased ventilation would result in decreased indoor $PM_{2.5}$ concentrations. The case when $d_{p_{Q^*}} = 0$ therefore marks the critical parameter combinations at which the effect of ventilation on potential for exposure changes sign. This sign is determined by the sum of two terms, each of which are best considered as the products of two factors (each bracketed in (6)).

[1] Although $\overline{p}_{in}$ is derived from an integrated solution of the ordinary differential equation (1), it is a smooth function of Q^* and this derivative provides a convenient measure of the sensitivity of the duration-averaged concentration to different fixed ventilation rates.

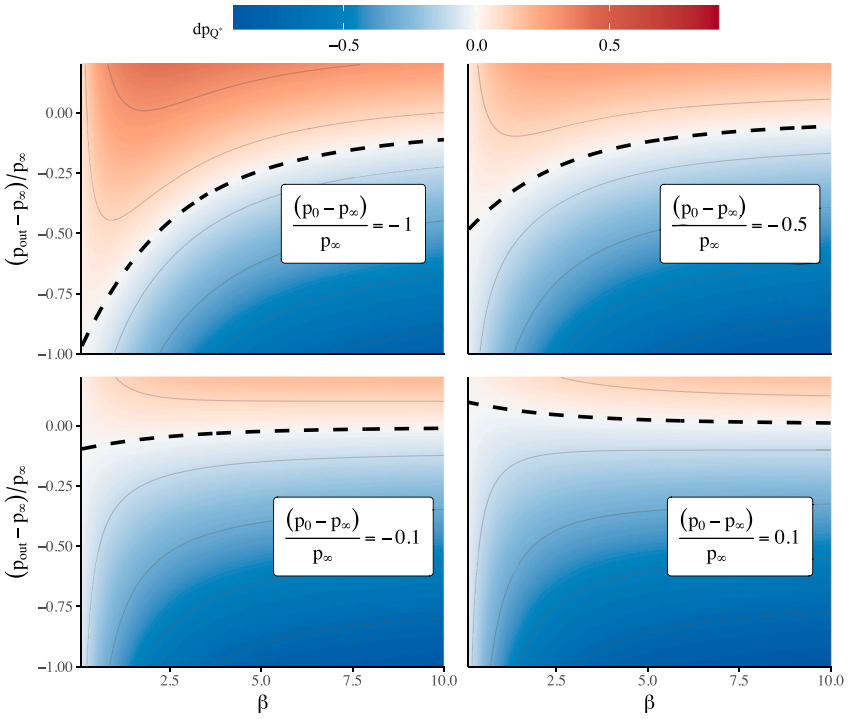

Fig. 2. The dimensionless gradient of duration-averaged indoor $PM_{2.5}$ with respect to removal rate and outdoor concentration, for different initial concentrations relative to the steady-state $PM_{2.5}$ concentration.

The first term is the product of a factor which is the relative difference between the initial and steady-state concentrations, and a factor that vanishes both for $\beta \to 0$ (i.e. very short durations of low ventilation and deposition rates) and for $\beta \to \infty$ (i.e. very long durations of large ventilation rates), and which is negative between these limits (taking a minimum value of approximately -0.3, for $\beta \approx 1.8$). We describe this first term as the 'transient effects term' since it incorporates the impact of the initial concentration on the dimensionless concentration gradient, and it can only be significant for intermediate durations and ventilation rates (i.e. intermediate β). This transient effects term offers the first mechanism, by which increasing the ventilation rate may lead to a decrease in the duration-averaged $PM_{2.5}$ concentration, but only if the initial concentration is larger than the steady-state concentration and the product of the removal rate and exposure-duration is of the order of the room volume.

In the second term, the second factor is relatively simple in that it depends only on β and is zero when $\beta = 0$ growing asymptotically to unity at large β (reaching a value of 0.5 at $\beta \approx 1.6$), i.e. this factor is only positive. The first factor in this second term is simply the relative difference between the outdoor and steady-state concentrations but its impact on the dimensionless concentration gradient is more subtle to interpret since the steady-state concentration itself depends on the outdoor concentration and the ventilation rate. However, whether the second term contributes positively or negatively to the dimensionless concentration gradient is simply determined by the sign of $(p_{\text{out}} - p_\infty)$ which is set only by whether or not the indoor sources are greater than the deposition divided by the I/O ratio, as described in the text following (2). Hence this second term offers a second mechanism, by which increasing the ventilation rate may lead to a decrease in the duration-averaged $PM_{2.5}$ concentration, but only if the indoor sources are greater than the deposition divided by the I/O ratio. This mechanism is most dominant for relatively long exposure-durations and large ventilation rates (i.e. $\beta \gg 1$) since then the full magnitude of the negative $(p_{\text{out}} - p_\infty)/p_\infty$ is realised, with half its magnitude being realised for the intermediate $\beta \approx 1.6$.

As such, increasing the ventilation rate can only decrease the duration-averaged $PM_{2.5}$ concentration in a given scenario if one of the two above mechanisms dominates, or if they act in combination. This, and the behaviour of dp_{Q^*} more generally, is explored visually for different parameter combinations in Fig. 2. Blue regions indicate parameter space in which the gradient is negative and increasing the ventilation rate is expected to decrease the duration-averaged $PM_{2.5}$ concentration. In the top-left pane the initial concentration indoors is zero, so the contribution to the gradient from the first term in (6) is as strongly positive as possible. As such, in this top left pane, the gradient can only be negative when the second mechanism dominates. This increasingly occurs for combinations of either larger β (long exposure-durations and large ventilation rates), or for greater magnitudes of $p_{\text{out}} < p_\infty$, i.e. $S > p_{\text{in}} \mathcal{R}_d V (p_{\text{out}}/p_{\text{in}})$, so the indoor sources are greater than the deposition divided by the I/O ratio. The bottom right pane illustrates a scenario in which the initial concentration exceeds the steady-state concentration, and this is the only scenario shown in the figure for which the first mechanism can contribute to a negative gradient. This first mechanism contributes most negatively to dp_{Q^*} at the intermediate $\beta \approx 1.8$. The remaining panes illustrate scenarios in which the initial concentration is more modest. Red regions indicate parameter combinations for which the gradient is positive, i.e. where increasing the ventilation rate is expected to increase the duration-averaged $PM_{2.5}$ concentration.

The above analysis, including its visual presentation in Fig. 2, and the particular example shown in Fig. 1, all confirm the variety of behaviours that can be expected even within these idealised rooms. In particular, whether increasing the ventilation rate can be expected to have a positive or negative impact on the potential for exposures within is not as simple as one might have expected; for example, increasing the ventilation rate can still lower the potential for exposures indoors even if the concentrations outside are above those inside. With this in mind, it is worth considering sets of scenarios, indicative of different building typologies, and examining which behaviours might be expected to arise in each typology. We consider the behaviour of these idealised

Table 1
Distributions for indoor $PM_{2.5}$ model parameters estimated from the literature for naturally ventilated (NV) classrooms and rooms in homes, and mechanically ventilated (MV) offices; showing mean, median, and 5th–95th percentile values used to define plausible parameter ranges for exposure-duration, i.e. T_{exp}. For classrooms, the exposure-duration is taken to be a uniform distribution between 6 and 8 h, for homes it is taken to be a uniform distribution between 1 and 4 h, and for offices to be between 6 and 10 h.

Environment	Parameter	Mean	Median	5%	95%
NV classrooms	S (μg h^{-1}) [2]	360	300	100	1000
	Air-change rate [14]	1.2	1.2	0.5	2.6
	$\mathcal{R}_d$ (h^{-1}) [15,16]	0.38	0.35	0.20	0.60
	V (m^3) [14]	240	242	203	282
NV homes	S (μg h^{-1}) [17]	2300	1500	300	6000
	Air-change rate [18]	0.80	0.79	0.33	1.9
	$\mathcal{R}_d$ (h^{-1})[15,16]	0.38	0.35	0.20	0.60
	V (m^3) [19]	50	51	28	82
MV offices	S (μg h^{-1}) [20,21]	400	300	50	1200
	Air-change rate [22]	2	1.7	0.5	3.0
	$\mathcal{R}_d$ (h^{-1}) [15,16]	0.38	0.35	0.20	0.60
	V (m^3) [22]	170	150	80	300

rooms now in the context of real spaces, by considering the range of parameters likely to be associated with naturally ventilated classrooms, mechanically ventilated office spaces and then rooms in naturally ventilated homes (these are selected to represent a range of domestic rooms, including kitchens, bedrooms and living spaces). We suggest parameter distributions relevant to the model above for UK schools, offices and homes in Table 1; these are intended to illustrate plausible operating conditions rather than provide quantitive predictions for individual spaces.

These are estimated to be broadly representative of these environments based on available measurements, and do not give a precise or exhaustive indication of the range of environments.

For each of these environments we may therefore generate a joint distribution of $((p_{out} - p_\infty)/p_\infty, (p_0 - p_\infty)/p_\infty, \beta)$, based on an initially clean space such that $p_0 = 0$. We draw from these distributions using Monte Carlo sampling. Based on these we may visualise the region containing 95% of the probability mass for each environment. We show these visualisations overlaid on to the dimensionless gradient calculated for the same conditions from Eq. (6), in Fig. 3. For classrooms and offices, it may be seen that the majority of the probability mass exists in regions of negative dp_Q, but a substantial fraction exists in regions of positive gradient; i.e. it might be expected that in classrooms and offices, increased ventilation might serve to either decrease or increase duration-averaged $PM_{2.5}$ concentration. In contrast, in domestic rooms, almost all of the probability mass lies in regions of negative gradient, suggesting that increased ventilation in homes is largely expected to reduce duration-averaged $PM_{2.5}$ concentration. This is in keeping with the knowledge that, in general, homes are known to be more influenced by indoor sources of particulate matter, whereas classrooms and offices have been shown to be dominated by outdoor air [2,16].

3. Data methods

To develop upon the theoretical insights above, this paper presents analysis of measurements of the particulate matter ($PM_{2.5}$) concentrations collected in 388 SAMHE schools —— those that participated in the Schools' Air Quality Monitoring for Health and Education (SAMHE) initiative, and provided data meeting quality assurance criteria (see Section 3.1). Schools volunteered to join SAMHE and collected air quality data following a citizen science based deployment, see [1]. The participating schools form a cohort that is broadly representative of the UK school estate in terms of national distribution and key demographic characteristics, including fee-paying status and the indices measuring relative levels of deprivation in the UK.

Each participating school was provided with an air quality monitor and access to the co-designed SAMHE Web App [23], which enables users to explore their air quality data and engage in educational activities. Guidance was given on where monitors should be placed in classrooms (at head height and away from windows, doors and other ventilation pathways), and schools were free to move monitors between rooms. However, the precise location of the monitors within buildings could not be known by the research team. The analysis herein assumes that, given the number of schools involved, most measurements are representative of the conditions in classrooms (as per [24]). The analysis in this work uses only the environmental and air quality measurements recorded by the SAMHE monitors; no additional contextual information on room occupancy or usage was incorporated nor required.

Environmental and air quality parameters were sampled continuously at one-minute intervals whenever the monitor was mains-powered and connected to the internet. Measurements were obtained using AirGradient One devices, which incorporate a Plantower PMS5003 optical particle counter (laser-scattering) to measure $PM_{2.5}$ concentrations (μg m^{-3}) and a SenseAir S8/S88 non-dispersive infrared sensor to measure CO_2 (ppm). The CO_2 sensors use an automatic baseline correction algorithm that compensates for slow drifts away from expected outdoor background concentrations during extended periods of inoccupancy (e.g. weekends). Twenty PMS5003 sensors in AirGradient One devices were validated against reference-grade instruments at four UK sites, yielding an MAE of 1.87 μg m^{-3} across 3254 observations of daily mean concentration, with a Spearman's rank correlation of approximately 0.74 ($p \ll 0.001$) between the PMS5003 sensors and reference grade instruments. These sensor errors were found to be small relative to the variability in μg m^{-3} observed across SAMHE schools, and no inter-device harmonisation was applied across the network. For full details, see Appendices G and H in [2]. While low-cost sensors have been shown to over-predict $PM_{2.5}$ at high humidities, at relative humidities below RH = 75% these over-predictions were deemed to be of no substantial effect [25]. Within the dataset of measurements in SAMHE schools, over 99% of all relative humidity values measured were below RH = 75%; as such, no humidity correction was deemed necessary.

The SAMHE $PM_{2.5}$ concentration was analysed alongside ambient outdoor $PM_{2.5}$ concentrations obtained from Defra AURN monitoring sites. These AURN sites report $PM_{2.5}$ concentrations at hourly resolution. For each SAMHE school included in the study, we identify its nearest background AURN site, and only data from these matched sites are used in the comparison. Our analysis was limited to AURN locations classified as background sites, which are intended to provide pollution measurements that characterise air quality over several square kilometres and are not dominated by any single local emission source (see [26]). Industrial and traffic stations are excluded, as they are deliberately positioned to capture emissions associated with specific industrial processes or nearby road traffic and therefore are not expected to reflect the outdoor air quality surrounding SAMHE school buildings.

3.1. Constructing the dataset

We define a school-data day as the measurements made during a normal teaching day within a single school over the period 09:00–16:00. Analysis of the air quality measurements spanning two academic years, 2023–2024 and 2024–2025, was included. Weekends were excluded from the dataset, due to lack of occupancy in schools, as were school holidays. Some school-data days were excluded from the dataset as the measured CO_2 concentration over the course of the day was indicative they were not occupied: this occurred when the mean CO_2 concentration over the course of the day was less than 480 ppm, or if at no point an increase in CO_2 concentration of 100 ppm or greater was observed over a fifteen-minute period. Due to the nature of the SAMHE dataset, individual schools could disconnect and reconnect over the course of

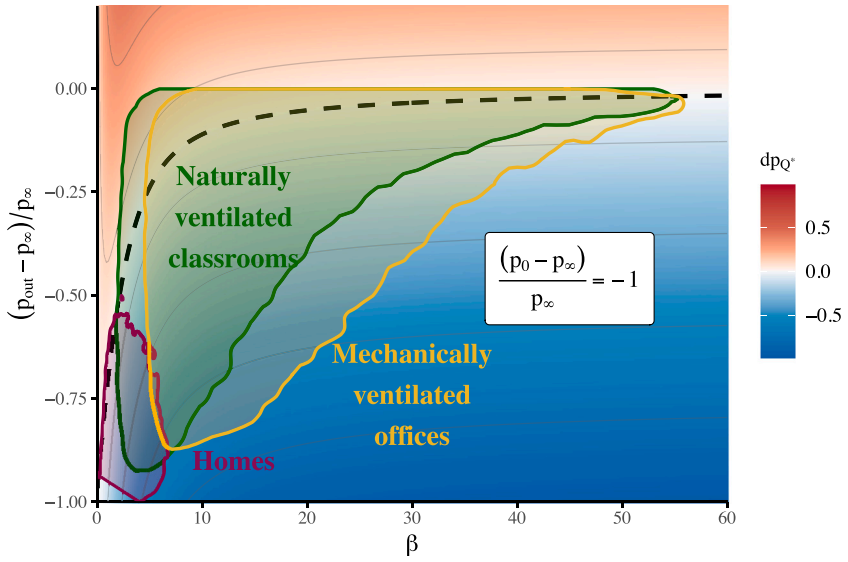

Fig. 3. The dimensionless gradient of duration-averaged indoor $PM_{2.5}$ with respect to removal rate, for different outdoor $PM_{2.5}$ concentrations, overlaid with representative shapes indicating 95% of the probability mass for different environments.

the year, and schools were included regardless of the number of days of data they provided, and on each included day there were between 73 and 201 schools included in the dataset. When, within a single day, a given monitor had data for less than 75% of hourly measurements (of either CO_2 or $PM_{2.5}$), that school's data was excluded from the dataset for that day.

The participating schools have been shown to be broadly representative of the UK school estate in terms of national distribution and key demographic characteristics [1]. The scale and distribution of the dataset reduces the influence of school day variability and provides broad coverage across observed operating conditions. While individual school-day estimates may carry uncertainty, we therefore have no specific reason to suspect systematic bias across the dataset. However, scale alone does not in itself remove possible systematic uncertainty or unmeasured confounding, and the results should be considered as empirical observations for the observed schools. As this is an in situ analysis of operational schools, uncertainty in individual parameters such as exact sensor placement, indoor sources, or occupancy is itself representative of the real-world conditions under which these schools operate, and is therefore an inherent feature of the dataset rather than a limitation to be corrected for.

Daily mean indoor $PM_{2.5}$ concentration was calculated for each school-data day included within the dataset by averaging over the period 09:00–16:00. The daily mean indoor $PM_{2.5}$ concentration serves as the indicator of potential for exposure within the statistical analysis, analogous to the duration-averaged concentration used in the physical model of Section 2, and similarly providing a practically relevant and directly comparable metric across schools and school-days.Daily mean outdoor $PM_{2.5}$ concentration for each school-data day was taken as the average $PM_{2.5}$ over the period 09:00–16:00 for the nearest AURN station with greater than 75% data completeness on that day. The average distance between a school and the nearest AURN station on a given day was 16.1 km. The large number of schools reduces sensitivity to differences between outdoor measurement sites and true outdoor conditions at a school, although it cannot exclude systematic differences between AURN background measurements and the outdoor $PM_{2.5}$ conditions immediately surrounding each school [2,27].

Direct measurement of ventilation rate for a naturally ventilated space can be challenging, and many commonly used methods require substantial contextual information or rely on simplifying assumptions. However, for schools, it has been demonstrated that reliable estimates of ventilation rates can be obtained from integral CO_2 measurements with very limited additional information [28], using only a single monitor per classroom. In line with [28], we estimate the daily average per-person ventilation rate, $|Q_{pp}|_{est}$, for each school on each day using a classroom occupancy duration T_{occ}, the mean per-person CO_2 generation rate $\overline{\langle G_N \rangle}_{est}$, and C, the integrated excess CO_2 concentration (i.e. the difference between the CO_2 measured by the SAMHE monitor and the baseline CO_2 concentration) across the occupied period. The daily mean for each SAMHE monitor is then estimated as

$$\overline{|Q_{pp}|}_{est} = M \frac{T_{occ}\,\overline{\langle G_N \rangle}_{est}}{\int_{t_0}^{t_{end}} C, \mathrm{d}t}, \tag{7}$$

where M is a bias-correction factor that reflects the amount of contextual information. For classrooms $1 < M \leq 1.05$, and for SAMHE we use $M = 1.048$ following [28].

Following [24], we take a $T_{occ} = 5\,\mathrm{h}$ (i.e. we assume each classroom was occupied for approximately five hours to exclude breaks during the school day) and per-person CO_2 generation rates of $4.32\,\mathrm{mL\,s^{-1}}$ in secondary schools and $3.1\,\mathrm{mL\,s^{-1}}$ in primary schools [29]. With these inputs, Eq. (7) gives the expected daily mean per-person ventilation rate for each school day. As shown in [28], the distribution of plausible daily averages is well described by a Gaussian with standard deviation equal to 12% of the mean under the limited contextual information available in SAMHE. We therefore assign each school-day a Gaussian distribution for Q_{pp}, and by sampling this distribution for each day we draw a plausible range of daily mean per-person ventilation rates for that day which accounts for the uncertainty in the estimation method. We note that the use of daily mean values necessarily averages over within-day variability; however, the daily mean represents the most practically actionable timescale for ventilation decisions in schools, where modulating ventilation in response to short-term fluctuations in outdoor $PM_{2.5}$ is not feasible in practice. Additionally, inference of ventilation over shorter timescales is not feasible with current methods. This integral method requires no assumptions of steady-state conditions, constant ventilation rate, or well-mixed air, and is therefore better suited to operational naturally ventilated classrooms than alternative

Table 2
For each school, on each day, the three summary statistics that provide the core of our analysis.

Measurement	Source	Explanation
Indoor $PM_{2.5}$	SAMHE school's monitored $PM_{2.5}$ data	Average $PM_{2.5}$ measurement between 09:00–16:00.
Outdoor $PM_{2.5}$	Closest background AURN station data	Average $PM_{2.5}$ measurement between 09:00–16:00.
Per-person ventilation rate	SAMHE school's monitored CO_2 data	Inferred based on integral CO_2 method of [28].

approaches such as the CO_2 decay method, which requires occupants to have vacated the space and may not be representative of ventilation conditions during occupied periods.

For each school-data day we therefore have three main statistics upon which we base our analysis, and these are summarised in Table 2. In total, there were 39 906 school-data days within the data set from 287 SAMHE monitors at different schools. For each of these, we randomly draw ten samples of ventilation rate in order to capture the uncertainty in inferred ventilation rate within the model. For completeness, the joint distribution of observed indoor and outdoor $PM_{2.5}$ across all school-data days is presented as a heatmap in Appendix, demonstrating broad and dense coverage across the range of interest.

3.2. Regression model specification

Generalised Additive Models (GAM) were used for the regression in order to produce the modelled indoor $PM_{2.5}$ concentrations, allowing for non-linear dependencies on, and between, predictors to be captured, whilst accounting for the hierarchical structure of the data wherein multiple observations arise from the same schools. The general form of the model is given by

$$p_m = f_1(p_{\text{out}}) + f_2(Q_{pp}) + f_{12}(p_{\text{out}}, Q_{pp}) + u_j + \epsilon \tag{8}$$

where $f_1(p_{\text{out}})$ is a smooth function of daily mean outdoor $PM_{2.5}$ concentration, $f_2(Q_{pp})$ is a smooth function of per-person ventilation rate, $f_{12}(p_{\text{out}}, Q_{pp})$ is a tensor interaction smooth capturing the joint non-linear effect of daily mean outdoor $PM_{2.5}$ concentration and per-person ventilation rate, $u_j \sim \mathcal{N}(0, \sigma_u^2)$ is a school-level random effect accounting for between-school differences in baseline indoor $PM_{2.5}$, and ϵ is the residual error. The school-level random effect ensures that the estimated effects of outdoor $PM_{2.5}$ and ventilation accounts for within-school variation. Models were fitted using Restricted Maximum Likelihood (REML) [30].

4. Results

To begin, we qualitatively explore the dataset of $PM_{2.5}$ concentrations measured in schools, before deploying the regression modelling to quantitatively analyse them — throughout, we draw comparison to the physical modelling of Section 2.

4.1. Qualitative exploration

To visually assess the relationships between outdoor $PM_{2.5}$ concentration, inferred per-person ventilation rate, and indoor $PM_{2.5}$ concentration within the dataset measured in schools (see Section 3.1), the colours plotted in Fig. 4 indicate the average daily mean indoor $PM_{2.5}$ concentrations measured when the dataset is subdivided into 36 brackets based on varied bands of daily mean outdoor $PM_{2.5}$ concentration values and inferred per-person ventilation rates. Firstly, the figure shows that the variation with per-person ventilation rate is slight compared to the more significant variation with daily mean outdoor $PM_{2.5}$ concentration – daily mean outdoor $PM_{2.5}$ concentration is far more strongly associated with daily mean indoor $PM_{2.5}$ concentration.

If any single band of per-person ventilation rate is considered (rows in the figure), indoor $PM_{2.5}$ always monotonically increases with increasing outdoor $PM_{2.5}$. However, considering bands of outdoor $PM_{2.5}$ concentration values (columns in the figure), varied and more complex trends in daily mean indoor $PM_{2.5}$ concentration are qualitatively evident when the inferred per-person ventilation rate increases. All of these behaviours are consistent with expectations based on the simplified model presented in Section 2; for example, the illustrative example in Fig. 1(a) shows simple monotonic increases in daily mean indoor concentrations when increasing outdoor levels (irrespective of the ventilation rate), whilst Fig. 2(b) shows that non-monotonic behaviour might be expected, particularly at relatively high concentrations of outdoor $PM_{2.5}$. Similar observations would be obtained if the inverse of CO_2 concentration were used as a proxy for ventilation instead of the per-person ventilation rate.

We note that, unlike the physical model presented in Section 2, this statistical framework does not assume fixed or steady-state parameters — the daily mean values used as predictors will themselves inherently reflect the integrated effect of transient fluctuations throughout the school day, including those arising from intermittent window opening, occupant activities, and other time-varying factors.

4.2. Regression modelling

Following qualitative exploration of the dataset (Section 4.1), a statistical evaluation, using regression analysis, compares the fits of modelled daily mean indoor $PM_{2.5}$ concentrations, p_m, to those observed within the dataset. We constructed two regression models: a 'baseline' model, which models indoor $PM_{2.5}$ solely as a function of outdoor $PM_{2.5}$; and an 'interaction' model which exploits both the outdoor $PM_{2.5}$ and per-person ventilation rate data, and the interaction between them. We compare the models by assessing the change in the coefficient of determination, R^2, (resulting from the improved fit of the model to the data when the additional predictor and interaction term are included), and by comparing the Akaike Information Criterion (AIC) between models, where a reduction in AIC of greater than 2 is taken as meaningful evidence of improved model fit accounting for model complexity [31]. We assess its statistical significance by means of an analysis of variance (ANOVA) of the residual sum of squares associated with the model fits.

Table 3 shows that adding per-person ventilation rate and its interaction with outdoor $PM_{2.5}$ improved model fit based on comparison of ANOVA, AIC and R^2.

The ANOVA comparison assesses whether the reduction in residual sum of squares achieved by adding the ventilation-related smooth terms is greater than would be expected purely from the addition of extra terms to the model. The very small p-value provides strong evidence that the additional terms significantly improved the model fit. AIC serves to compare the relative fit of the baseline and interaction models, while penalising the interaction model for its additional complexity. The reduction in AIC was substantial (ΔAIC = 1505), indicating that the interaction model provided a meaningfully improved fit even after accounting for the additional model complexity.

The increase in R^2 was small: $\Delta R^2 = 0.00532$, equivalent to 0.53% of the total observed variance in daily mean indoor $PM_{2.5}$ concentration. Thus, while the additional predictor terms improved the fit, they account for only a modest additional proportion of the total variation between observed school-days. This result can also be expressed relative to the variation not explained by the baseline model. For the baseline model, 43.7% of the observed variance was unexplained. Adding ventilation and its interaction with outdoor $PM_{2.5}$ reduces this remaining unexplained variance by $0.0053/0.437 \approx 1.2\%$. This remains a modest improvement to overall explanatory power, and indicates that daily mean outdoor $PM_{2.5}$ is the dominant predictor of daily mean indoor $PM_{2.5}$.

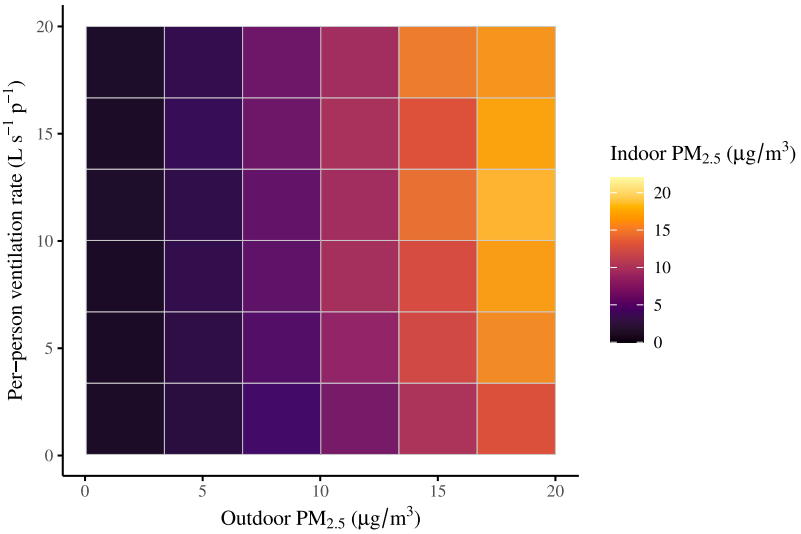

Fig. 4. The daily mean indoor $PM_{2.5}$ concentration averaged over schools and school-days that fall in one of 36 brackets based on bands of associated daily mean outdoor $PM_{2.5}$ concentration values and inferred per-person ventilation rates.

Table 3
Statistics of the two GAM regressions. The coefficient of determination R^2 is shown for each model, with the improvement ΔR^2 relative to the baseline model. The p-value indicates the statistical significance of the improvement in model fit resulting from the addition of the ventilation term. $\overline{\Delta p}_{Q1\text{–}9}$ denotes the mean change in predicted daily mean indoor $PM_{2.5}$ concentration, p_m, between the 1st and 9th deciles of per-person ventilation.

Model	Predictors	R^2	ΔR^2	AIC	ΔAIC	p-value	$\overline{\Delta p}_{Q1\text{–}9}$
Baseline	p_{out}	0.5627	–	231 634	–	–	–
Interaction	$p_{out} + Q_{pp}$ + interaction	0.5680	0.00532	230 129	1505	7.52×10^{-41}	2.1 µg m^{-3}

Results from the GAM regression including both predictors but excluding their interaction term resulted in reductions in the coefficient of determination approximately 3.7 times smaller than the reduction when the interaction term was included — indicating that, if both ventilation and outdoor $PM_{2.5}$ are to be used as predictors, the interaction between the two is also important. Hereafter, we therefore choose to only include the results from the 'interaction' GAM regression.

The relatively small impact when modelling daily mean indoor $PM_{2.5}$ concentrations of including data for per-person ventilation rates (i.e. a modest reduction in the unexplained variance, even when including interaction terms) might be taken to suggest that this data is not particularly valuable. Indeed, for applications when the data originates from large numbers of similar indoor environments, with not dissimilar outdoor environments (here UK schools), one might consider that it is appropriate to only include outdoor $PM_{2.5}$ concentration data (and ignore the effect of per-person ventilation rate) when attempting to model indoor $PM_{2.5}$. However, this should be considered with caution as it is partly a reflection of the relatively narrow distribution of per-person ventilation rates observed within these UK schools and should not be expected to hold, necessarily, for other indoor environments. Furthermore, as ventilation is inferred rather than experimentally controlled within this observational dataset, the results reflect statistical associations and model predictions rather than directly observed causal effects.

While the increase in the explained variance from the data relating to ventilation might appear small, it is statistically significant and the resulting model reveals substantial differences in predicted indoor $PM_{2.5}$ across realistic contrasts in per-person ventilation rate, conditional on daily mean outdoor $PM_{2.5}$ concentration. These differences are not well captured by global goodness-of-fit metrics such as the coefficient of determination, R^2. Thus we additionally consider a metric indicating the change in the modelled indoor $PM_{2.5}$, p_m, associated with a relatively large change in ventilation, taking the 1st and 9th deciles measured for Q_{pp} as being indicative of a large change in ventilation. Since the GAM regression is bivariate, i.e. $p_m = p_m(p_{out}, Q_{pp})$, we further require that this metric reflects the dependence on p_{out} appropriately, but desire to avoid the extrema of p_{out}, which we achieve by reflecting only p_m values corresponding to values of p_{out} that fall within the middle 80%, i.e. between the 1st and 9th deciles, of observed p_{out} values. This metric, which we denote $\overline{\Delta p}_{Q1\text{–}9}$, is then calculated as the difference in two values corresponding to the mean of p_m values that correspond to all of the middle 80% of the values of observed p_{out} within the dataset, and to either Q_{pp} taking the value of the 9th decile or the 1st decile, respectively. The resulting metric takes a value $\overline{\Delta p}_{Q1\text{–}9} = 2.1$ µg m^{-3}, which is 36% of the mean indoor $PM_{2.5}$ concentration across the dataset, and therefore indicates that there is a considerable difference in model predicted indoor $PM_{2.5}$ measurements between poorly-ventilated and well-ventilated schools.

Further investigation is enabled by taking the indoor $PM_{2.5}$ modelled (as a function of outdoor $PM_{2.5}$ and per-person ventilation, Q_{pp}) from the 'interaction GAM' regression and querying the model to provide curves of modelled indoor $PM_{2.5}$ as a function of outdoor $PM_{2.5}$, at constant ventilation rates. Plotted in Fig. 5 are the nine curves, each illustrating the model predictions for one of the nine deciles within the dataset of observed Q_{pp}. At low levels of outdoor $PM_{2.5}$ (less than approximately 3 µg m^{-3}), Fig. 5 shows that the modelled indoor $PM_{2.5}$ corresponding to the lowest ventilation rates displays only slightly

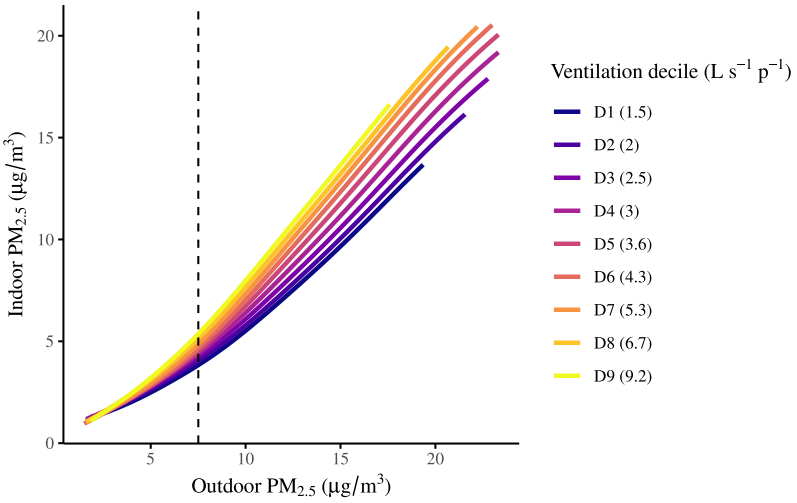

Fig. 5. Modelled daily mean indoor $PM_{2.5}$ concentrations from the GAM regression. Nine curves mark the modelled indoor $PM_{2.5}$ as a function of outdoor $PM_{2.5}$, one each for the deciles of per-person ventilation (within the observed dataset), with the deciles shown in brackets in the legend. For context, the average daily mean outdoor $PM_{2.5}$ in the observed dataset is shown as a dashed vertical line.

higher daily mean indoor $PM_{2.5}$ concentrations than those corresponding to the highest ventilation rates. This suggests that, as might be expected at these low outdoor concentrations, increasing ventilation may reduce indoor concentrations but the impact is of small magnitude. Beyond this, as the daily mean outdoor $PM_{2.5}$ concentrations increase, this dependence on ventilation rate inverts so that the highest indoor $PM_{2.5}$ concentrations are associated with the highest ventilation rates, and the effect increases so that the separation between the predictions for the 1st and 9th deciles Q_{pp} values is much greater. Finally we note there is very little separation between the daily mean indoor $PM_{2.5}$ concentrations associated with the top three deciles, irrespective of the outdoor $PM_{2.5}$ concentration — therefore, the model suggests that for the, approximately, 30% most highly-ventilated school days the indoor $PM_{2.5}$ concentrations were relatively insensitive to the absolute level of ventilation rates achieved, and only really impacted by the concentration in the outdoor environment.

The quantitative predictions from the GAM regression (informed by daily averaged measurements in operational schools), as illustrated by the curves in Fig. 5, may be compared to the analytical predictions for the duration-averaged indoor particulate matter concentration in idealised rooms (i.e. well-mixed spaces with all parameters held constant for the averaging duration), presented in Section 2. Illustrative predictions for idealised rooms show a linear relationship between $\overline{p}_{\text{in}}$ and p_{out}; Fig. 5 shows an approximately linear relationship over a wide range of data. The curves in Fig. 5 reflect fits to data composed of measurements in hundreds of different UK schools and with this in mind, the level of similarity shown between the data presented in Figs. 1(a) and 5 can be taken as indicating some reassurance in the modelling approaches that each reflect.

Beyond Fig. 5, the results of the GAM regression can be used to provide predictions of how the daily mean indoor $PM_{2.5}$ concentrations might be expected to change if the ventilation rate were increased by any particular amount. To do this, we take all of the pairs of daily mean outdoor $PM_{2.5}$ concentration and inferred per-person ventilation rates observed within the dataset from measurements, and for each we use the GAM regression to evaluate the modelled daily mean indoor $PM_{2.5}$ concentration, namely $p_m = p_m(p_{\text{out}}, Q_{pp})$. We then evaluate how the model predicts daily mean indoor $PM_{2.5}$ concentration might change if the per-person ventilation rate were increased by some amount ΔQ, giving $p_m + \Delta p_m = p_m(p_{\text{out}}, Q_{pp} + \Delta Q)$. Subtracting the original modelled indoor $PM_{2.5}$ concentration provides the modelled estimate of the change, Δp_m, associated with an increase in ventilation of ΔQ. For each pair of (p_{out}, Q_{pp}) within the observed dataset, we repeat this process for ΔQ values ranging from $0.1\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$ to $10\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$, in increments of $0.1\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$. This yields 100 values of Δp_m for each observed pair in the dataset.

The curve connecting the mean across all pairs of Δp_m values for each ΔQ is plotted in both panels of Fig. 6 as the thick solid black line. This data indicates relatively rapid increases in the modelled daily mean indoor $PM_{2.5}$ concentration for changes in ventilation up to about $\Delta Q = 5\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$ ($\Delta p_m \approx 0.74\,\mu\text{g}\,\text{m}^{-3}$), with more gradual increases thereafter (e.g. for $\Delta Q = 10\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$, $\Delta p_m \approx 0.76\,\mu\text{g}\,\text{m}^{-3}$). However, much is masked within the mean and examining the modelled data more broadly is worthwhile. Taking again the example of an increase of $\Delta Q = 5\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$ in the per-person ventilation rate, the modelled data predicts an increase in daily mean indoor $PM_{2.5}$ concentration for approximately 85% of observed (p_{out}, Q_{pp}) pairs, with reductions predicted for the remaining 15%. When this data is taken and then the modelled daily mean indoor concentrations associated with each school are averaged over the study period then 99% of the schools are predicted by the model to show an increase in their mean indoor $PM_{2.5}$ concentrations over the period. Still taking the example of an increase of $\Delta Q = 5\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$ in the per-person ventilation rate, the modelled data predicted that for the single most influenced school $\Delta p_m \approx 1.7\,\mu\text{g}\,\text{m}^{-3}$, which was about 15% of its annual mean value.

To investigate the variations within the predictions of the model, the dataset is first subdivided into four brackets based on the value of each observation's outdoor $PM_{2.5}$ concentration (i.e. that associated with each school on each day). The division is such that a quarter of all observations lie within each of the four brackets. The four dashed curves in Fig. 6(a) mark the loci of the mean Δp_m values within each bracket, and the shaded region highlights the interquartile range within the bracket.

Within the lowest quartile bracket of $PM_{2.5}$, only a very slight increase in daily mean $PM_{2.5}$ is seen even for large increases in ventilation rate. At higher quartile brackets, larger initial increases are seen with each increase in $PM_{2.5}$ bracket, though for the quartile bracket with the highest daily mean outdoor $PM_{2.5}$ concentrations, increases of more than approximately $5\,\text{L}\,\text{s}^{-1}\,\text{p}^{-1}$ are no longer associated with increases in daily mean indoor $PM_{2.5}$ concentration. Interestingly, this trend is

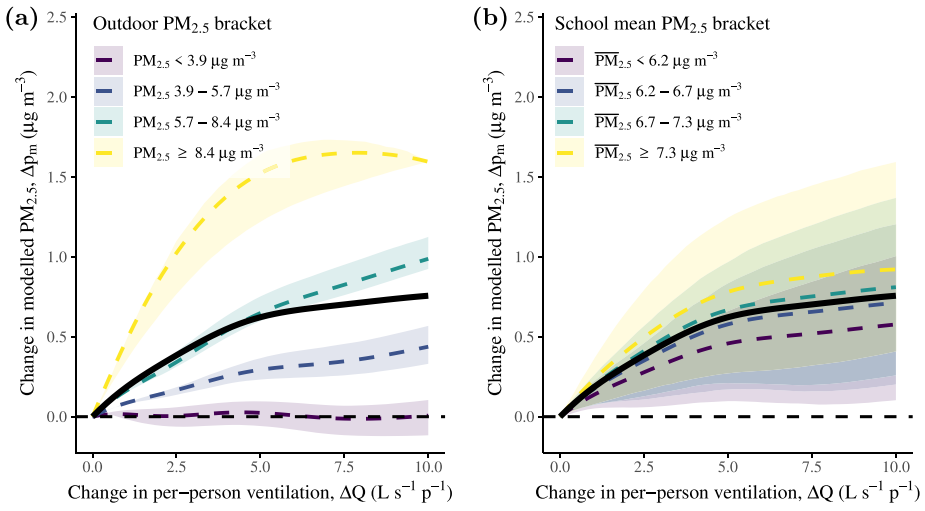

Fig. 6. The changes in modelled daily mean indoor $PM_{2.5}$ concentrations predicted by the GAM regressions for increasing per-person ventilation rates: (a) with the resulting data divided into four brackets based on the local outdoor $PM_{2.5}$ measurement recorded on each day at each school, and (b) with the data associated with each school divided into four brackets based on each schools' mean outdoor $PM_{2.5}$ measurements throughout the study period.

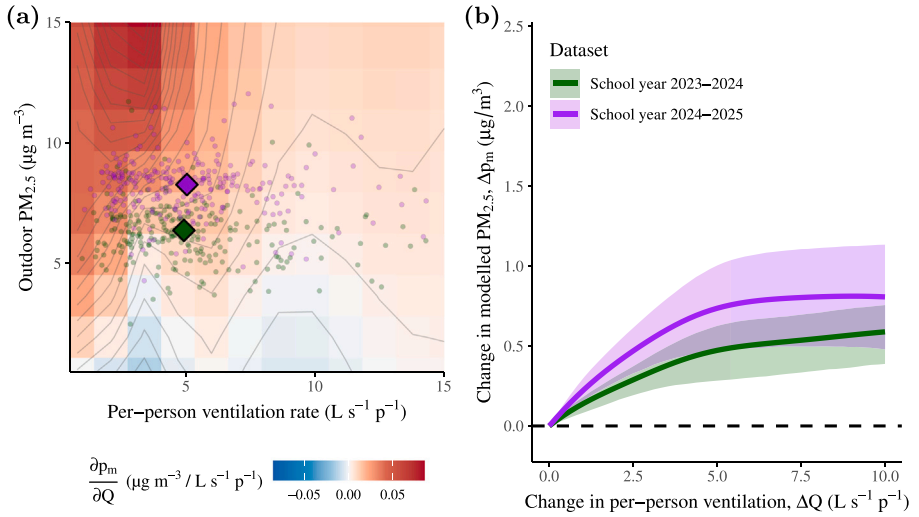

Fig. 7. (a) Colour map indicating the gradient, with respect to ventilation rate, in the modelled GAM $PM_{2.5}$— contours mark lines of constant gradient. Overlain are points marking annual averages of the input data to the GAM regression; namely, average $PM_{2.5}$ and per-person ventilation rate measurements for each school over the academic year 2023–2024 (in green) and the academic year 2024–2025 (in purple), with the mean of these data for each year marked by the large diamond. (b) The change in modelled indoor daily mean $PM_{2.5}$ concentration expected when per-person ventilation rates are increased with the data divided by academic year, the shaded regions indicate the interquartile range within the data.

qualitatively somewhat similar to that of the illustrative example of an idealised room shown at the higher outdoor concentrations (p_{out} > 5 μg m^{-3}), see Fig. 1(b).

The previous method of subdivision results in the data gathered from any given school being distributed within any number of the quartile brackets. To provide an alternate presentation of this data, Fig. 6(b) shows results for four different subsets, this time based on the value of annual mean outdoor $PM_{2.5}$ concentration associated with each school, such that each subset contains data from a quarter of the schools included in this study. Again, each of the four dashed curves in Fig. 6(b) marks the loci of the mean Δp_m values within each subset, and each exhibits a qualitatively similar shape — suggesting that on average a small increase in ventilation rate is expected to result in a slight increase in the daily mean indoor $PM_{2.5}$ concentration, irrespective of a schools' mean outdoor $PM_{2.5}$ concentration.

From the GAM regression, we can explore the parameter space of daily mean outdoor $PM_{2.5}$ concentrations and per-person ventilation rates investigating whether increasing ventilation can be expected to increase indoor $PM_{2.5}$ concentrations, or not, by evaluating the gradient of modelled $PM_{2.5}$ with respect to ventilation rate, i.e. $\partial p_m/\partial Q$. Fig. 7(a) presents a colour map indicating the sign and magnitude of $\partial p_m/\partial Q$, over reasonable ranges of outdoor $PM_{2.5}$ concentrations and per-person ventilation rates; with contours indicating lines of constant $\partial p_m/\partial Q$. Overlaid within the figure are coloured dots, one marking the mean of the data from each school in each academic year; coloured green for the academic year 2023–2024, and purple for 2024–2025.

It can be seen that the majority of the data in both school years is concentrated in regions where $\partial p_m/\partial Q$ is positive. Broadly, the distribution of observed ventilation rates is similar between the two years, but the particulate matter concentrations measured during the academic year 2024–2025 are seen to be significantly higher, by about 1.9 μg m^{-3}. This difference in outdoor $PM_{2.5}$ motivates examining Fig. 7(b), which shows the change in mean modelled indoor $PM_{2.5}$ with incremental increases in per-person ventilation rate, akin to the data shown in Fig. 6 but here with the data bracketed by academic year. In the academic year 2024–2025, a significantly greater increase in indoor $PM_{2.5}$ concentration would have been expected in the case of increased ventilation rates when compared to the academic year 2023–2024. In fact, the difference in the mean curves for the two years is similar in magnitude to the interquartile range exhibited in either year (as indicated by the shaded regions), highlighting the requirement for sufficient and recent data when attempting to draw inference from such modelling.

Some qualitative comparison may be made between the gradient with respect to ventilation rate that emerges from the GAM regression (Fig. 7(a)) to that seen for the dimensionless gradient, with respect to removal rate, in the duration-averaged $PM_{2.5}$ concentration for idealised naturally ventilated rooms (Fig. 2). This is particularly true for scenarios of the idealised room where the initial particulate matter concentration is significantly below the steady state condition (e.g. the top right pane of Fig. 2) — operational classrooms might be expected to exhibit lower particulate matter concentrations than outdoors at the start of the day as ingress is expected to be lessened overnight and deposition to have occurred.

5. Discussion

The physical model of idealised ventilated rooms shows that duration-averaged indoor $PM_{2.5}$ responds linearly to changes in outdoor concentration, but more complexly to changes in ventilation, including the potential for a non-monotonic response. The effect of increasing ventilation depends on the balance between outdoor concentration, the indoor sources, and the removal rates. Analysis of the dimensionless gradient of $\bar{p}_{in}$ with respect to the removal rate identified only two mechanisms by which increases in the ventilation rate can reduce duration-averaged $PM_{2.5}$ concentration. Firstly, if the initial indoor concentration exceeds the steady-state concentration, and the room removal rate is comparable to the inverse of the exposure-duration, then transient effects can result in increased ventilation resulting in reduced $\bar{p}_{in}$. The second mechanism exists when the steady-state concentration exceeds the outdoor concentration, which occurs only when the indoor sources are greater than the deposition divided by the I/O ratio. The mechanism exists when this condition is satisfied, but it is most likely to dominate when the ventilation rate is large, or for long exposure-durations. This second mechanism implies that increasing ventilation can still lower the potential for indoor exposures even when the outdoor concentration exceeds the indoor concentration. How these two mechanisms contribute (individually or in combination) to changes in the potential for pollutant exposure with ventilation is non-trivial. However, when neither mechanism is present, increasing ventilation is expected to increase the potential for particulate matter exposure indoors.

It is worth noting that, in this idealised model, deposition has the same mathematical form as in-room filtration — the effects of which can, unlike deposition, be easily controlled. Meaningful in-room filtration would therefore reduce the steady-state concentration without simultaneously increasing the ingress of outdoor $PM_{2.5}$, making the condition required for the second beneficial mechanism less readily satisfied and hence increasing the likelihood that increasing ventilation would increase indoor $PM_{2.5}$ concentrations, albeit from a lower baseline than would exist in the absence of filtration. In mechanically ventilated environments where the incoming air is filtered (or those where the penetration factor through the building envelope is low), the same basic physical model could be applied, with the addition of a factor to account for the reduction in incoming $PM_{2.5}$. This illustrates a broader utility of the physical model: even where data is unavailable, it can support qualitative predictions to be made between spaces with different distributions of parameters.

Exploiting the SAMHE dataset, from predominantly naturally ventilated UK schools, a statistical model was built using GAM regression informed by the measured daily mean indoor and outdoor $PM_{2.5}$ concentrations, and inferred daily mean ventilation rate. The model indicated that outdoor $PM_{2.5}$ concentration was by far the strongest predictor of indoor $PM_{2.5}$, with ventilation (and its interaction with outdoor $PM_{2.5}$) showing a comparatively modest association with indoor $PM_{2.5}$ concentration. This model showed an approximately linear dependence on the outdoor $PM_{2.5}$ concentration, consistent with the physical model, but further showed that the data associated with the 30% most-highly-ventilated school days exhibited very little dependence on the precise ventilation rate. The statistical model further provided predictions of the expected effect on daily mean $PM_{2.5}$ concentration when increasing ventilation rate (within this population of school measurements). In most cases indoor concentrations were predicted to increase with increased ventilation (with an average increase of 0.74 μg m^{-3} for ventilation increasing by 5 L s^{-1} p^{-1}) with the rate of increase predicted being greatest for ventilation rate increases of up to 5 L s^{-1} p^{-1}. This modelling highlighted that predictions vary strongly with outdoor concentrations, a fact that was also apparent upon grouping the data by academic year. Finally, comparing the gradient of modelled indoor $PM_{2.5}$ concentration with respect to ventilation rate, to that of the dimensionless gradient from the physical model (especially for intermediate initial concentrations), illustrates that simple physical models can provide insights that may be valuable when gathering the data for statistical modelling is more challenging.

6. Conclusion

The relationship between the potential for particulate matter exposures, and its rate of removal via ventilation, was investigated using both physical and statistical modelling. The physical modelling provided generalised findings and results for a range of space types. Both models were applied to UK school classrooms, which are typically

relatively poorly naturally ventilated spaces with relatively low to modest indoor sources of $PM_{2.5}$ present, and provided the conclusion that increasing the ventilation provision is expected to result in an increased potential for particulate matter exposure therein. Notably, should the ventilation be increased by a few litres-per-second-per-person then the increase in $PM_{2.5}$ concentration is expected to be only a fraction of a microgram-per-cubic-metre. The physical model indicates that opposite results are expected for environments like homes, i.e. increasing the ventilation provision is expected to result in a decreased potential for particulate matter exposure therein, because these spaces typically contain greater indoor sources of $PM_{2.5}$ (e.g. cooking, etc.). Mechanically ventilated offices, assuming the incoming air is not filtered, are likely to fall somewhere in between and more data on these spaces is required to draw any firm conclusions. A strength of combining physical and statistical modelling is that, whilst the observational dataset is necessarily limited to the schools in which measurements were made, the physical model provides a general framework that can inform expectations across a broader range of environments. For spaces characterised by larger indoor source terms, such as kitchens or workshops, or by greater effective removal rates Q^* for the same level of ventilation, from, for example, the presence of in-room filtration, the physical model can be used to anticipate qualitatively how the relationship between ventilation and indoor $PM_{2.5}$ might differ from that observed here, and to guide the design of future targeted data collection in such settings.

The findings presented herein should be set in the context of the broader context of indoor air quality management. The statistical modelling was applied specifically to the naturally ventilated UK schools studied, which are typically densely occupied, without strong indoor sources and rely on natural ventilation. Importantly, the findings of this study should not be interpreted as suggesting that ventilation is generally detrimental to indoor air quality nor occupant exposure and health. Ventilation serves multiple critical functions beyond particulate matter dilution, including the removal of exhaled CO_2, the reduction of potential airborne pathogen concentrations and associated infection risk, the dilution of indoor-generated pollutants, and the maintenance of thermal comfort — all of these benefits remain significant and worthy of consideration. The relatively small $PM_{2.5}$ increases associated with increased ventilation reported herein are specific to unfiltered ventilation in environments where outdoor $PM_{2.5}$ concentrations are the greatest contributor to indoor $PM_{2.5}$, and should be considered as only one factor among many to consider within the design and management of indoor air quality strategies.

CRediT authorship contribution statement

Samuel G.A. Wood: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alice E.E. Handy:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Henry C. Burridge:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

The "School Air quality Monitoring for Health and Education" project, SAMHE — an extension of the CO-TRACE project, was funded by the EPSRC under grant number EP/W001411/1, and received additional funding from the UK's Department for Education. This work was supported by UK Research and Innovation Building a Green Future strategic theme, The National Institute for Health and Care Research (NIHR), Medical Research Council (MRC), Natural Environment Research Council (NERC) and Engineering and Physical Sciences Research Council (EPSRC) APP28660 MR/Z50645X/1.

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.

Acknowledgments

The authors would like to thank all of the schools, teachers and pupils who have participated in the SAMHE project, and acknowledge the contribution of everyone involved, including members of the SAMHE Steering Committee and SAMHE Engagement Panel, for their continued support and guidance.

Appendix. Data density

Fig. 8 shows the density of the observed data used within the statistical modelling. The dataset contains 39 906 school-data days from 287 SAMHE monitors, providing dense coverage across the central range of observed outdoor $PM_{2.5}$ concentrations and inferred per-person ventilation rates.

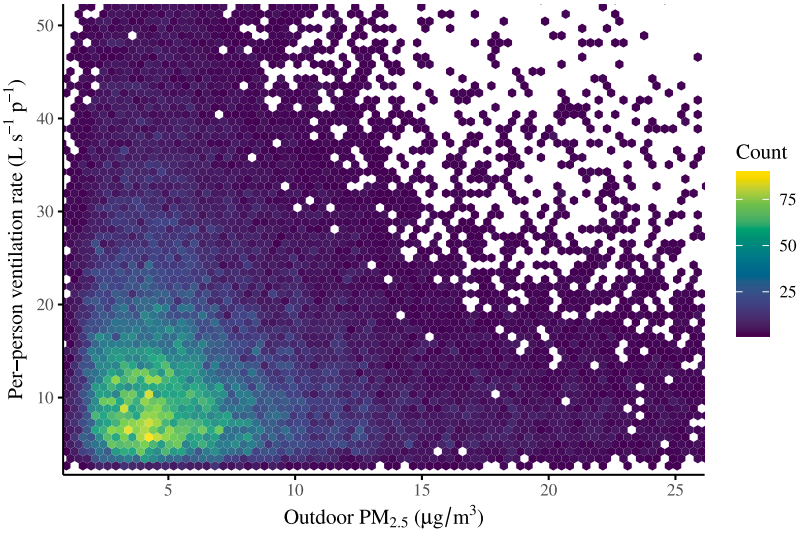

Fig. 8. Data density indicated as a heatmap of per-person ventilation rate and outdoor $PM_{2.5}$ measurements. White regions indicate those hexagons with no measurements.

References

[1] L. Chatzidiakou, R. Archer, V. Beale, S. Bland, H. Carter, C. Castro-Faccetti, H. Edwards, J. Finneran, S. Hama, R.L. Jones, et al., Schools' air quality monitoring for health and education: Methods and protocols of the SAMHE initiative and project, Dev. Built Environ. 16 (2023) 100266, http://dx.doi.org/10.1016/j.dibe.2023.100266.

[2] A.E.E. Handy, S.G.A. Wood, K. Roberts, C.S. Malley, H.C. Burridge, SAMHE Project Consortium, et al., Particulate matter concentrations in UK schools: A nationwide study into the influence of ambient $PM_{2.5}$ and the resulting exposure potentials, Sci. Total Environ. 970 (2025) 178875.

[3] S. Liu, R. Song, T.T. Zhang, Residential building ventilation in situations with outdoor $PM_{2.5}$ pollution, Build. Environ. 202 (2021) 108040.

[4] D. Park, T. Kim, S. Kim, K. Kang, J. Lee, B. Jeong, D. Kim, Impact of ventilation type on BC, $PM_{2.5}$, and CO_2 concentrations in urban residential buildings: A case study in South Korea, Energies 15 (3) (2022) 847.

[5] S. Zhang, D. Chen, X. Li, Seasonal variation of air purifier effectiveness and natural ventilation behavior: Implications for sustainable indoor air quality in London nurseries, Sustainability 17 (15) (2025) 7093, http://dx.doi.org/10.3390/su17157093.

[6] A. Ścibor, L. Samek, et al., Impact of wind speed and apartment ventilation on indoor concentrations of PM_{10} and $PM_{2.5}$ in Kraków, Poland, Air Qual. Atmos. & Health 13 (2020) 703–712, http://dx.doi.org/10.1007/s11869-020-00816-8.

[7] C. Chen, B. Zhao, Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor, Atmos. Environ. 45 (2) (2011) 275–288.

[8] D.H. Choi, D.H. Kang, Infiltration of ambient $PM_{2.5}$ through building envelope in apartment housing units in Korea, Aerosol Air Qual. Res. 17 (2) (2017) 598–607, http://dx.doi.org/10.4209/aaqr.2016.06.0287.

[9] B.R. Park, Y.S. Eom, D.H. Choi, D.H. Kang, Estimation of outdoor $PM_{2.5}$ infiltration into multifamily homes depending on building characteristics using regression models, Sustainability 13 (10) (2021) 5708, http://dx.doi.org/10.3390/su13105708.

[10] E. Barnes, et al., Characterizing outdoor infiltration and indoor contribution of $PM_{2.5}$ with citizen-based low-cost monitoring data, Environ. Pollut. 276 (2021) 116763, http://dx.doi.org/10.1016/j.envpol.2021.116763.

[11] Department for Education, BB 101: guidelines on ventilation, thermal comfort and indoor air quality in schools, 2018.

[12] A.C. Lai, W.W. Nazaroff, Modeling indoor particle deposition from turbulent flow onto smooth surfaces, J. Aerosol Sci. 31 (4) (2000) 463–476.

[13] J.-H. Kim, M.-S. Yeo, Effect of flow rate and filter efficiency on indoor $PM_{2.5}$ in ventilation and filtration control, Atmosphere 11 (10) (2020) 1061, http://dx.doi.org/10.3390/atmos11101061.

[14] P. Kumar, S. Hama, R.A. Abbass, K.V. Abhijith, A. Tiwari, D. Grassie, C. Mitsakou, Environmental quality in sixty primary and secondary school classrooms in London, J. Build. Eng. 91 (2024) 109549, http://dx.doi.org/10.1016/j.jobe.2024.109549.

[15] B. Zhao, J. Wu, Particle deposition in indoor environments: Analysis of influencing factors, J. Hazard. Mater. 147 (1–2) (2007) 439–448.

[16] M. Mohammadi, J. Calautit, Quantifying the transmission of outdoor pollutants into the indoor environment and vice versa—Review of influencing factors, methods, challenges and future direction, Sustainability 14 (17) (2022) 10880.

[17] S. Wei, S. Semple, Exposure to fine particulate matter ($PM_{2.5}$) from non-tobacco sources in homes within high-income countries: A systematic review, Air Qual. Atmos. & Health 16 (3) (2023) 553–566.

[18] M.H. Sherman, et al., Residential air-change rates: A critical review, Indoor Air 31 (4) (2021) 921–933, http://dx.doi.org/10.1111/ina.12814.

[19] Ministry of Housing, Communities & Local Government, English housing survey 2023–24: Headline report, 2025, https://www.gov.uk/government/statistics/chapters-for-english-housing-survey-2023-to-2024-headline-findings-on-housing-quality-and-energy-efficiency. (Accessed 18 Jebruary 2025).

[20] D. Bousiotis, D.S. Sanghera, J. Carrington, G. Hodgkiss, F. Jajarmi, K.Z. Rajab, F.D. Pope, Parameterising the effect of human occupancy and kinetic energy on indoor air pollution, Npj Clim. Atmos. Sci. 9 (1) (2026) 4.

[21] A.R. Ferro, R.J. Kopperud, L.M. Hildemann, Source strengths for indoor human activities that resuspend particulate matter, Environ. Sci. Technol. 38 (6) (2004) 1759–1764.

[22] ANSI/ASHRAE standard 62.1-2022: Ventilation for acceptable indoor air quality, 2022, American Society of Heating, Refrigerating and Air-Conditioning Engineers.

[23] S.E. West, L. Way, R. Archer, V.J. Beale, S. Bland, H. Burridge, C. Castro-Faccetti, L. Chatzidiakou, P. Kumar, C. Vouriot, N. Williams, SAMHE Project Consortium, Co-designing an air quality web app with school pupils and staff: The SAMHE web app, Citiz. Sci.: Theory Pr. 8 (1) (2023) 64, http://dx.doi.org/10.5334/cstp.620.

[24] S.G.A. Wood, A.E.E. Handy, K. Roberts, H.C. Burridge, Assessing classroom ventilation rates using CO_2 data from a nationwide study of UK schools and identifying school-wide correlation factors, Dev. Built Environ. 19 (2024) 100520.

[25] R. Jayaratne, X. Liu, P. Thai, M. Dunbabin, L. Morawska, The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog, Atmos. Meas. Tech. 11 (8) (2018) 4883–4890.

[26] Department for Environment, Food & Rural Affairs (DEFRA), AURN (automatic urban and rural network) site classification, 2024, https://uk-air.defra.gov.uk/networks/site-types. (Accessed 16 July 2024).

[27] A.E. Handy, S.G. Wood, H.C. Burridge, SAMHE Project Consortium, et al., Analysing longitudinal indoor air quality data: robust principal component analysis of the "Schools' air quality monitoring for health and education" timeseries data, Indoor Environ. 3 (1) (2026) 100146.

[28] J. Finneran, H. Burridge, Inferring ventilation rates with quantified uncertainty in operational rooms using point measurements of carbon dioxide: Classrooms as a case study, Build. Environ. 253 (2024) 111309.

[29] A.K. Persily, Field measurement of ventilation rates, Indoor Air 26 (1) (2016) 97–111.

[30] T.J. Hastie, R.J. Tibshirani, Generalized Additive Models, Chapman and Hall/CRC, 1990.

[31] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.