

RESEARCH

Open Access



# Assessing the health estimation capacity of air pollution exposure prediction models

Jenna R. Krall<sup>1\*</sup> , Joshua P. Keller<sup>2</sup> and Roger D. Peng<sup>3</sup>

## Abstract

**Background:** The era of big data has enabled sophisticated models to predict air pollution concentrations over space and time. Historically these models have been evaluated using overall metrics that measure how close predictions are to monitoring data. However, overall methods are not designed to distinguish error at timescales most relevant for epidemiologic studies, such as day-to-day errors that impact studies of short-term health associations.

**Methods:** We introduce frequency band model performance, which quantifies health estimation capacity of air quality prediction models for time series studies of air pollution and health. Frequency band model performance uses a discrete Fourier transform to evaluate prediction models at timescales of interest. We simulated fine particulate matter (PM<sub>2.5</sub>), with errors at timescales varying from acute to seasonal, and health time series data. To compare evaluation approaches, we use correlations and root mean squared error (RMSE). Additionally, we assess health estimation capacity through bias and RMSE in estimated health associations. We apply frequency band model performance to PM<sub>2.5</sub> predictions at 17 monitors in 8 US cities.

**Results:** In simulations, frequency band model performance rates predictions better (lower RMSE, higher correlation) when there is no error at a particular timescale (e.g., acute) and worse when error is added to that timescale, compared to overall approaches. Further, frequency band model performance is more strongly associated ( $R^2 = 0.95$ ) with health association bias compared to overall approaches ( $R^2 = 0.57$ ). For PM<sub>2.5</sub> predictions in Salt Lake City, UT, frequency band model performance better identifies acute error that may impact estimated short-term health associations.

**Conclusions:** For epidemiologic studies, frequency band model performance provides an improvement over existing approaches because it evaluates models at the timescale of interest and is more strongly associated with bias in estimated health associations. Evaluating prediction models at timescales relevant for health studies is critical to determining whether model error will impact estimated health associations.

**Keywords:** Ambient air pollution, Chemical transport model, Epidemiology, Exposure assessment, Fourier transform, Health assessment, Particulate matter

## Background

The United States has for decades monitored air pollution levels via the Environmental Protection Agency's network of monitors as well as state and local monitors

[36]. These monitors tend to be sited around large urban centers or around significant sources of pollution, and as a result, large swaths of the country are typically not monitored and knowledge of air pollution concentrations in those areas has historically been minimal. In the past, one major consequence of this lack of monitoring is that many areas of the country could not be included in studies examining the health associations of air pollutants [2, 39, 44, 55]. The era of big data along with novel machine

\*Correspondence: jkrall@gmu.edu

<sup>1</sup> Department of Global and Community Health, George Mason University, 4400 University Drive, MS 5B7, Fairfax, VA 22030, USA  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

learning techniques and statistical models have allowed us to predict ambient air pollution concentrations with greater accuracy and precision than in the past [3, 18, 19, 28, 32]. This new generation of models provides air pollution prediction at finer spatial and temporal scales by leveraging multiple sources of data such as satellite data, computer weather models, chemistry models, land use data, emissions source information, and/or pollution monitoring data. These input data sources have varying strengths, for example models may incorporate data from monitors that provide ground truth observations, but have generally limited spatial coverage, and from chemical transport models such as the Community Multiscale Air Quality (CMAQ) model, which have good spatial coverage but are often biased [25, 51]. Predicting air pollution concentrations for epidemiologic applications is challenging and must balance the shortcomings of each input data source.

Given the development of modern prediction models, it is natural to want to evaluate their performance. However, the nature of model evaluation depends critically on the application in which the model will be applied [10, 32]. Without information about study-specific context, it is impossible to provide an unqualified assessment of a model that is informative about the specific application. For example, studies of the acute health associations of ambient air pollution typically focus on day-to-day variation in pollutants and health outcomes [2, 44, 55], suggesting that models predicting pollution concentrations for such studies need to do well predicting the higher frequency temporal components. Studies of chronic health associations of pollution often make broad comparisons across larger geographies [15, 20, 29, 33], suggesting that prediction models there need to reproduce spatial variations in pollution at larger scales, but not finer scale fluctuations.

To evaluate exposure prediction models, metrics such as  $R^2$ , root mean square error (RMSE), or normalized root mean square error are commonly used to quantify how predictions vary from ground truth observations [7, 10, 11, 14, 17, 25, 49, 51]. More recent studies have proposed correlation [35] and variance ratios [7], which also capture overall deviations in variability, as being more relevant for evaluating prediction models for use in health studies. Although all these metrics evaluate similarity between model predictions and observations, they do not focus on those errors in prediction models that will most impact estimated health associations. As an example, suppose model predictions correlate well with the observed data at the seasonal and monthly temporal scales, but correlations are low at the day-to-day scale. Metrics such as  $R^2$  are impacted by performance at all temporal scales and will not highlight poor performance

at the day-to-day scale most relevant for acute epidemiologic studies. Conversely, for a model that performs well primarily at the day-to-day scale, these metrics may be overly pessimistic for model performance in an acute epidemiologic study.

Existing methods for *evaluating* prediction models do not incorporate temporal and spatial scales of interest. This presents a gap in the use of novel methods, such as machine learning approaches, for exposure estimation in epidemiologic studies. Therefore, the objective of the present research is to propose a new approach to evaluate prediction models that focuses on temporal scales that will most impact estimated health associations. We demonstrate how usual model performance metrics such as RMSE can fail to capture errors in prediction models that are relevant to epidemiologic studies. Additionally, we propose frequency band model evaluation to determine whether a given prediction model will provide good estimates of health associations. Our approach evaluates the health estimation capacity of prediction models by focusing model evaluation on the timescale of interest of the health effect. We illustrate model evaluation using particulate matter air pollution less than  $2.5 \mu\text{m}$  in aerodynamic diameter ( $\text{PM}_{2.5}$ ), though our methods are applicable to prediction models of air pollutants generally (e.g., ozone and nitrogen dioxide ( $\text{NO}_2$ )).

## Methods

We observe a time series of ground-truth observations  $z(t)$ , which may represent air pollution measurements (e.g., for  $\text{PM}_{2.5}$ ) taken at a central site monitor. The goal of an exposure prediction model is to accurately replicate such observations with a predicted time series  $z^*(t)$  that can be computed at locations and times without monitoring data. The predicted series  $z^*(t)$  may represent output from a regression model, a machine learning algorithm, a computer simulation model such as the CMAQ model, or any combination of these approaches. While we will focus on temporal data series  $z(t)$  and predictions  $z^*(t)$  indexed by time  $t$ , we discuss extensions of these ideas to the spatial domain in the Discussion.

Our goal is to compare a time series of model predictions,  $z^*(t)$ , with a reference time series of ground-truth observations,  $z(t)$ , for times  $t=1, \dots, n$ . Existing approaches include quantifying prediction accuracy using correlation  $r = \sqrt{R^2} = \text{Cor}(z^*(t), z(t))$  and  $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (z^*(t) - z(t))^2}$  [7, 11, 14, 35]. Additional existing approaches include the log variance ratio (*LVR*), defined as  $\log\left(\frac{\text{Var}(z^*(t))}{\text{Var}(z(t))}\right)$  [7]. The *LVR* captures differences in temporal variation between models, which can impact precision in estimated health associations. We refer to these approaches as *overall model*

performance measures, which we denote by  $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ .

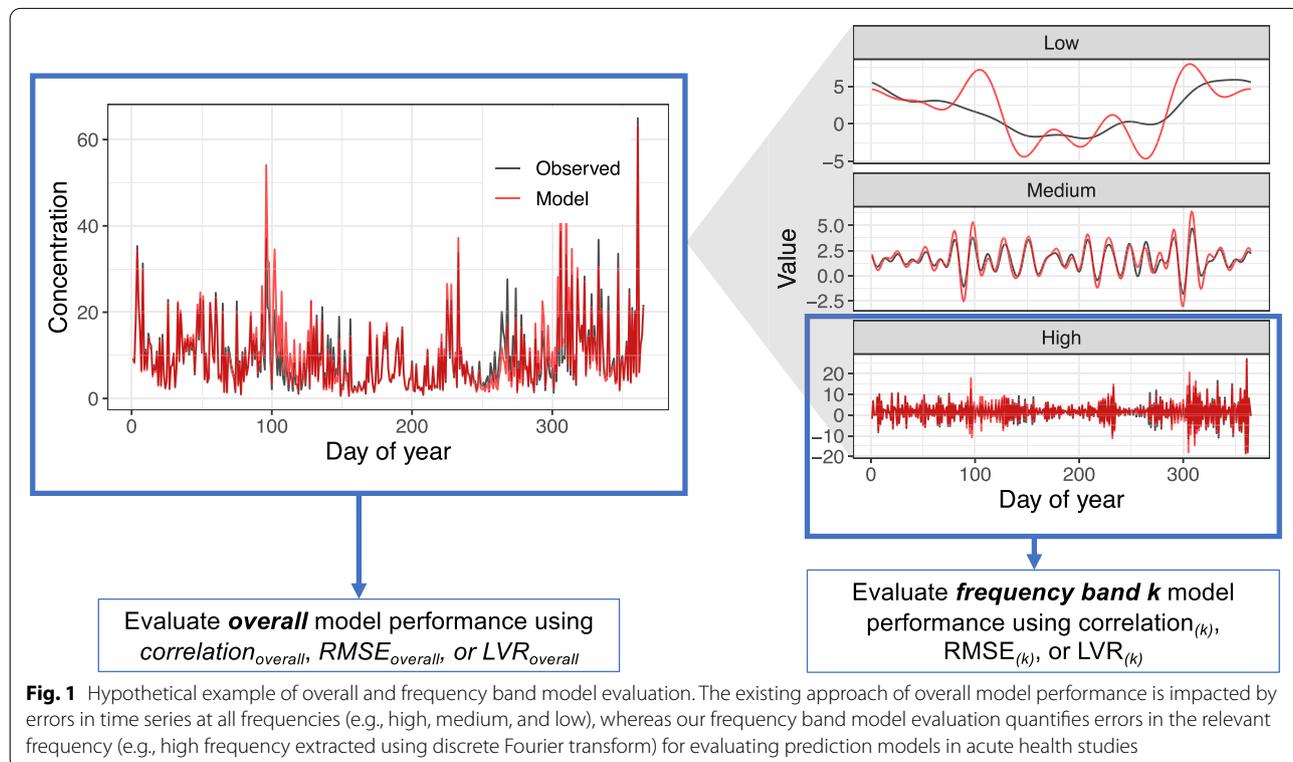
Overall model performance can be impacted by model errors at timescales different than the timescale of interest. Therefore to better capture health estimation capacity, we propose *frequency band model performance*. Frequency band model performance differs from existing approaches by restricting the model predictions  $z^*(t)$  and observations  $z(t)$  to their timescale-specific components to create measures  $r_{(k)}$ ,  $RMSE_{(k)}$ , and  $LVR_{(k)}$ , which represent correlation, RMSE, and LVR for a chosen range of frequencies denoted as band  $k$ . To extract the frequency band  $k$  components, we use a discrete Fourier transform [6, 23, 40]. For our reference time series,  $z(t)$ , we partition the range of frequencies  $[0, n/2)$  into non-overlapping frequency bands  $k=1, \dots, K$  such that

$$z(t) = \sum_{k=1}^K z_k(t) \tag{1}$$

The same approach is used to partition  $z^*(t)$  into components  $z_k^*(t)$ . Then,  $r_{(k)} = \text{Cor}(z_k^*(t), z_k(t))$ ,  $RMSE_{(k)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_k^*(t) - z_k(t))^2}$ , and  $LVR_{(k)} = \log\left(\frac{\text{Var}(z_k^*(t))}{\text{Var}(z_k(t))}\right)$ . To facilitate comparisons between overall and frequency band  $k$  RMSE, we scale  $RMSE_{overall}$  and  $RMSE_{(k)}$  by the standard deviation of the reference time series  $z(t)$  and  $z_k(t)$ , respectively.

To understand the advantage of frequency band model evaluation, consider a hypothetical time series of  $PM_{2.5}$  observations and model predictions (Fig. 1). Existing overall model performance metrics ( $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ ) are applied to the time series on the left, which is impacted by differences between observations and model predictions at all frequencies. In contrast, frequency band  $k$  model performance restricts evaluation of model performance to the frequency of interest using a discrete Fourier transform (e.g., high frequency, such as day-to-day variation) and is not impacted by errors at other frequencies (e.g., medium frequency, such as monthly variation, and low frequency, such as seasonal variation) (Fig. 1).

As in previous work estimating health associations of PM air pollution [23], we set  $K=6$  and consider frequency bands of  $k=1$ : [1,6) cycles per year corresponding to seasonal components,  $k=2$ : [6,12) cycles per year,  $k=3$ : [12,26) cycles per year,  $k=4$ : [26,52) cycles per year,  $k=5$ : [52,104) cycles per year, and  $k=6$ : [104,183) cycles per year corresponding to acute components. Our particular interest for acute health associations of air pollution is in the acute timescale captured by the frequency band  $k=6$  model performance ( $k=6$ : [104,183) cycles per year), corresponding to a timescale of a few days or less. However, frequency band model performance can be applied to any timescales that are of interest.



**Fig. 1** Hypothetical example of overall and frequency band model evaluation. The existing approach of overall model performance is impacted by errors in time series at all frequencies (e.g., high, medium, and low), whereas our frequency band model evaluation quantifies errors in the relevant frequency (e.g., high frequency extracted using discrete Fourier transform) for evaluating prediction models in acute health studies

We first conduct a simulation study to determine how errors at specific frequency bands impact overall model performance measures  $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$  as well as frequency band  $k$  model performance measures  $r_{(k)}$ ,  $RMSE_{(k)}$ , and  $LVR_{(k)}$ . We hypothesize that errors at a specific frequency band  $k'$  impact the overall model performance ( $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ ) and the model performance at the same frequency band  $k'$  ( $r_{(k')}$ ,  $RMSE_{(k')}$ , and  $LVR_{(k')}$ ), but not model performance at different frequency bands  $k \neq k'$ . For example, seasonal error that occurs within the first frequency band ( $k=1$  corresponding to [1,6] cycles per year) would impact model performance measured by  $r_{overall}$ ,  $RMSE_{overall}$ ,  $LVR_{overall}$ ,  $r_{(1)}$ ,  $RMSE_{(1)}$ , and  $LVR_{(1)}$ , but not model performance at higher frequency bands relevant for estimating acute health associations (e.g.,  $r_{(6)}$ ,  $RMSE_{(6)}$ , and  $LVR_{(6)}$ ). This implies that if we are primarily interested in model predictions at frequency band  $k'$ , model performance measured by  $r_{(k')}$ ,  $RMSE_{(k')}$ , and  $LVR_{(k')}$  will better reflect errors relevant for frequency band  $k'$  while limiting the influence of errors at other frequency bands  $k \neq k'$ .

To simulate observed air pollution time series that are non-negative and right-skewed, let  $z(t) = \exp(x(t)\sigma_x + \mu_x)$ , where  $x(t)$  is a time series with  $\text{Var}(x(t)) = 1$ , and where  $\mu_x$  and  $\sigma_x$  represent the mean and standard deviation of the log-transformed time series. We specify  $\mu_x = 1.9 \log \mu \text{ g/m}^3$  and  $\sigma_x = 0.6 \log \mu \text{ g/m}^3$  to reflect log-transformed concentrations of  $\text{PM}_{2.5}$  in New York City from 2010 to 2018. We simulate the time series  $x(t) = \frac{\sum_{k \in \{1,2,6\}} x_k(t)}{\sqrt{\text{Var}(\sum_{k \in \{1,2,6\}} x_k(t))}}$  consisting of frequency bands  $k = 1, 2, 6$  to approximately reflect seasonal, monthly, and acute time trends found in air pollution concentrations. Each  $x_k(t)$  is specified using cosine functions with varying wavelengths (details in [Additional file, Section A](#)) and  $\text{Var}(x_k(t)) = 1$ . As a sensitivity analysis, we increase the relative seasonal variation to reflect observed relative variability across timescales in  $\text{PM}_{2.5}$  data, i.e.  $\text{Var}(x_{(1)}(t)) \in \{1.5, 2\}$ . We simulate 100 observed time series with 3 years of data each ( $n = 1095$ ).

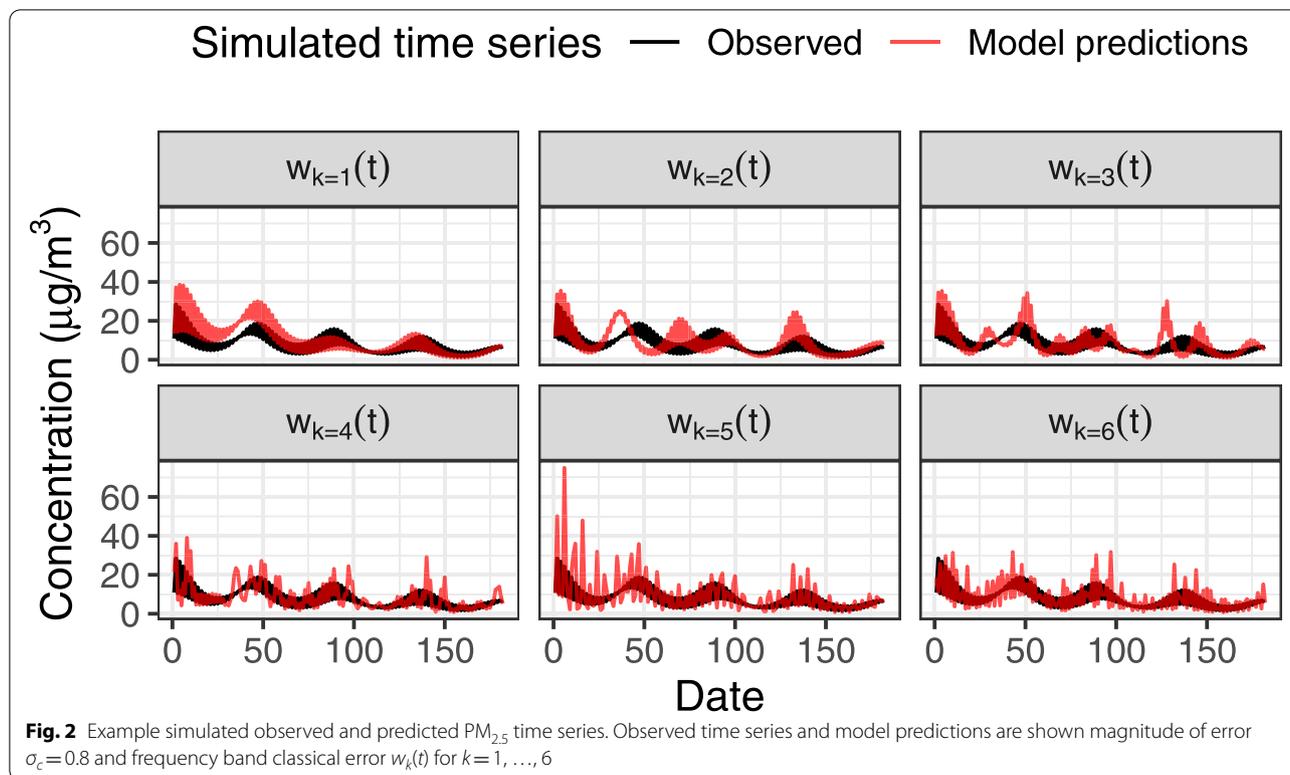
We simulate model predictions by incorporating classical measurement error at varying timescales into simulated observed time series. Our model predictions are simulated as  $z^*(t) = \exp(x^*(t)\sigma_x + \mu_x)$ , with  $\mu_x = 1.9$  and  $\sigma_x = 0.6$  as in the simulated observed time series. The log-transformed simulated model predictions  $x^*(t) = x(t) + w_k(t)\sigma_c$ , where  $x(t)$  is the simulated observed time series (log-transformed with  $\text{Var}(x(t)) = 1$ ) and  $w_k(t)$  is the standardized error component at frequency band  $k$  with  $\text{Var}(w_k(t)) = 1$ . The magnitude of error is represented by the standard deviation  $\sigma_c \in \{0.2, 0.4, 0.6, 0.8\}$ . We obtain  $w_k(t)$  as the standardized  $k$  frequency band component from a discrete Fourier transform of standard normal error. Therefore, on the logarithmic scale, our simulated model predictions are the simulated

observed time series with classical error at frequency band  $k$ . As  $k$  varies from 1, ..., 6, the timescale of the error changes from seasonal ( $k=1$ ) to acute ( $k=6$ ) (Fig. 2). For each of our 100 simulated observed time series with  $n = 1095$ , we simulate 24 model prediction time series with varying classical measurement error (four varying magnitudes of error  $\sigma_c$  and six error frequencies  $w_k(t)$   $k=1, \dots, 6$ ).

Health counts (e.g., number of deaths or emergency department (ED) visits for day  $t$ ) are simulated as  $\text{Poisson}(\mu(t))$  where  $\mu(t) = \exp(\beta_0 + \beta_1 z^{\text{residual}}(t) + 0.03 z^{\text{fitted}}(t))$ . The residuals  $z^{\text{residual}}(t)$  and fitted  $z^{\text{fitted}}(t)$  are obtained from a regression of the simulated observed time series  $z(t)$  on a natural spline of time with 24 degrees of freedom to capture seasonal and monthly trends. Therefore, the residuals  $z^{\text{residual}}(t)$  represent the sub-monthly frequency components of  $z(t)$  relevant for acute health associations. We specify the acute health association  $\beta_1 = \log(1.1)/10$  corresponding to a relative risk of 1.1 per  $10 \mu \text{ g/m}^3$  increase in  $\text{PM}_{2.5}$ . The base rate of health  $\beta_0 = 5$  yields approximately 200 health counts per day (i.e., deaths or ED visits) and is selected to reflect cardiorespiratory ED visits observed in large U.S. cities [34].

We estimate health associations for simulated model predictions with varying magnitudes of classical error ( $\sigma_c$ ) and error  $w_k(t)$  at frequencies 1, ..., 6. The health association model is an overdispersed Poisson time series regression model controlling for non-acute temporal trends with a natural spline of time with 280 degrees of freedom per year to isolate the acute time series. This number of degrees of freedom is much larger than generally used in the epidemiologic literature to properly adjust for  $w_k(t)$  error added at  $k=5$  [52, 104] cycles per year. In practice, confounding in epidemiologic studies of acute health associations is controlled for by both a smooth function of time (generally with <10 degrees of freedom per year) as well as smooth functions for meteorology including temperature and humidity [37, 38, 44, 45, 50]. To evaluate the estimated health associations for the model predictions, we compute the estimated health association RMSE and percent relative mean bias in the estimated health association (mean bias/ $\beta_1 \times 100$ ). Using scatterplots and  $R^2$ , we examine associations of both mean overall model performance and frequency band  $k$  model performance with health association RMSE and percent relative mean bias across 100 simulated datasets.

We also compare observations of  $\text{PM}_{2.5}$  to predictions from an exposure model to demonstrate both overall model performance and frequency band  $k=6$  model performance. To represent “ground-truth” observations, we develop a dataset of observed daily  $\text{PM}_{2.5}$  data in  $\mu \text{ g/m}^3$  from 17 monitors across 8 US cities from 2010 to 2017 using the US Environmental Protection Agency’s (US EPA) Air Quality System. We select monitors from



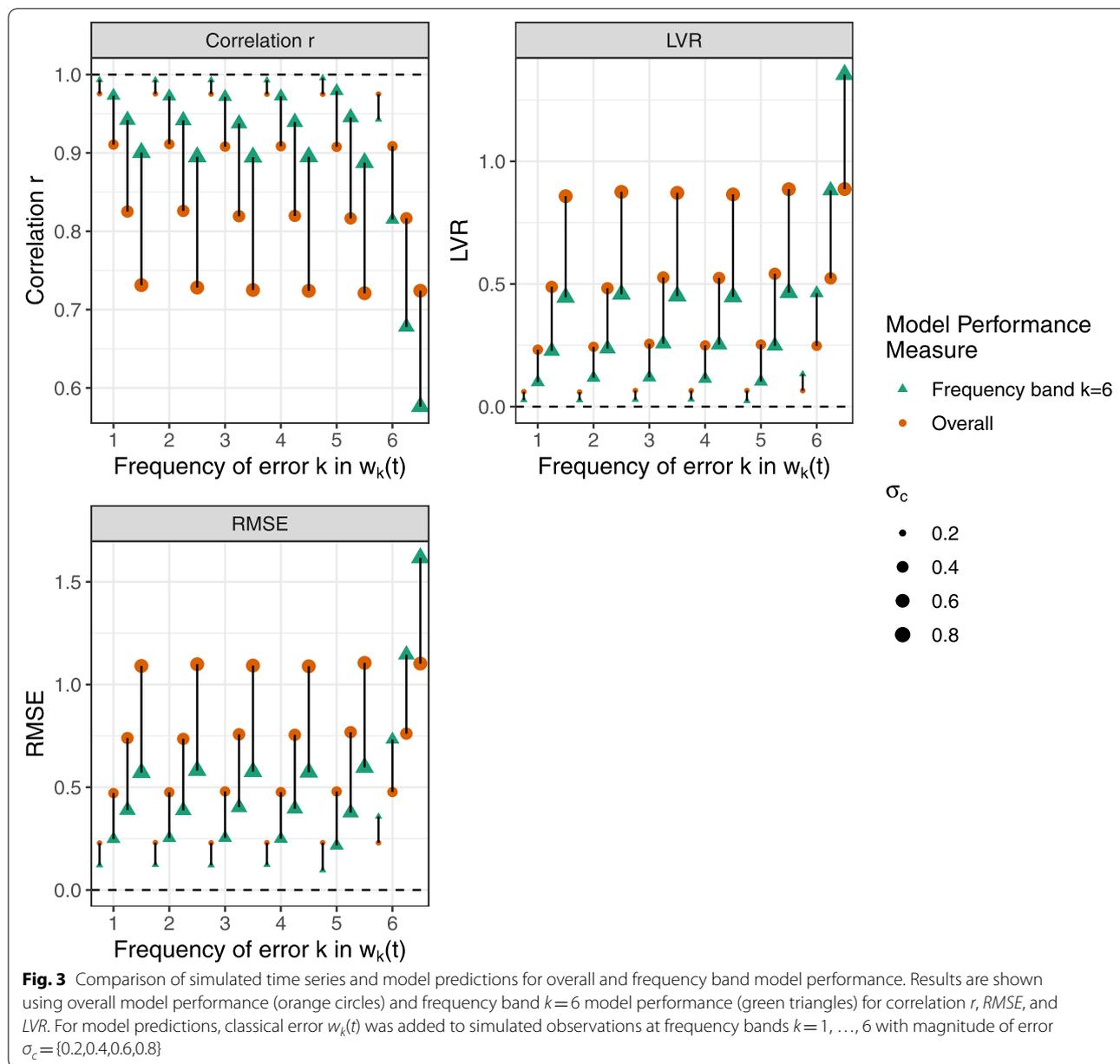
cities based on geographic locations throughout the continental US and based on previous studies of air pollution and health [23, 34, 44, 55], including Atlanta, GA (number of monitoring sites:  $n = 4$ ); Dallas, TX ( $n = 1$ ); Houston, TX ( $n = 1$ ); Los Angeles, CA ( $n = 5$ ); New York City, NY ( $n = 1$ ); Pittsburgh, PA ( $n = 2$ ); Seattle/Tacoma, WA ( $n = 1$ ); and Salt Lake City, UT ( $n = 2$ ). For each monitor, we interpolate  $PM_{2.5}$  means for short gaps (<10 days) in the data for each city to create uninterrupted time series. We include only monitors that had at least 1 year of daily concentrations after interpolation. We utilize the longest complete time series for each monitor.

As our prediction model, we utilize predictions from the Fused Air Quality Surface Using Downscaling (FAQSD) approach at 2010 US census tracts from the US EPA [52]. FAQSD uses a Bayesian space-time model to fuse monitoring data with CMAQ model predictions and develop predictions at 2010 US census tracts [3–5, 42]. CMAQ is an atmospheric chemical transport model that provides predictions  $12 \times 12$  km resolution grids across the US [12, 13] and may be calibrated or fused with observed monitoring data [21]. We link FAQSD predictions to monitors using the census tract where the monitor is located. We compare the observed  $PM_{2.5}$  data and FAQSD using overall performance measures ( $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ ) and frequency band  $k=6$  model performance ( $r_{(6)}$ ,  $RMSE_{(6)}$ , and  $LVR_{(6)}$ ), where

$k=6$  corresponds to [104,183) cycles per year and represents variation relevant for acute health associations at timescales <3.5 days [23]. All analyses were conducted using R version 4.0 [41].

### Results

In our simulation study, we evaluate model performance by comparing  $z(t)$  and  $z^*(t)$  using overall model performance ( $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ ) and frequency band  $k$  model performance ( $r_{(k)}$ ,  $RMSE_{(k)}$ , and  $LVR_{(k)}$ ). We focus on frequency band  $k=6$  model performance, ( $r_{(6)}$ ,  $RMSE_{(6)}$ , and  $LVR_{(6)}$ ), which evaluates the high frequency of the component of the model and is most relevant for estimating acute health associations. Fig. 3 shows the mean across 100 simulated datasets for overall model performance measures (orange circles) and frequency band  $k=6$  model performance measures (green triangles) for error  $w_k(t)$ ,  $k=1, \dots, 6$ . With error  $w_k(t)$  at frequency bands  $k=1, \dots, 5$ , the frequency band  $k=6$  model performance ( $r_{(6)}$ ,  $RMSE_{(6)}$ , and  $LVR_{(6)}$ ) rated the prediction model better compared to the overall model performance ( $r_{overall}$ ,  $RMSE_{overall}$ , and  $LVR_{overall}$ ). When error is added using the high frequency band  $k=6$ , i.e., component  $w_{(6)}(t)$ , the frequency band  $k=6$  model performance rated the prediction model worse compared to the overall model performance. The results are consistent across model performance measures of  $r$ ,  $RMSE$  and



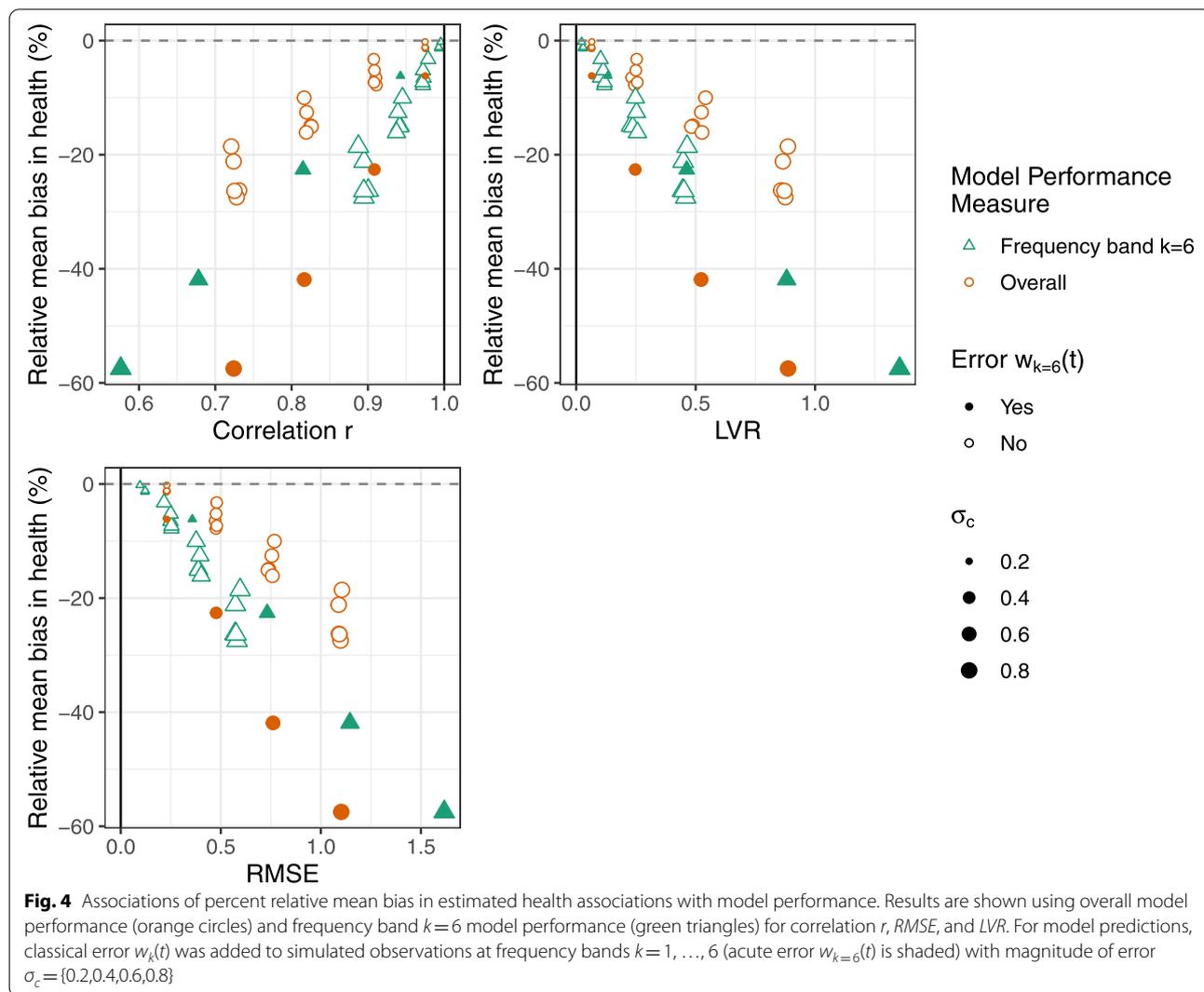
$LVR$ . Further, the results are consistent for frequency band  $k=1, 2$  model performance (Additional file, Fig. S1), with frequency band  $k$  model performance better reflecting errors at frequency band  $k$ . We did not examine frequency band  $k=3, \dots, 5$  model performance because the simulated observed time series did not have variation at those frequencies. In summary, overall model performance can be both overly pessimistic when the prediction model has error at timescales not relevant to the study design and overly optimistic when the prediction model has error at relevant timescales. Frequency band  $k$  model performance better reflects error at timescales of interest.

As a sensitivity analysis, we increase the relative variance of the seasonal component of the simulated observed time series  $\text{Var}(x_{(1)}(t)) \in \{1^2, 1.5^2, 2^2\}$  while keeping  $\text{Var}(x_{(2)}(t)) = \text{Var}(x_{(6)}(t)) = 1$  and the total variance  $\text{Var}(x(t)) = 1$  to reflect observed timescale variability in  $PM_{2.5}$  concentrations. Changing the variance of the seasonal component alone did not impact overall model performance. However the frequency band  $k=6$  model performance rates the models worse with increasing  $\text{Var}(x_{(1)}(t))$  (Additional file, Fig. S2). Because the entire time series is scaled such that  $\text{Var}(x(t)) = 1$ , the acute component of the simulated time series has decreasing

variance as  $\text{Var}(x_{(1)}(t))$  increases. In other words, holding the total variability of the time series constant, the same magnitude of error  $\sigma_c$  has a stronger impact on the acute component of the time series when the acute component has lower relative variance.

To determine whether overall model performance or frequency band  $k$  model performance better capture health estimation capacity, we examine associations of model performance measures ( $r$ ,  $RMSE$ ,  $LVR$ ) with bias in estimated health associations as well as estimated health association RMSE. We focus on frequency band  $k=6$  model performance that evaluates the high frequency component of the time series relevant for estimating acute health associations. We expect larger bias and RMSE in estimated acute health associations for model predictions with high frequency error  $w_{k=6}(t)$ . Figure 4 shows the percent relative mean bias against

the mean overall model performance (orange circles) and the mean frequency band  $k=6$  model performance (green triangles). The solid points indicate when acute frequency band error  $w_{(6)}(t)$  is added and therefore expected to strongly impact bias in the acute health association, and outlined otherwise. The size of the point indicates the magnitude of error added. For the frequency band  $k=6$  model performance,  $r_{(6)}$ ,  $RMSE_{(6)}$ , and  $LVR_{(6)}$  are more strongly associated with bias compared to the overall performance measures ( $r_{\text{overall}}$ ,  $RMSE_{\text{overall}}$ , and  $LVR_{\text{overall}}$ ). Similarly, the association between frequency band  $k=6$  model performance and health RMSE is also stronger compared to the overall performance (Fig. 5). The linear association  $R^2$  with acute health association measures (percent relative mean bias and health RMSE) is stronger for frequency band  $k=6$  model performance compared to overall model performance



(Table 1). For example, percent relative mean bias was more strongly associated with frequency band  $k=6$   $RMSE_{(6)}$  ( $R^2=0.95$ ) compared to overall  $RMSE_{overall}$  ( $R^2=0.57$ ).

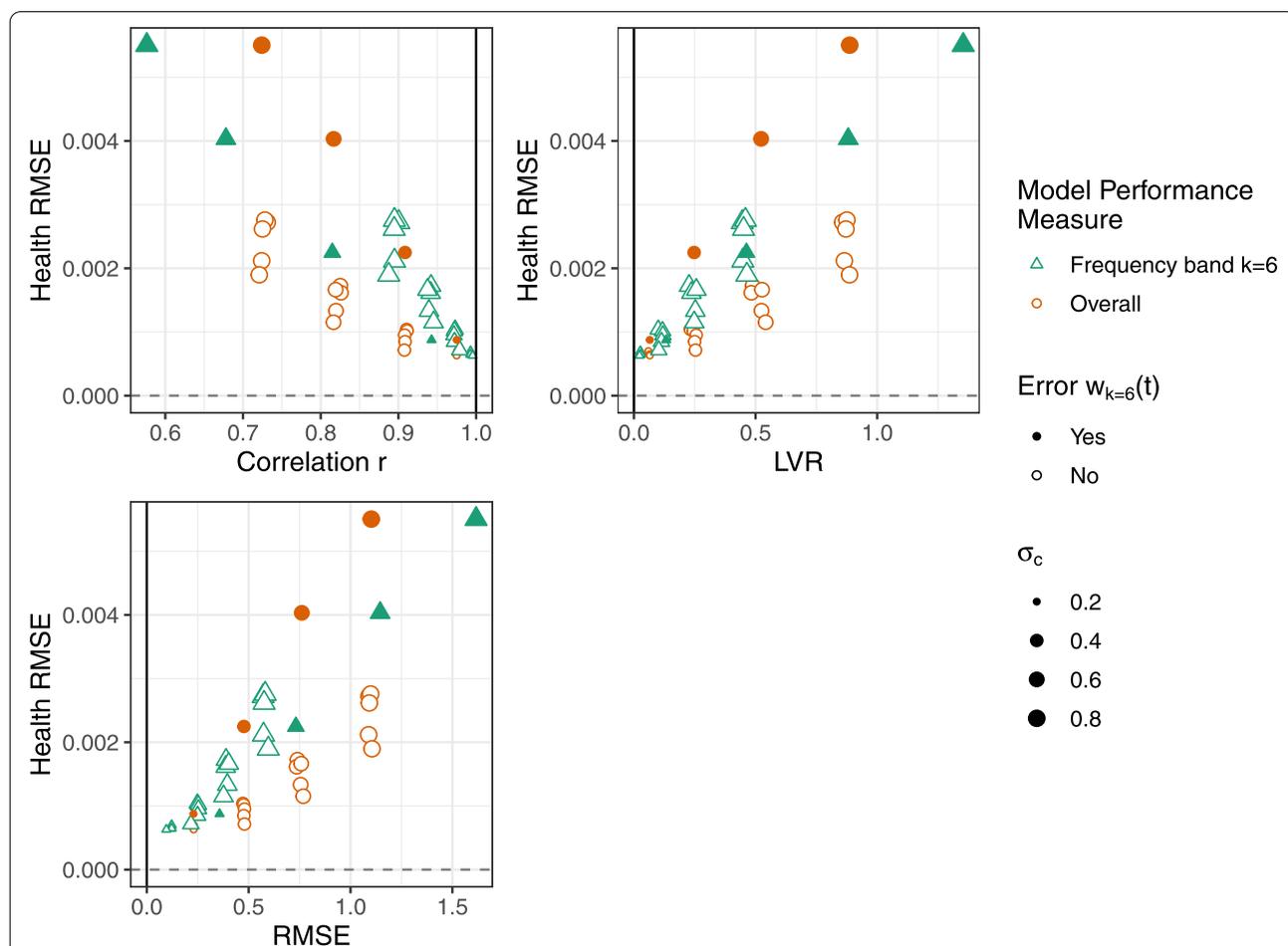
For the analysis of  $PM_{2.5}$  predictions, the locations of the selected  $PM_{2.5}$  monitors were spread across the US (Additional file, Fig. S3). The available daily observations range from 593 days (monitor site 420030008 in Pittsburgh) to 2351 days (monitor site 360810124 in New York City) (Additional file, Table S1). The lowest median  $PM_{2.5}$  concentration is in Seattle/Tacoma ( $5.4 \mu g/m^3$ ) and the highest in Los Angeles ( $12 \mu g/m^3$ ). Figure 6 compares the overall concentration time series and three frequency band components from a discrete Fourier transform (Eq. 1):  $k=1$  or the seasonal component,  $k=2$  or the monthly component, and  $k=6$  the acute component. The monitors include 060374008 in Los Angeles where FAQSD and the monitor differ considerably at all timescales, 131210032 in Atlanta where FAQSD and the

**Table 1**  $R^2$  for the linear relationship of exposure metrics with health measures

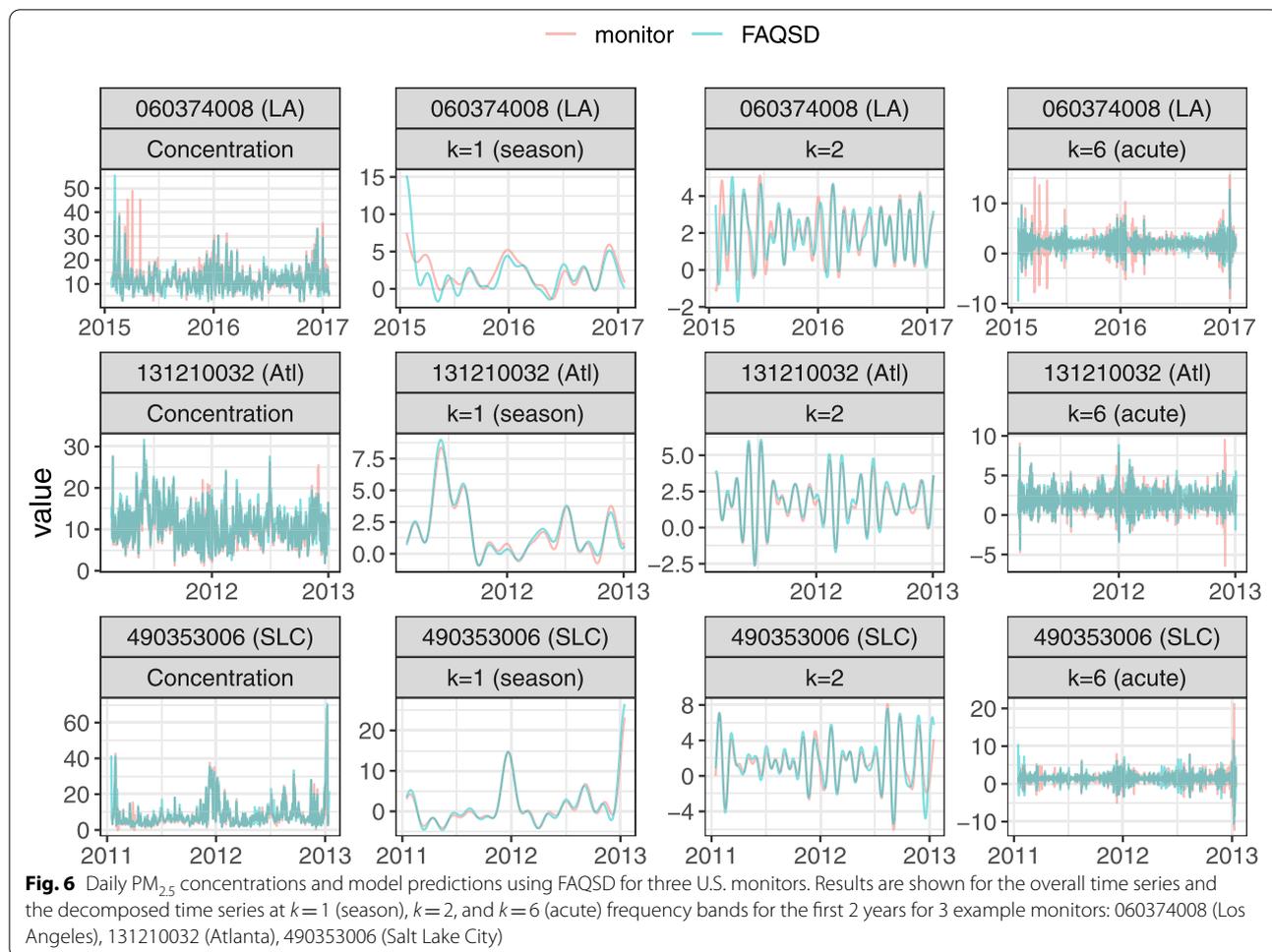
Health measure	Metric	Frequency band $k=6$	Overall
Rel. mean bias (%)	Correlation $r$	0.89	0.57
	LVR	0.96	0.58
	RMSE	0.95	0.57
Health RMSE	Correlation $r$	0.90	0.52
	LVR	0.96	0.53
	RMSE	0.94	0.52

monitor are similar at all timescales, and 490353006 in Salt Lake City where FAQSD performs similarly to the monitor at longer timescales (monthly, seasonal), but not at shorter timescales (acute).

Comparing overall performance and frequency band  $k=6$  model performance of FAQSD for all 17 monitors,



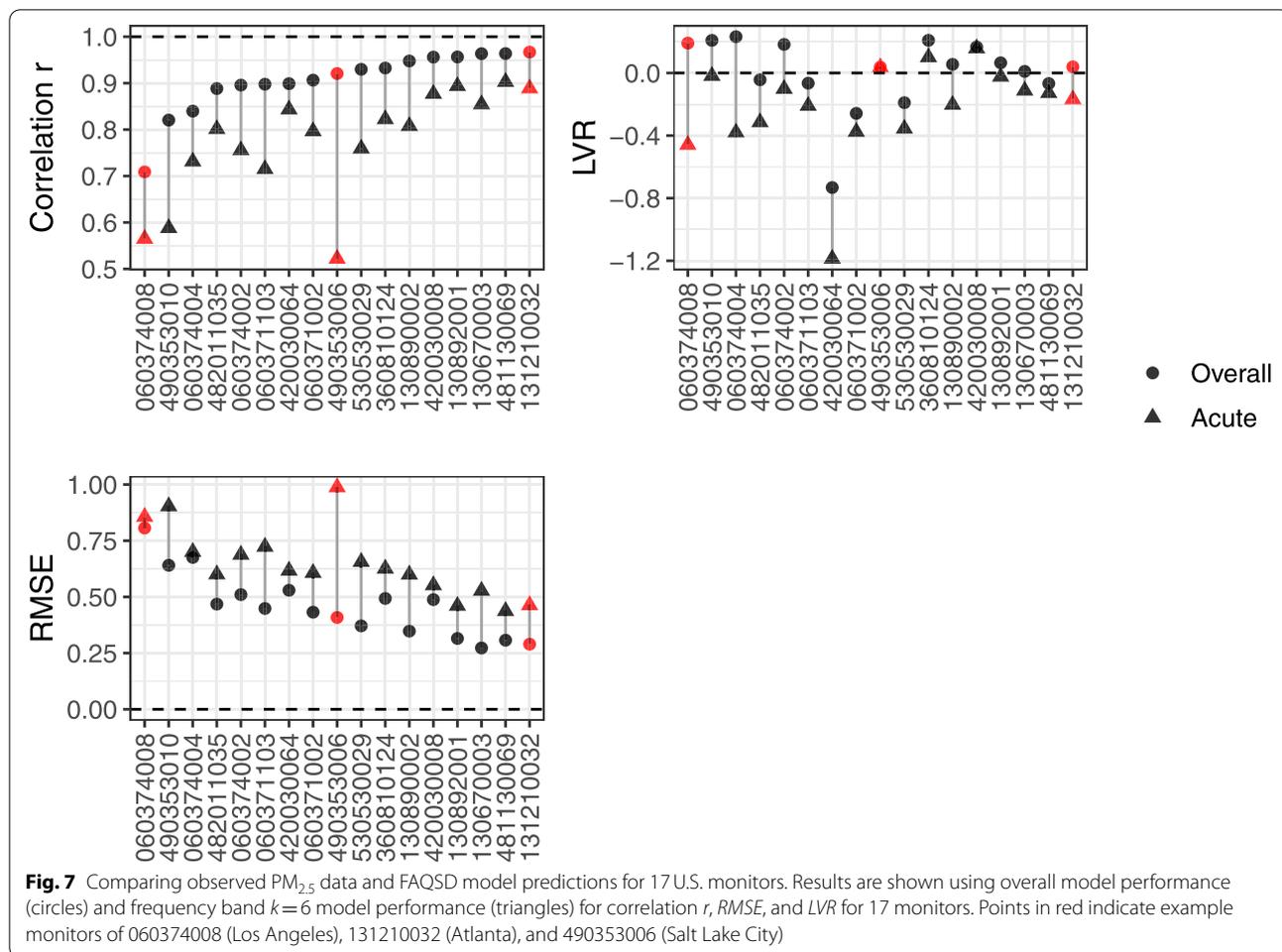
**Fig. 5** Associations of health association RMSE with model performance. Results are shown using model performance (orange circles) and frequency band  $k=6$  model performance (green triangles) for correlation  $r$ ,  $RMSE$ , and  $LVR$ . For model predictions, classical error  $w_k(t)$  was added to simulated observations at frequency bands  $k=1, \dots, 6$  (acute error  $w_{k=6}(t)$  is shaded) with magnitude of error  $\sigma_c = \{0.2, 0.4, 0.6, 0.8\}$



the overall performance measures rate FAQSD better compared to frequency band  $k=6$  model performance (Fig. 7). The three example monitors are shown in red. For overall model performance, correlations below  $r=0.89$  ( $R^2=0.8$ ) are considered low. For 060374008 in Los Angeles, both the overall and frequency band  $k=6$  measures rate FAQSD low ( $r=0.71$  and  $r=0.56$ , respectively). For 131210032 in Atlanta, both the overall and frequency band  $k=6$  model performance measures rate FAQSD well ( $r=0.97$  and  $r=0.89$ , respectively). However, at 490353006 in Salt Lake City, there is a large discrepancy between the overall and frequency band  $k=6$  correlation and RMSE, where the performance of FAQSD may be overrated using the overall approach and may be overly optimistic about its acute health estimation capacity (e.g.,  $r=0.92$  and  $r=0.52$ , respectively). This is likely driven by the good performance of FAQSD at this monitor at the seasonal timescale, but not at the acute timescale (Fig. 6).

### Discussion

We propose frequency band model performance for evaluating health estimation capacity of air quality prediction models. When comparing model predictions to truth in simulations, frequency band  $k$  correlation  $r_{(k)}$ ,  $RMSE_{(k)}$ , and  $LVR_{(k)}$  better reflect error at specific timescales compared to overall metrics. Of particular relevance to acute epidemiologic studies, frequency band  $k=6$  model performance penalizes models for error at acute timescales, with lower correlation  $r$  and higher RMSE, while reporting higher model performance when error is not present at the acute timescale. Furthermore in simulations of estimated acute health associations, frequency band  $k=6$  model performance is more strongly associated with relative mean bias and RMSE in estimated health associations. In a study of 8 US cities, overall model performance and frequency band  $k=6$  model performance can rate prediction models differently, emphasizing the need for a model performance tool that is best suited to the proposed analysis.



Recent studies have evaluated or compared the performance of air quality prediction models [11, 14, 32, 35]. Although many previous studies used primarily overall  $RMSE$  and correlation ( $r$ ) for model performance [7, 11, 14, 35],  $LVR$  can capture differences relevant for precision of estimated health associations [7]. Whether overall or frequency band model performance is used, examining multiple metrics such as  $r$ ,  $RMSE$ , and  $LVR$ , can help elucidate different aspects of model performance. Further, while we demonstrate frequency band model evaluation using  $PM_{2.5}$ , our approach can be directly applied to evaluate prediction models for other pollutants examined in previous studies such as  $NO_2$  [7, 11, 14, 35] and ozone [7, 35].

Effects of measurement error in time series studies of air pollution and health has been extensively examined using simulation studies [7, 8, 22, 27, 48]. Studies have examined spatial errors [48], error type [27], as well as effects of measurement error in multipollutant models [22] and multi-level models [7, 8]. Our work adds to this literature by simulating measurement error

at varying timescales. This can better reflect practice where a prediction model may have errors in the seasonal component, but not the acute component, or vice versa.

Previous epidemiologic studies have used timescale decomposition approaches to determine health associations of air pollution at varying timescales [23, 46]. As in previous work [23], we decompose the time series into components corresponding to different timescales using a discrete Fourier transform [6, 40]. More recent epidemiologic studies utilized distributed lag models to estimate health associations at varying timescales [26, 47, 53, 54]. For analyses of health associations at different time scales, frequency band model performance can be applied in the planning stage of an analysis before health data are collected to evaluate whether a prediction model can be effectively used for the proposed timescales of interest.

Aside from epidemiologic studies, there are analyses for which overall model performance will be more appropriate compared to frequency band model performance.

For analyses estimating burden of disease due to air pollution, representing the true concentrations, and not short or long-scale variability, is most important [24, 30]. Furthermore, both frequency band and overall model performance represent “operational evaluation” of the model biases [16] that describe deviations of the predictions from the truth. Depending on use, air quality prediction models should be evaluated with respect to multiple performance features, including operational evaluation, “diagnostic evaluation” of whether model errors are driven by inputs, and other features [16]. A recent study of CMAQ, along with other air quality models and statistical modeling approaches, did not find substantial differences in regional and national scale spatial PM<sub>2.5</sub> predictions from these models, but the authors note that the best model may vary with how the model will be used [32].

We incorporated classical measurement error in our simulation study. Additive classical measurement error biases estimated health associations [9, 56]. In practice, the error may be a more complex combination of both Berkson and classical error [7]. Additive error on the log scale, as we simulated in this work, can introduce bias in estimated health associations when the error type is classical or Berkson [27]. The goal of the present study was to demonstrate that error at varying timescales was better captured by frequency band model performance compared to overall model performance, and not to examine measurement error types.

Frequency band model performance can be directly applied to time series studies of health associations, such as studies of acute health associations. Future work could extend our method to epidemiologic studies of long-term exposure to air pollution and health. Extensions of frequency band model evaluation to long-term epidemiologic studies will need to address several challenges. Assessing effects of long-term exposure on health relies on accurately representing spatial contrasts in concentrations across different locations. The natural extension of frequency band model performance to a spatial setting would require applying a two-dimensional Fourier transformation, which necessitates gridded data. Prediction model output is often available over spatial grids [18, 43], and a recent study estimated associations of PM<sub>2.5</sub> and birthweight at varying spatial scales using wavelet decomposition [1]. Similarly, an assessment of unmeasured spatial confounding at different scales compared Fourier and wavelet decompositions of exposure [31]. However, the spatial distribution of monitoring locations do not follow a regular grid, limiting the application of the frequency band approach to these contexts. Increasing the spatial density of the monitoring network, perhaps through the use of low-cost monitors, could mitigate this issue and

allow for decomposition of the spatial scales of variation in monitoring data.

## Conclusions

Frequency band model performance can be applied to predictions from air quality models to evaluate performance at timescales of interest for epidemiologic studies. Compared to commonly-used overall model evaluation approaches, frequency band model evaluation better reflects error at timescales of interest and is more strongly associated with bias in estimated health associations. Multiple metrics should be used to evaluate the performance of air quality models. When the model predictions will be used for health analyses, it is important to evaluate the model performance at timescales that will impact the estimated health associations.

## Abbreviations

CMAQ: Community multiscale air quality model; RMSE: Root mean square error; PM<sub>2.5</sub>: Particulate matter less than 2.5  $\mu$  m; NO<sub>2</sub>: Nitrogen dioxide; LVR: Log variance ratio; ED: Emergency department; FAQSD: Fused air quality surface using downscaling.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12940-022-00844-0>.

**Additional file 1.** Methods for simulation study and additional figures and tables.

## Acknowledgements

Not applicable.

## Authors' contributions

JRK conceptualized the simulation study, performed all analyses, and contributed substantially to the manuscript. JPK advised on the methods, simulation design, and interpretation of results, and contributed substantially to the manuscript. RDP conceptualized the study, advised on the methods, and contributed substantially to the manuscript. All authors read and approved the final manuscript.

## Funding

Dr. Krall was supported in part by the Thomas F. and Kate Miller Jeffress Memorial Trust, Bank of America, Trustee. Dr. Peng was supported in part by the US Environmental Protection Agency (EPA) through award RD835871. This work has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the agency. EPA does not endorse any products or commercial services mentioned in this publication.

## Availability of data and materials

The datasets and code for the data analysis are available as an R package on GitHub [<https://github.com/kralljr/tsfreqband>].

## Declarations

### Ethics approval and consent to participate

This work does not include any data on human subjects.

### Consent for publication

This work does not include any data on human subjects.

**Competing interests**

The authors declare no competing interests.

**Author details**

<sup>1</sup>Department of Global and Community Health, George Mason University, 4400 University Drive, MS 5B7, Fairfax, VA 22030, USA. <sup>2</sup>Department of Statistics, Colorado State University, 1877 Campus Delivery, Fort Collins, CO 80523, USA. <sup>3</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St, Baltimore, MD 21205, USA.

Received: 2 September 2021 Accepted: 25 February 2022

Published online: 17 March 2022

**References**

- Antonelli J, Schwartz J, Kloog I, Coull BA. Spatial multiresolution analysis of the effect of PM<sub>2.5</sub> on birth weights. *Ann Appl Stat*. 2017;11(2):792–807. <https://doi.org/10.1214/16-AOAS1018>.
- Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. Ozone and short-term mortality in 95 US urban communities, 1987–2000. *JAMA*. 2004;292(19):2372–8. <https://doi.org/10.1001/jama.292.19.2372>.
- Berrocal VJ, Gelfand AE, Holland DM. A bivariate space-time downscaler under space and time misalignment. *Ann Appl Stat*. 2010a;4(4):1942–75. <https://doi.org/10.1214/10-aos351>.
- Berrocal VJ, Gelfand AE, Holland DM. A spatio-temporal downscaler for output from numerical models. *J Agric Biol Environ Stat*. 2010b;15(2):176–97. <https://doi.org/10.1007/s13253-009-0004-z>.
- Berrocal VJ, Gelfand AE, Holland DM. Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics*. 2012;68(3):837–48. <https://doi.org/10.1111/j.1541-0420.2011.01725.x>.
- Bloomfield P. *Fourier analysis of time series: an introduction*. New York, NY: Wiley; 2004.
- Butland BK, Samoli E, Atkinson RW, Barratt B, Beevers SD, Kitwiroon N, et al. Comparing the performance of air pollution models for nitrogen dioxide and ozone in the context of a multilevel epidemiological analysis. *Environ Epidemiol*. 2020;4(3):e093. <https://doi.org/10.1097/EE9.00000.0000000093>.
- Butland BK, Samoli E, Atkinson RW, Barratt B, Katsouyanni K. Measurement error in a multi-level analysis of air pollution and health: a simulation study. *Environ Health*. 2019;18(1):13. <https://doi.org/10.1186/s12940-018-0432-8>.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement error in nonlinear models: a modern perspective*. New York, NY: Chapman & Hall/CRC; 2006.
- Chang JC, Hanna SR. Air quality model performance evaluation. *Meteorol Atmos Phys*. 2004;87:167–96. <https://doi.org/10.1007/s00703-003-0070-7>.
- Chen J, de Hoogh K, Gulliver J, Hoffmann B, Hertel O, Ketzler M, et al. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environ Int*. 2019;130:104934. <https://doi.org/10.1016/j.envint.2019.104934>.
- CMAQ: The Community Multiscale Air Quality Modeling System. (2021). U.S. Environmental Protection Agency.
- Community Modeling and Analysis System. (2021). UNC Institute for the environment.
- Cowie CT, Garden F, Jegasothy E, Knibbs LD, Hanigan I, Morley D, et al. Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian maximum entropy model, in estimating NO<sub>2</sub> for a birth cohort in Sydney, Australia. *Environ Res*. 2019;174:24–34. <https://doi.org/10.1016/j.envres.2019.03.068>.
- Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. Ambient PM<sub>2.5</sub>, O<sub>3</sub>, and NO<sub>2</sub> exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment cohort (CANHEC). *Environ Health Perspect*. 2015;123(11):1180–6. <https://doi.org/10.1289/ehp.1409276>.
- Dennis R, Fox T, Fuentes M, Gilliland A, Hanna S, Hogrefe C, et al. A framework for evaluating regional-scale numerical photochemical modeling systems. *Environ Fluid Mech*. 2010;10(4):471–89. <https://doi.org/10.1007/s10652-009-9163-2>.
- Derwent D, Fraser A, Abbott J, Jenkin M, Willis P, Murrells T. Evaluating the performance of air quality models. DEFRA report, vol. 3; 2010. [https://uk-air.defra.gov.uk/assets/documents/reports/cat05/1006241607\\_100608\\_MIP\\_Final\\_Version.pdf](https://uk-air.defra.gov.uk/assets/documents/reports/cat05/1006241607_100608_MIP_Final_Version.pdf)
- Di Q, Amiri H, Shi L, Kloog I, Silvern R, Kelly J, et al. An ensemble-based model of PM<sub>2.5</sub> concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int*. 2019;130:104909. <https://doi.org/10.1016/j.envint.2019.104909>.
- Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing PM<sub>2.5</sub> exposures with high spatiotemporal resolution across the continental United States. *Environ Sci Technol*. 2016;50(9):4712–21. <https://doi.org/10.1021/acs.est.5b06121>.
- Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. Air pollution and mortality in the Medicare population. *N Engl J Med*. 2017;376(26):2513–22. <https://doi.org/10.1056/NEJMoa1702747>.
- Diao M, Holloway T, Choi S, O'Neill SM, Al-Hamdan MZ, Van Donkelaar A, et al. Methods, availability, and applications of PM<sub>2.5</sub> exposure estimates derived from ground measurements, satellite, and atmospheric models. *J Air Waste Manag Assoc*. 2019;69(12):1391–414. <https://doi.org/10.1080/10962247.2019.1668498>.
- Dionisio KL, Chang HH, Baxter LK. A simulation study to quantify the impacts of exposure measurement error on air pollution health risk estimates in copollutant time-series models. *Environ Health*. 2016;15(1):114. <https://doi.org/10.1186/s12940-016-0186-0>.
- Dominici F, McDermott A, Zeger SL, Samet JM. Airborne particulate matter and mortality: timescale effects in four US cities. *Am J Epidemiol*. 2003;157(12):1055–65. <https://doi.org/10.1093/aje/kwg087>.
- Ford B, Heald CL. Exploring the uncertainty associated with satellite-based estimates of premature mortality due to exposure to fine particulate matter. *Atmos Chem Phys*. 2016;16(5):3499–523. <https://doi.org/10.5194/acp-16-3499-2016>.
- Friberg MD, Kahn RA, Holmes HA, Chang HH, Sarnat SE, Tolbert PE, et al. Daily ambient air pollution metrics for five cities: evaluation of data-fusion-based estimates and uncertainties. *Atmos Environ*. 2017;158:36–50. <https://doi.org/10.1016/j.atmosenv.2017.03.022>.
- Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med*. 2010;29(21):2224–34. <https://doi.org/10.1002/sim.3940>.
- Goldman GT, Mulholland JA, Russell AG, Strickland MJ, Klein M, Waller LA, et al. Impact of exposure measurement error in air pollution epidemiology: effect of error type in time-series studies. *Environ Health*. 2011;10(6):1. <https://doi.org/10.1186/1476-069X-10-61>.
- Hu X, Belle JH, Meng X, Wildani A, Waller LA, Strickland MJ, et al. Estimating PM<sub>2.5</sub> concentrations in the conterminous United States using the random forest approach. *Environ Sci Technol*. 2017;51(12):6936–44. <https://doi.org/10.1021/acs.est.7b01210>.
- Jerrett M, Shankardass K, Berhane K, Gauderman WJ, Künzli N, Avol E, et al. Traffic-related air pollution and asthma onset in children: a prospective cohort study with individual exposure measurement. *Environ Health Perspect*. 2008;116(10):1433–8. <https://doi.org/10.1289/ehp.10968>.
- Jin X, Fiore AM, Civerolo K, Bi J, Liu Y, Van Donkelaar A, et al. Comparison of multiple PM<sub>2.5</sub> exposure products for estimating health benefits of emission controls over New York State, USA. *Environ Res Lett*. 2019;14(8):084023. <https://doi.org/10.1088/1748-9326/ab2dcb>.
- Keller JP, Szpiro AA. Selecting a scale for spatial confounding adjustment. *J R Stat Soc Ser A, (Statistics in Society)*. 2020;183(3):1121–43. <https://doi.org/10.1111/rssa.12556>.
- Kelly JT, Jang C, Timin B, Di Q, Schwartz J, Liu Y, et al. Examining PM<sub>2.5</sub> concentrations and exposure using multiple models. *Environ Res*. 2021;196:110432. <https://doi.org/10.1016/j.envres.2020.110432>.
- Kioumourtzoglou M-A, Schwartz JD, Weisskopf MG, Melly SJ, Wang Y, Dominici F, et al. Long-term PM<sub>2.5</sub> exposure and neurological hospital admissions in the northeastern United States. *Environ Health Perspect*. 2016;124(1):23–9. <https://doi.org/10.1289/ehp.1408973>.
- Krall JR, Chang HH, Waller LA, Mulholland JA, Winquist A, Talbott EO, et al. A multicity study of air pollution and cardiorespiratory emergency department visits: comparing approaches for combining estimates across cities. *Environ Int*. 2018;120:312–20. <https://doi.org/10.1016/j.envint.2018.07.033>.

35. Lin C, Heal MR, Vieno M, MacKenzie IA, Armstrong BG, Butland BK, et al. Spatiotemporal evaluation of EMEP4UK-WRF v4.3 atmospheric chemistry transport simulations of health-related metrics for NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, and PM<sub>2.5</sub> for 2001–2010. *Geosci Model Dev*. 2017;10(4):1767–87. <https://doi.org/10.5194/gmd-10-1767-2017>.
36. National Research Council. Air quality management in the United States. Washington, D.C.: National Academies Press; 2004.
37. Ostro B, Roth L, Malig B, Marty M. The effects of fine particle components on respiratory hospital admissions in children. *Environ Health Perspect*. 2009;117(3):475–80. <https://doi.org/10.1289/ehp.11848>.
38. Peng RD, Bell ML, Geyh AS, McDermott A, Zeger SL, Samet JM, et al. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environ Health Perspect*. 2009;117(6):957–63. <https://doi.org/10.1289/ehp.0800185>.
39. Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*. 2002;287(9):1132–41. <https://doi.org/10.1001/jama.287.9.1132>.
40. Priestley M. Spectral analysis and time series. In: *Multivariate series prediction and control*, vol. 2; 1981.
41. R Core Team. (2020). R: a language and environment for statistical computing. R Foundation for statistical computing. <https://www.R-project.org/>.
42. Reff A, Phillips S, Eyth A, Mintz D. Bayesian space-time downscaling fusion model (downscaler) -derived estimates of air quality for 2010. U.S. Research Triangle Park, NC: Environmental Protection Agency; 2014.
43. Reich BJ, Chang HH, Foley KM. A spectral method for spatial downscaling. *Biometrics*. 2014;70(4):932–42. <https://doi.org/10.1111/biom.12196>.
44. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *N Engl J Med*. 2000;343(24):1742–9. <https://doi.org/10.1056/NEJM200012143432401>.
45. Samoli E, Stafoggia M, Rodopoulou S, Ostro B, Declercq C, Alessandrini E, et al. Associations between fine and coarse particles and mortality in Mediterranean cities: results from the MED-PARTICLES project. *Environ Health Perspect*. 2013;121(8):932–8. <https://doi.org/10.1289/ehp.1206124>.
46. Schwartz J. Harvesting and long term exposure effects in the relation between air pollution and mortality. *Am J Epidemiol*. 2000a;151(5):440–8. <https://doi.org/10.1093/oxfordjournals.aje.a010228>.
47. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*. 2000b;11(3):320–6. <https://doi.org/10.1097/00001648-200005000-00016>.
48. Strickland MJ, Gass KM, Goldman GT, Mulholland JA. Effects of ambient air pollution measurement error on health effect estimates in time-series studies: a simulation-based analysis. *J Expo Sci Environ Epidemiol*. 2015;25:160–6. <https://doi.org/10.1038/jes.2013.16>.
49. Thunis P, Pederzoli A, Pernigotti D. Performance criteria to evaluate air quality modeling applications. *Atmos Environ*. 2012;59:476–82. <https://doi.org/10.1016/j.atmosenv.2012.05.043>.
50. Tolbert PE, Klein M, Peel JL, Sarnat SE, Sarnat JA. Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. *J Expo Sci Environ Epidemiol*. 2007;17(S2):S29–35. <https://doi.org/10.1038/sj.jes.7500625>.
51. Tong DQ, Mauzerall DL. Spatial variability of summertime tropospheric ozone over the continental United States: implications of an evaluation of the CMAQ model. *Atmos Environ*. 2006;40(17):3041–56. <https://doi.org/10.1016/j.jatmosenv.2005.11.058>.
52. US EPA. (2020, December 15). RSIG-related downloadable data files [data and tools]. <https://www.epa.gov/hesc/rsig-related-downloadable-data-files>.
53. Welty LJ, Peng RD, Zeger SL, Dominici F. Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality. *Biometrics*. 2009;65(1):282–91. <https://doi.org/10.1111/j.1541-0420.2007.01039.x>.
54. Wilson A, Chiu Y-HM, Hsu H-HL, Wright RO, Wright RJ, Coull BA. Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children's health. *Biostatistics*. 2017;18(3):537–52. <https://doi.org/10.1093/biostatistics/kxx002>.
55. Zanobetti A, Schwartz J. The effect of fine and coarse particulate air pollution on mortality: a national analysis. *Environ Health Perspect*. 2009;117(6):898–903. <https://doi.org/10.1289/ehp.0800108>.
56. Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al. Exposure measurement error in time-series studies of air pollution:

concepts and consequences. *Environ Health Perspect*. 2000;108(5):419–26. <https://doi.org/10.1289/ehp.00108419>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

