RESEARCH

Association between ambient air pollutants and upper respiratory tract infection and pneumonia disease burden in Thailand from 2000 to 2022: a high frequency ecological analysis

Esther Li Wen Choo[1,2†], A. Janhavi[1,2†], Joel Ruihan Koo[1], Steve H. L. Yim[3,4,5], Borame L Dickens[1*] and Jue Tao Lim[1,3]

Abstract

Background A pertinent risk factor of upper respiratory tract infections (URTIs) and pneumonia is the exposure to major ambient air pollutants, with short term exposures to different air pollutants being shown to exacerbate several respiratory conditions.

Methods Here, using disease surveillance data comprising of reported disease case counts at the province level, high frequency ambient air pollutant and climate data in Thailand, we delineated the association between ambient air pollution and URTI/Pneumonia burden in Thailand from 2000 – 2022. We developed mixed-data sampling methods and estimation strategies to account for the high frequency nature of ambient air pollutant concentration data. This was used to evaluate the effects past concentrations of fine particulate matter ($PM_{2.5}$), sulphur dioxide (SO_2), and carbon monoxide (CO) and the number of disease case count, after controlling for the confounding meteorological and disease factors.

Results Across provinces, we found that past increases in CO, SO_2, and $PM_{2.5}$ concentration were associated to changes in URTI and pneumonia case counts, but the direction of their association mixed. The contributive burden of past ambient air pollutants on contemporaneous disease burden was also found to be larger than meteorological factors, and comparable to that of disease related factors.

Conclusions By developing a novel statistical methodology, we prevented subjective variable selection and discretization bias to detect associations, and provided a robust estimate on the effect of ambient air pollutants on URTI and pneumonia burden over a large spatial scale.

Keywords Upper respiratory tract infections, Pneumonia, Ambient air pollutants, Weather, Environmental epidemiology, Mixed data sampling methods

†Esther Li Wen Choo and A. Janhavi contributed equally to this work.

*Correspondence:
Borame L Dickens
ephdbsl@nus.edu.sg
Full list of author information is available at the end of the article

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 2 of 13

Introduction

Upper Respiratory Tract Infections (URTIs), such as pneumonia and influenza, are usually characterised by irritation and swelling of the upper airways. They are caused by a variety of bacteria and viruses and the infection can vary from a mild cold to life-threatening pneumonia [1]. Being one of the most common diseases, the global burden of URTI in 2019 is estimated to be 17.2 billion [2]. Although a majority of URTIs are quite harmless, the estimated economic burden of non-influenza related viral URTIs in the United States alone is estimated to be 22.5 billion USD [3] and an estimated 2 billion USD is spent on over-the-counter treatments for URTIs [4].

A pertinent risk factor of URTIs is the exposure to major ambient air pollutants, such as carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), particulate matter (PM) and ozone (O_3) [5]. Short term exposures to different air pollutants have shown to have detrimental immunological effects, and can exacerbate several respiratory conditions, including URTIs [6–10]. Whereas long-term exposures to different air pollutants, accounted through averaging pollutant measurements of past years and looking across large spatial scales, also found similar relationships between URTIs and ambient air pollutants [11]. Apart from linear relationships, previous research has also found non-linear effects of PM on URTI relative risk [12, 13].

Environmental variables, such as ambient air pollutant and climate measurements, can be collected almost instantaneously, but disease case counts are collated on daily or lower frequencies due to inherent limitations of disease surveillance systems. Exposures to environmental variables are also continuously occurring in nature, while the exposure and generation time to disease is disjointed. The differences in sampling frequencies and the relationships between exposure and response variables therefore complicates the analysis of environmental variables with that of disease case counts.

As evidenced from previous work looking at disease cases and ambient air pollution, a common way to account for variables of different frequencies comprise of averaging the values of higher frequency variables to that of a lower frequency [11]. In the case of time series analysis, to further account for the different ways each lagged covariate influence disease case counts, multiple lags of the covariate of interest may be included into a regression specification [11]. Crucially, the former leads to the loss of valuable information within the higher frequency variable, and the latter may lead to overfitting and errors in estimation due to the inclusion of many correlated parameters. It is thus important to bridge this discrepancy in data sampling frequencies in some optimal way.

To circumvent overfitting and model over-specification, mixed data sampling (MIDAS) models deal with resolving data collected at different frequencies by taking the weighting scheme of the higher frequency variable as unknown and parameterized by a function to be estimated. It also has several attractive properties, which include (1) model parsimony – requiring a small number of parameters to incorporate a large number of high frequency lagged observations and (2) flexibility – using a flexible, data driven weighing scheme for high frequency variables which reduces the possibility for omitted variable bias and model misspecification.

When frequencies are resolved among variables, a central related problem is inferring the exposure–response functions or curves. In literature, this translates to (1) choosing the number of lags in time-series models and (2) the number of terms to place into splines or kernels within non-linear regression models [14] through diagnostic tests [15], model selection criteria [16] and prior epidemiological information. Centrally, these measures tradeoff between a modeller's prior on disease and being agnostic about assumptions on disease transmission dynamics – due to them comparing the discrepancy between the estimated model and observed data. MIDAS incorporates the full range of observed exposures as explanatory variables into the regression specification and allows the estimation procedure to decide appropriately on the temporal importance of exposures, thereby circumventing either problem.

In this paper, using disease surveillance and ambient air pollutant data collected for 2 decades, we employ MIDAS to understand the short to medium exposure–response between URTIs, pneumonia and major ambient air pollutants. We compared the ability of MIDAS to fit to disease case data versus standard time series models, map the subsequent exposure–response curves between URTIs, pneumonia and major ambient air pollutants across all provinces. By exploiting the large contiguous scale on which the data is collected, we also examined how these exposure-responses are heterogenous.

Methods

Disease surveillance data

Disease surveillance data was obtained from Thailand's Ministry of Public Health (MOPH) disease surveillance system [17] where we obtained reported disease case counts from 2000 to 2022 at the province level. Suspected cases of pneumonia and influenza in provincial public health offices, hospitals and all health stations were reported through the disease surveillance system to the Bureau of Epidemiology, MOPH. Suspected cases were defined as patients who had symptoms matching the clinical criteria. Private hospitals were not covered by the surveillance system. Diagnoses were recorded using

10th revision of the International Classification of Disease (ICD) codes. As Bueng Khan was split from Nong Khai in 2011 to form a separate province, we have merged disease case counts from Bueng Khan back to Nong Khai from 2011 onwards, to allow consistent analysis over the time-frame of the dataset.

Demographic data

Data on annual population size for each province from 2000 to 2022 was obtained from the Official Statistics Registration Systems of Thailand [18]. Similarly, as Bueng Khan was split from Nong Khai in 2011 to form a separate province, we have merged the population numbers from Bueng Khan back to Nong Khai from 2011 onwards, to allow consistent analysis over the timeframe of the dataset.

Climate data

Climate data was obtained from ERA5, published by the European Centre for Medium-Range Weather Forecasts [19]. Each data point covers a 30 km grid,

$$y_{\tau+1} = \alpha + \sum_{j=0}^{p_y-1} p_{j+1} y_{\tau-j} + \sum_{j=0}^{p_z-1} \sum_{s=1}^{S} \gamma'_{j+1,s} z_{\tau-j,s} + \sum_{v=1}^{V} \beta_v B\left(L^{\frac{1}{m}}; \theta_v\right) x_{\tau,v}^{(m)} + \epsilon$$

which we spatially averaged across each province. Mean, median and maximum of total precipitation, vegetation index, air temperature at 2 m and dew point temperature at 2 m was collected. Relative humidity and average humidity were calculated using standard formula.

Ambient air pollutants data

Ambient air pollutant data was obtained from NASA's Goddard Earth Sciences Data and Information Services Center, GES DISC. The 1-Hourly CO Column Burden, CO Surface Concentration, and 1-Hourly Aerosol diagnostics were obtained from Global Modeling And Assimilation Office [20] to derive $PM_{2.5}$ surface concentration. The surface model layer of the 3d 3-Hourly Aerosol Mixing Ratio was obtained from Global Modelling and Assimilation Office [21] to derive PM_1 and PM_{10} surface concentration. These datasets are a part of Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA 2), which is a reanalysis of the modern satellite data, and is produced by NASA's Global Modelling and Assimilation Office (GMAO) [22].

Mixed data sampling (MIDAS)

Consider reported disease case counts $y_{\tau+1}$ which is observed at the discrete time point between τ and $\tau+1$ and supposed that we have additional information arising from a set of V predictors $\boldsymbol{x}_\tau^{(m)} = (\boldsymbol{x}_1^{(m)}, \boldsymbol{x}_\tau^{(m)}, \ldots, \boldsymbol{x}_V^{(m)})$ which are observed m times between τ and $\tau+1$. The variables $y_{\tau+1}$ and $\boldsymbol{x}_\tau^{(m)}$ can thus be said to be observed at different frequencies. The traditional and simplest manner of dealing with mixed frequency data would be averaging the high frequency predictors to the same frequency of the dependent variable $y_{\tau+1}$. However, this approach may result in omitted variable bias and model mis-specification if the true weighting scheme is not a simple average [11]. The mixed frequency data sampling (MIDAS) framework provides an alternative estimator that can account for different weighting schemes in the high frequency predictors. Specifically, the approach plugs in the high frequency lagged terms of predictors $\boldsymbol{x}_\tau^{(m)}$ in a regression for the low frequency dependent variable $y_{\tau+1}$ as follows:

where $y_{\tau-j}$ denotes past disease case counts for a maximum of p_y-1 lags. We can additionally place S exogenous variables $z_{\tau-j,s}$ recorded at time $\tau-j$ with the same frequency as y for a maximum of p_z-1 lags. The high frequency predictors $x_{\tau,v}^{(m)}$ are shrunk to the same frequency as y and z by a polynomial term B to be defined later. α is an intercept term, p_y autoregressive terms, $\gamma'_{j+1,s}$ coefficients denoting the effect of past z_s on y and β_v are coefficients which capture the overall effect of $B\left(L^{\frac{1}{m}}; \theta_v\right)$ on $y_{\tau+1}$. We further assume i.i.d and normally distributed errors, with mean zero and finite variance σ_ϵ^2:

$$\epsilon \sim N(0, \sigma_\epsilon^2)$$

The polynomial term is given by:

$$B\left(L^{\frac{1}{m}}; \theta_v\right) = \sum_{k=0}^{K-1} B(k, \theta_k) L^{k/m}$$

where $L^{k/m}$ is a lag operator such that $L^{1/m} = x_{\tau-1/m}^{(m)} . \boldsymbol{\theta}_v$ are parameters which provide the shape of the polynomial term. We use the normalized beta probability density function on the polynomial term $B(k, \boldsymbol{\theta}_v)$:

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 4 of 13

$$B(k, \boldsymbol{\theta}_v) = \frac{x_k^{\gamma_1 - 1}(1 - x_k)^{\gamma_2 - 1}}{\sum_{k=1}^{K-1} x_k^{\gamma_1 - 1}(1 - x_k)^{\gamma_2 - 1}}$$

where $x_k = (k - 1)/(K - 2)$. The beta polynomial was used as it only requires two parameters $\boldsymbol{\theta}_v \in \{\gamma_1, \gamma_2\}$ to specify and generate a large variety of weighting shapes [23]. In particular, further restrictions such as $\{\gamma_1 = 1, \gamma_2\}$ or $\{1 + \gamma_1, 1 + \gamma_1 + \gamma_2\}$ allow the weighting structure to form only downward sloping and hump-shaped weights respectively, which were also explored.

Implementing the MIDAS specification

In order to implement MIDAS, we consider the Bayesian paradigm and estimate parameters based on Markov chain Monte Carlo (MCMC) methods, which alleviates certain issues in frequentist parameter estimation. First, uncertainty and standard errors can be characterized easily using credible intervals. Second, the MCMC approach facilitates easy modification to allow shrinkage and sparsity, thereby allowing unimportant variables to be shrunk closer to zero.

Briefly, we first placed diffuse, conjugate priors on both the parameters of the observation equation $\{p_{j+1}, \gamma'_{j+1,s}, \beta_v\}$ and error variance σ_ϵ^2. Gibbs sampling was used as conditional posterior distributions for these parameters were well defined. We also modified the baseline procedure of sampling parameters in the observation equation to induce shrinkage, through incorporating the structure of Bayesian Lasso [24]. This comprised changing the prior for the variance of the error term, to enable more probability weight to be placed on 0 for the parameters of the observation equation.

Following Ghysels [23], a Gamma distribution is placed as a prior on the parameters of the polynomial terms $\{\gamma_1, \gamma_2\}$ with both shape and scale parameters being 1, which amounts to a flat weighting scheme that puts equal weight on the high frequency data. The polynomial terms were sampled using a random walk Metropolis-in-Gibbs step. The Metropolis step is an accept-reject step which takes a candidate draw from some proposal distribution, with acceptance for that draw is given by a probability that depends on the likelihood, parameter's prior distribution and the proposal density. In this case, we use the Gamma proposal distribution as a proposal density, as it corresponds to the functional form of the MIDAS weighting polynomial. Full details for the estimation strategy are outlined in the supplementary information.

Assessing model performance

Convergence of MCMC chains was first assessed by visual inspection of trace plots and Gewecke convergence diagnostic checks. Residual autocorrelation is computed for up to 20 week lags to ensure that the transmission dynamics are properly accounted for in each specification, across models. Quantile–quantile plots are used to see whether the specifications adequately account for the data structure. We additionally computed 4 other statistics on in-sample model fit, namely, the R^2 and in-sample mean-squared error to look at whether overall variations in data are captured, the adjusted R^2 to look at whether overall variations in data are captured while penalizing larger models. Lastly, the deviance information criterion [25] look at whether overall variations in data are captured while penalizing larger models with larger uncertainty in posterior distributions.

Results

We first merged at the province level, with either monthly URTI or pneumonia case counts, climate confounders such as mean temperature total precipitation, absolute and relative humidity for the past 2 months, as well as ambient air pollutants for the past 40 days, recorded at the 6 hourly level (Table 1). MIDAS models were estimated with each ambient air pollutant taken separately in respective polynomial terms, while controlling for past climate measurements and past disease case counts for up to two months. We estimated (1) a baseline autoregressive model consisting of no MIDAS terms and only disease case counts and climate measurements for the past 2 months (2) MIDAS models with no variable shrinkage (3) MIDAS models with variable shrinkage. For (2), (3), we also explored separately the utility of placing no restrictions on MIDAS weights, downward sloping restrictions on MIDAS weights and hump-shaped restrictions on MIDAS weights. These help us understand whether constraints or the lack thereof in MIDAS weights would provide superior model performance compared to alternatives.

Impact of major ambient air pollutants on URTI and pneumonia

After adjustment for confounders, SO_2, $PM_{2.5}$, and CO over a 40-day period were estimated to be positively

Table 1 Summary statistics for dependent variables of interest and covariates used for MIDAS specification

| Variable (unit)	Mean	Range
Influenza Case Counts (Monthly)	102.13	(0, 13,214)
Pneumonia Case Counts (Monthly)	203.88	(0, 3144)
Absolute Humidity	26.78	(15.7, 40.98)
Relative Humidity (%)	76.36	(45.87, 91.68)
Total Precipitation (mm)	0.18	(0, 1.24)
Temperature (K)	300.08	(290.79, 307.65)
CO Surface Concentration (ppb)	172.19	(50.56, 1090.3)
SO2 Surface Concentration (mu/g)	5.07	(0.05, 59.52)
PM2.5 Surface Concentration (mu/g)	19.05	(1.51, 367.24)
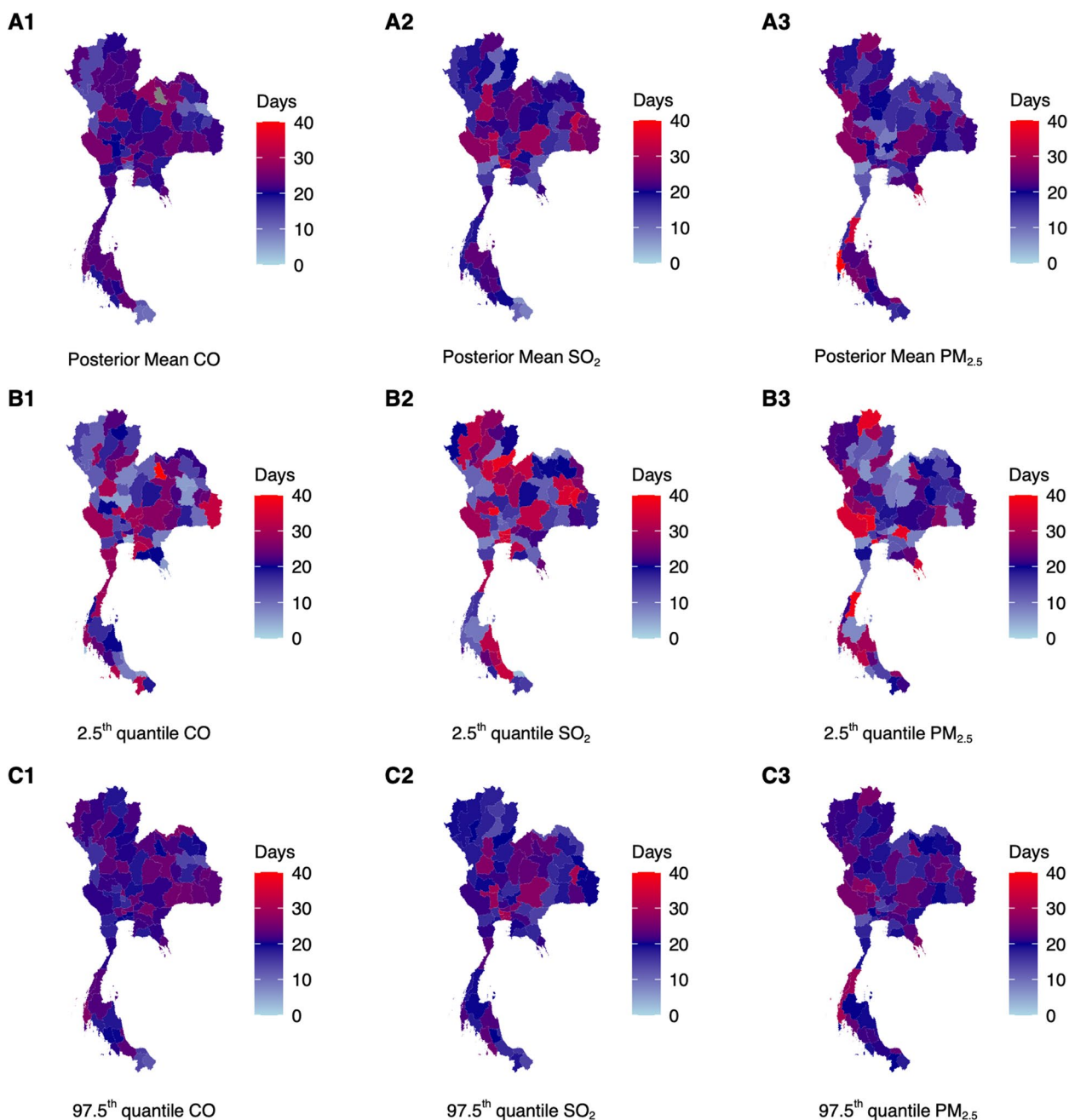
Fig. 1 **A1** – **A3** Posterior mean estimates of duration where the highest MIDAS weights were placed on the importance of respective ambient air pollutant measurements for each province on influencing URTI disease case counts the following month. Darker red and blue shaded regions represent that ambient air pollutant measurements beyond 20 – 40 days are deemed more important in determining URTI disease case counts. **B1** – **B3** 2.5th quantile value for MIDAS weights drawn from MCMC samples **C1** – **C3** 97.5th quantile value for MIDAS weights drawn from MCMC samples

associated to monthly URTI case counts in 41, 15 and 31 of 76 provinces, with their 95% credible intervals away from zero. In other provinces, ambient air pollutants were negatively associated to monthly URTI case counts. On average, greatest influence of ambient air pollutants were estimated to be around the 20th day mark across provinces (Fig. 1A1-3), but the range of values differed greatly, where the 7 – 36, 6 – 41 and 7 – 40 day measurements

for SO_2, CO and $PM_{2.5}$ (Fig. 1) were ascribed the greatest influence on case counts across provinces respectively. Exposure–response shapes were also highly heterogenous across provinces but were mainly hump-shaped across the 40-day period (See supplementary material).

Whereas $PM_{2.5}$, SO_2 and CO over a 40-day period were positively associated to monthly pneumonia case counts in 15, 31 and 49 of 76 provinces (Fig. 2A1-3, B1-3, C1-3). Similar to URTI, we found on average that the greatest influence of ambient air pollutants were
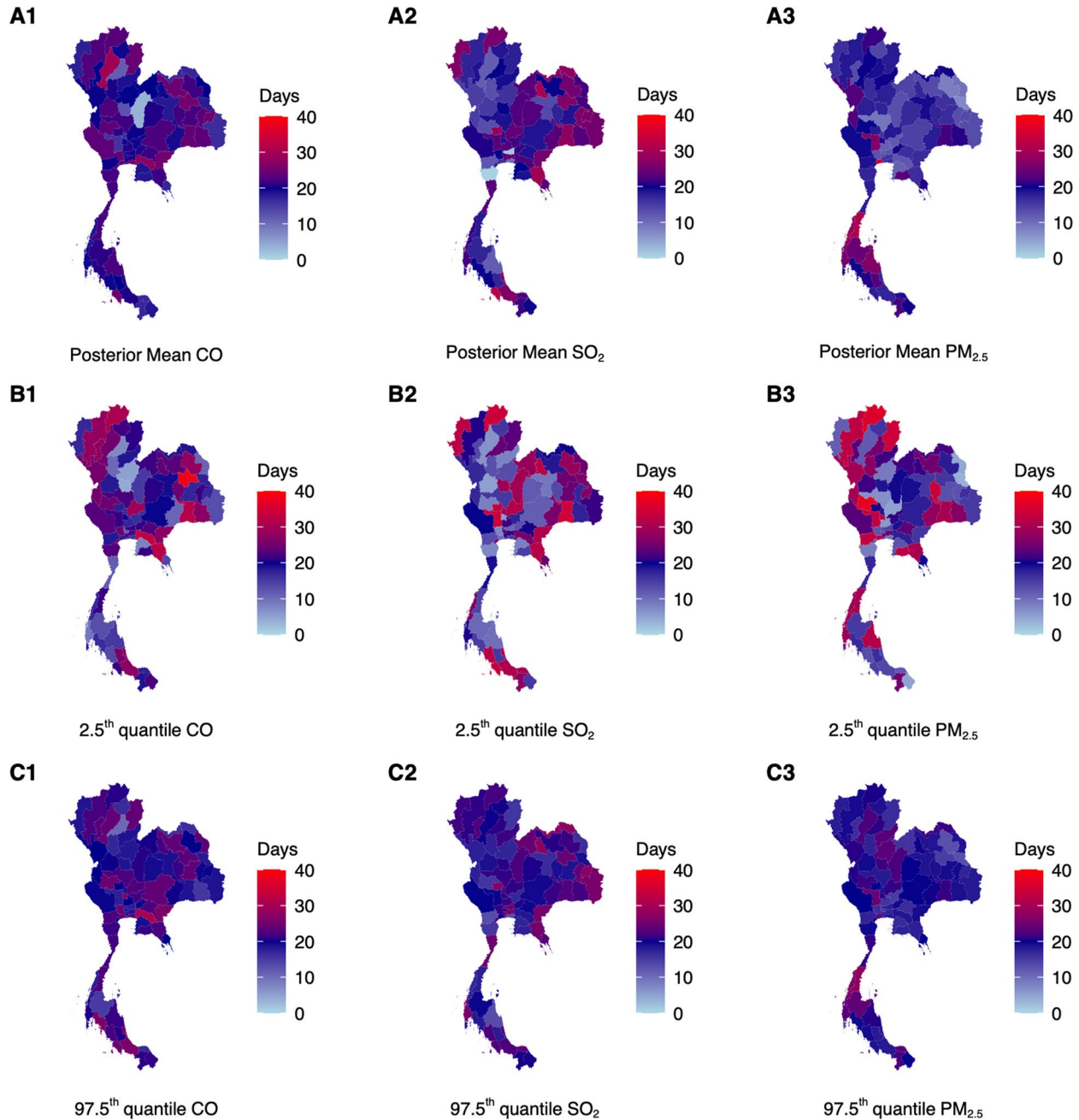
Fig. 2 A1 – A3 Posterior mean estimates of duration where the highest MIDAS weights were placed on the importance of respective ambient air pollutant measurements for each province on influencing pneumonia disease case counts the following month. Darker red and blue shaded regions represent that ambient air pollutant measurements beyond 20 – 40 days are deemed more important in determining pneumonia disease case counts. **B1 – B3** 2.5th quantile value for MIDAS weights drawn from MCMC samples **C1 – C3** 97.5th quantile value for MIDAS weights drawn from MCMC samples

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 7 of 13

estimated to be in the 20th day mark across provinces, but the range was noticeably narrower compared to case of URTIs, at the 0 – 32, 3 – 30 and 7 – 33 day mark for SO$_2$, CO and PM$_{2.5}$ respectively. Similarly, exposure–response shapes were highly heterogenous across provinces for ambient air pollutants and pneumonia

case counts. These shapes were however mainly hump-shaped across the 40-day period (See supplementary material).

Controlling for all other variables, on average, a 1 µg/m^3 rise in SO$_2$ over the past 40 days would lead to an expected 1624.1 (Fig. 3, A2, B2, C2) increase in URTI
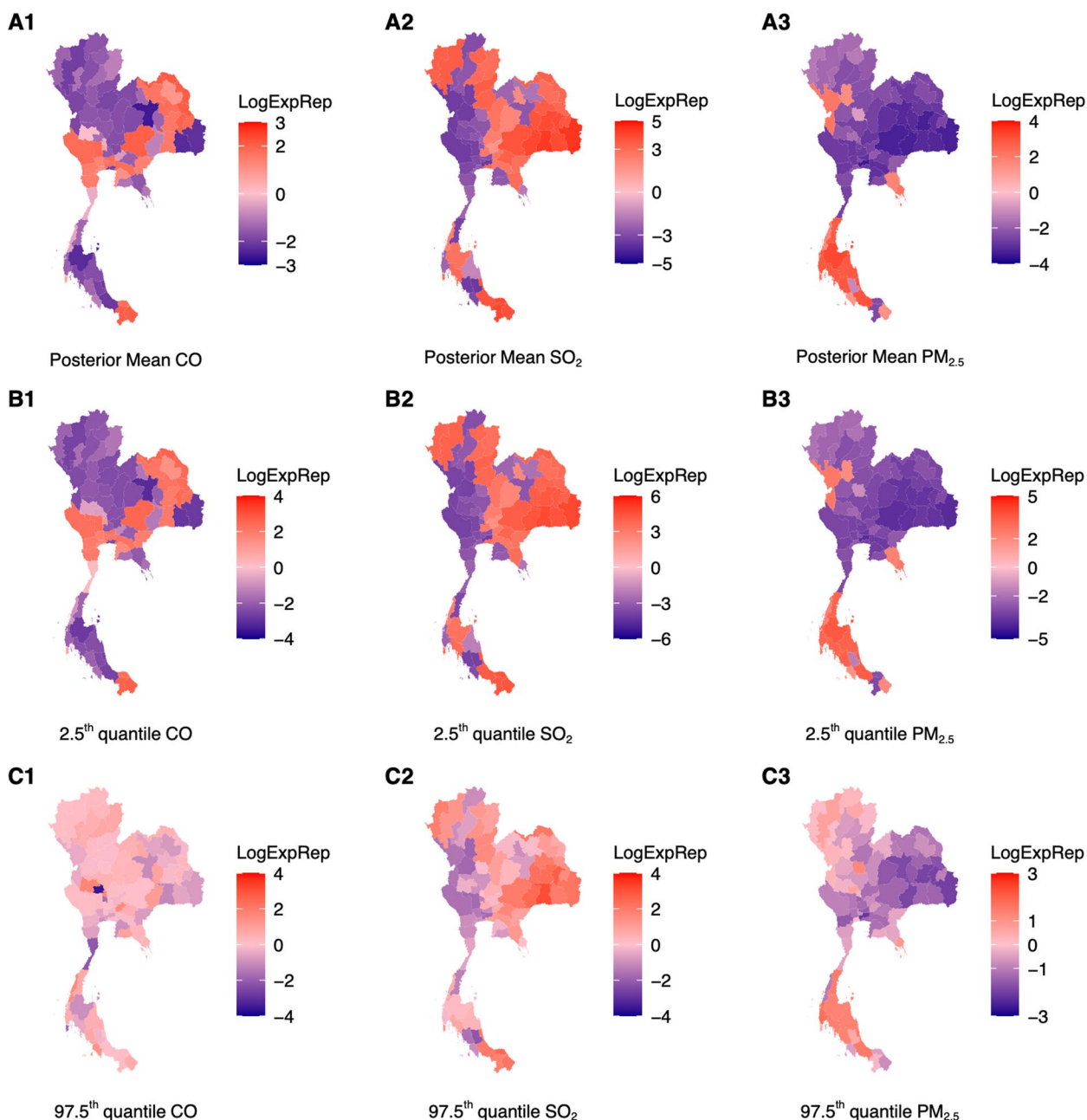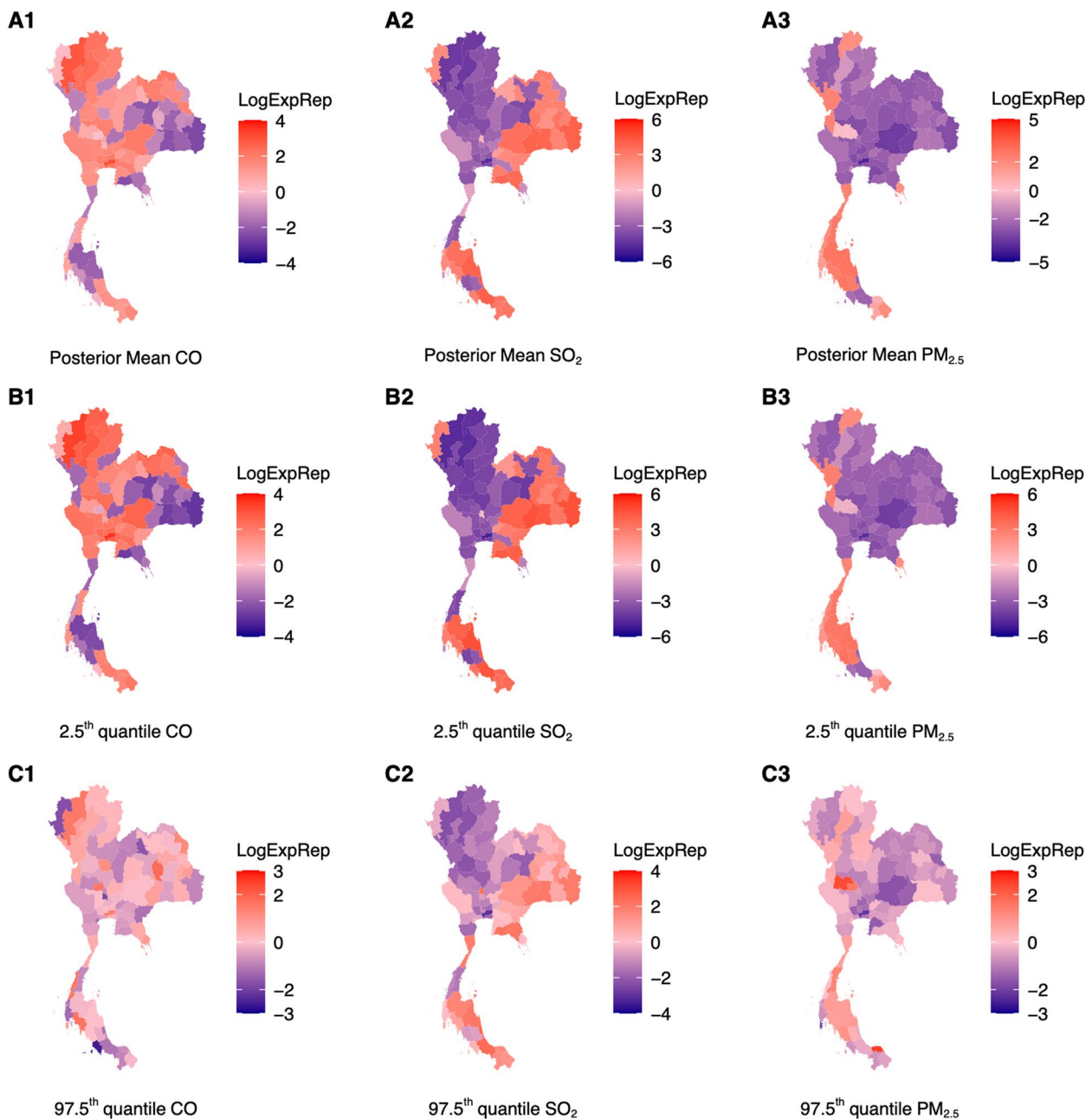

Fig. 3 **A1** – **A3** Posterior mean estimates of mean exposure–response of respective ambient air pollutants over the past 40 day period on contemporaneous URTI disease case counts for a specific region. Red shades represent regions where a one unit increase in ambient air pollutant surface concentrations in the respective region for the past 40 days are associated to an increase in contemporaneous, monthly URTI disease case counts. **B1** – **B3** 2.5th quantile value for exposure response drawn from MCMC samples of MIDAS weights **C1** – **C3** 97.5th quantile value for exposure response drawn from MCMC samples of MIDAS weights

Fig. 4 **A1** – **A3** Posterior mean estimates of mean exposure–response of respective ambient air pollutants over the past 40 day period on contemporaneous pneumonia case counts for a specific region. Red shades represent regions where a one unit increase in ambient air pollutant surface concentrations in the respective region for the past 40 days are associated to an increase in contemporaneous, monthly pneumonia case counts. **B1** – **B3** 2.5th quantile value for exposure response drawn from MCMC samples **C1** – **C3** 97.5th quantile value for exposure response drawn from MCMC samples

case counts. However, an expected change of -15.1 (Fig. 3, A1, B1, C1) and -406.3 (Fig. 3, A3, B3, C3) in URTI case counts respectively are expected to occur with a 1 ppb rise in CO and 1 μg/m³ rise in $PM_{2.5}$ surface concentrations over the past 40 days respectively across provinces.

Whereas a 1 ppb rise in CO surface concentrations over 40 days would lead to an expected 39.3 change (Fig. 4, A1, B1, C1) in monthly pneumonia case counts, a 1 μg/m³ increase in SO_2 and $PM_{2.5}$ surface concentration would lead to an expected -943 (Fig. 4A2, B2, C2) and -401.8 (Fig. 4) change in monthly pneumonia

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 9 of 13

case counts across provinces respectively. Comparing the range of posterior mean effect sizes and quantile values across provinces for major ambient air pollutants on monthly pneumonia and URTI case counts showed that associations between ambient air pollutants on pneumonia were far more variable compared to that of URTI (Figs. 3 and 4).

Contributive effects and burden of climate, past disease and ambient air pollutants on URTI and pneumonia

Across provinces, controlling for past disease case counts and the effects of ambient air pollutants, increases in climate measurements the past two months have mixed associations to monthly URTI and pneumonia case counts. In particular, an increase in mean absolute humidity over the past 1 and 2 month was associated to decreases in contemporaneous URTI case counts in 22 and 9 of 76 provinces respectively. Whereas an increase in mean relative humidity 1 and 2 months prior was associated to decreases in contemporaneous URTI case counts in only 2 and 1 provinces respectively. However, in 8 provinces, increases in mean relative humidity 1 month prior was associated to increases in contemporaneous URTI case counts. In all provinces, higher temperature 1 and 2 months prior were associated to decreases in URTI in all provinces. Lastly, increases in total precipitation the past month were associated to increases and decreases in contemporaneous conjunctivitis case counts in 21 and 10 provinces respectively.

Similarly, increases in temperature for the past 1 and 2 months were associated to decreases in pneumonia case counts in 74 and 73 provinces respectively. Increases in absolute humidity the past 1 and 2 months were associated to decreases in contemporaneous pneumonia case counts in 16 and 11 provinces respectively. Whereas in only 3 provinces, changes in relative humidity were associated to any increase or decrease in contemporaneous pneumonia case counts.

Comparing the contributive effect of past disease case counts, ambient air pollutants and climate on disease case counts suggests that the mean impact of past climate variables across provinces on contemporaneous URTI case counts is minor, ranging from -0.7 – 0.17% (Fig. 5A2), while the mean impact of past climate variables across provinces on pneumonia case counts is higher at -27.8% – 15.8% (Fig. 5B2). Whereas the impact of ambient air pollutants were greater in influencing contemporaneous URTI and pneumonia case counts at -14.8% – 4.96% (Fig. 5A1, B1) and -79.78% – 107.5% (Fig. 5A3, B3) respectively. Lastly, the impact

of past case counts in influencing contemporaneous URTI and pneumonia case counts at -17.4% – -0.16% (Fig. 5A1, B1) and -118.4% – 17.9% (Fig. 5A3, B3) respectively. For the case of pneumonia, wide estimates for contributive effect were partially attributed to the timepoints where case counts were low in some provinces (See supplementary material) and model predictions were pulled in separate directions by case count and ambient air pollutant measurements.

Effects of normalizing disease case counts by annual population

All analysis was repeated using case counts normalized by population size. That is, disease case counts per 100,000 individuals was used as the dependent variable of interest for the analysis. Similar to the analysis using raw disease case counts as the dependent variable of interest, we found that CO and SO_2 surface concentrations the past 40 days were more likely to be positive associated to increases in disease case counts for the contemporaneous month. Mean duration where the pollutants were taken to be most important by MIDAS weights were also around the 18 – 21th day mark. The contributive burdens of different confounders on disease case counts were also similarly found to be higher on past disease case counts and ambient air pollutant measurements, rather than climate (See supplementary material).

Effect of higher frequency data and variable shrinkage on model performance

Quantile–quantile plots showed that models considered, with or without MIDAS terms, in general could capture distributional characteristics of disease case counts well. Visual diagnostics of trace-plots also showed that MCMC procedures have converged for all parameters in each model.

Overall, we found that incorporating MIDAS weights provided overall improvements over the baseline model where no MIDAS weights were incorporated, with overall lower model mean absolute errors, deviance information criterion and coefficients of determination. Furthermore, the model incorporating no restrictions on MIDAS weights worked better compared to models incorporating downward sloping or hump-shaped weights. These models had better model fit as demonstrated by the mean-squared error as well as the deviance information criterion in 40.8% and 37.5% of provinces across all models considered. Imposing restrictions for MIDAS weights to be downward sloping weights also performed
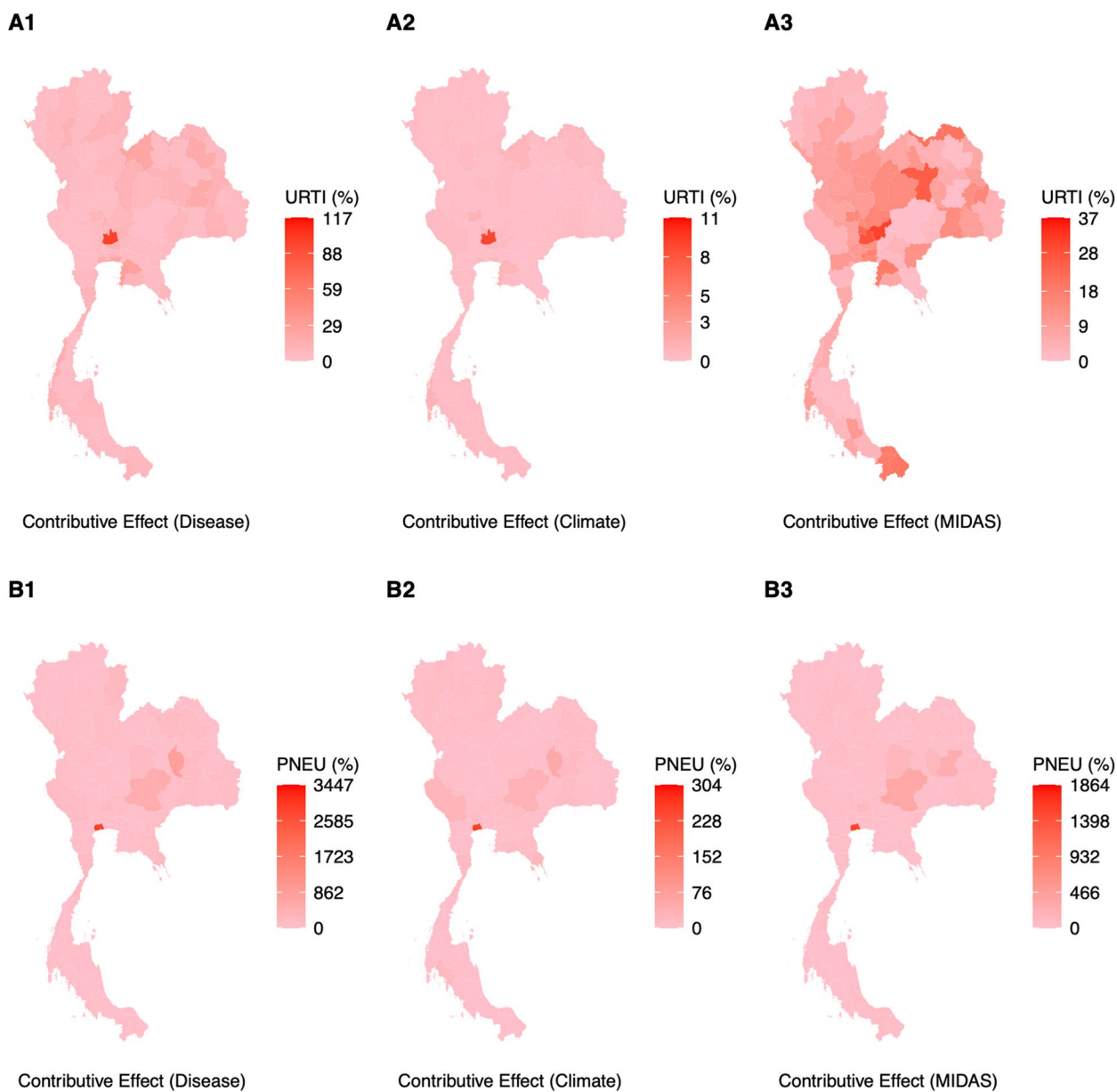
Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 10 of 13

A1

Contributive Effect (Disease)

A2

Contributive Effect (Climate)

A3

Contributive Effect (MIDAS)

B1

Contributive Effect (Disease)

B2

Contributive Effect (Climate)

B3

Contributive Effect (MIDAS)

Fig. 5 **A1** – **A3** Average absolute contribution of disease case counts the past two months, climate measurements the past two months (i.e. temperature, total precipitation, absolute humidity, relative humidity) as well as ambient air pollutants as weighted under MIDAS over the past 40 day period on predicted contemporaneous URTI case counts **B1** – **B3** Average absolute contribution of disease case counts the past two months, climate measurements the past two months (i.e. temperature, total precipitation, absolute humidity, relative humidity) as well as ambient air pollutants as weighted under MIDAS over the past 40 day period on predicted contemporaneous pneumonia case counts

reasonably well, as demonstrated by their comparatively better mean-squared error as well as the deviance information criterion in 34.9% and 40.8% of provinces across all models considered.

Incorporating variable shrinkage to the overall regression specification also improved model performance in terms of mean-squared error, unadjusted and adjusted coefficients of determination as well as the deviance

information criterion in almost all provinces. These tests are reported in the supplementary information section.

Given these diagnostic tests, the model incorporating shrinkage in regression parameters, as well as no restrictions on MIDAS polynomials was taken as representative to infer exposure–response curves between ambient air pollutants and URTI/pneumonia case counts across all provinces.

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 11 of 13

Discussion

Heightened exposure to ambient air pollutants the past 40 days was shown to have mixed associations with contemporaneous monthly URTI and pneumonia case counts across provinces (Figs. 3 and 4). Specifically, heightened SO_2 and CO concentrations the past 40 days were shown to have a positive association with URTI and pneumonia case counts in around half the regions, with or without controlling for population size, and controlling for past disease case counts and climate data (Figs. 3 and 4). These associations also had credible intervals which crossed the zero bound.

SO_2 and $PM_{2.5}$ have been shown to have negative effects on the respiratory system and exacerbate breathing problem [26, 27]. In mouse models, SO_2 exposure was observed to cause DNA damage and oxidative damage in the lungs [28, 29]. $PM_{2.5}$ also leads to inflammation and cell damage, through the production of free radicals and reactive oxygen species [30]. The potential excess inflammation and respiratory system damage from PM could increase one's susceptibility and severity to pneumonia and URTI [31, 32]. However, we found that the association of past SO_2 and $PM_{2.5}$ exposure on disease case counts were mixed across different provinces. Although SO_2 is mostly known as an air pollutant, several studies suggested that SO2 had beneficial impacts on mammals, including anti-inflammatory effects [33–35]. Low levels of SO_2 was suggested to have a protective effect on bacteria-induced pulmonary infections [36]. Furthermore, a study in mice found that low doses of SO_2 exposure reduced the amount of pneumonia [37].

CO has been shown to be toxic on a systemic level [38]. In excess, CO causes hypoxia by decreasing the oxygen carrying capacity and O_2 release in tissues [39]. Excessive CO can also lead to systemic immunological or inflammatory damage [39, 40]. While there is no clear direct biological mechanism between CO exposure and pneumonia, CO is a product of incomplete combustion [41]. Hence, CO might be a marker for other harmful inflammatory combustion products such as soot, which potentially explains the positive association found between CO and pneumonia [42]. However, an inverse association was also found in some provinces. While CO is toxic at high levels, low levels of CO have been found to possess anti-inflammatory effects and have been suggested for therapeutic use [38]. Low-dose CO could reduce lung inflammation and possibly decrease one's susceptibility to pneumonia, resulting in the mixed associations observed [43].

In contrast, there is a stronger consensus in literature that $PM_{2.5}$ has a positive relationship with URTI and pneumonia [44, 45]. Exposure to heightened $PM_{2.5}$ can cause the development and progression of acute and chronic lung diseases, such as tracheal and pulmonary inflammation [46, 47]. In outpatient, emergency, and hospitalization-related data on respiratory infections showed that $PM_{2.5}$ exposure was positively associated with the increased respiratory infections. Other studies showed that $PM_{2.5}$ exposure was positively correlated with outpatient visits for upper respiratory tract infection [44]. While in certain provinces, especially in highly urbanized central Thailand (Fig. 3A1, 4A1) our analysis demonstrated a positive exposure–response between heightened $PM_{2.5}$ concentrations and URTI/Pneumonia case counts, in other provinces we estimated converse associations between $PM_{2.5}$ concentrations and URTI/Pneumonia case counts. To discern the reason for the difference in direction and effect sizes for associations between SO_2 and $PM_{2.5}$ and URTI/pneumonia case counts, the surface concentrations of SO_2 and $PM_{2.5}$ across the time period over the provinces could be compared to the National Ambient Air Quality Standards established by the United States Environmental Protection Agency (EPA) [48]. SO_2 and $PM_{2.5}$ levels that are consistently lower than the EPA requirements were likely to have less or varying effects on URTI and pneumonia case counts.

Lastly, past disease case counts were found to have a significant contributive effect on disease case count. This is unsurprising as case counts tend to co-move. Additionally, we found that the effects of weather had little contribution on URTI and pneumonia case counts. This is perhaps due to the lack of large variations in weather in our study setting, which has a tropical climate. This is in contrast to climates with stark variations in weather, where colder months would lead to physiological and behavioural changes, such as spending more time indoors in poorly ventilated spaces, which would increase respiratory disease transmission risk [49, 50]. This changing behaviour and the resultant changing transmission risk would not be observed in tropical climates, explaining the lack of contribution of weather variables to URTI and pneumonia cases.

Our proposed MIDAS methodology has several strengths, it enables us to incorporate higher frequency ambient air pollutant measurements to infer their respective associations with future disease burden. This is conducted by first defining a flexible specification where the importance of each ambient air pollutant's lagged high frequency observations are decided by data, rather than taking a simple average to aggregate these variables to the same frequency as disease case counts. This prevents subjective lag selection in a model specification to detect associations [51–53] as all past pollutant measurements can be placed in a model, with their importance post-hoc decided. Furthermore, having a flexible weighting scheme

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 12 of 13

reduces discretization bias due to simple averaging and increases the statistical efficiency of the model. Empirically, this is demonstrated from overall better model fit to observed data using the MIDAS framework across all provinces (See supplementary material).

Furthermore, while we used data over a large spatial scale, collected over a long period to delineate to historical impact of air pollutant concentrations on overall ambient air pollutant burden, the air pollutant concentrations in the study could not be representative of individual exposures, and the ecological design may cause ecological fallacy. Subgroup analyses by age and sex were not undertaken and should be considered in future studies. Also, while a primary strength of this study is the delineation of association on the provincial rather than national scale, conducting the analysis with national-level case counts could produce a generalized insight into the nationwide association of pneumonia and URTI with air pollutant exposure. Results could however be biased by spatial confounding. Unmeasured variables may be potentially confounded with the exposures of interest, such as income levels. Future directions could also comprise using more structured penalties, such as the group or fused LASSO [54, 55] to shrink MIDAS terms in a more informative manner, as well as consider MIDAS as a means to forecast instead of explain disease burden. National level analyses could also be undertaken, when additional spatial covariates are measured to adjust for unobserved confounding.

Conclusion
In summary, we developed a novel statistical methodology to delineate the impacts of air pollutant concentrations on disease burden. Over a large spatial region, we found mixed but generally positive associations between increased O_3, SO_2, and $PM_{2.5}$ concentrations on URTI and pneumonia case counts.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12879-023-08185-0.

> **Additional file 1.**

Authors' contributions
Esther LW Choo: Conceptualization, Original draft preparation, Writing-Reviewing and Editing, A Janhavi: Original draft preparation, Writing- Reviewing and Editing Joel Ruihan Koo: Visualization, Methodology, Original draft preparation, Writing- Reviewing and Editing, Steve HL Yim: Writing- Reviewing and Editing, Borame Dickens: Writing- Reviewing and Editing, Jue Tao Lim: Data curation, Writing- Original draft preparation, Investigation, Writing- Reviewing and Editing, Conceptualization, Methodology. The author(s) read and approved the final manuscript.

Availability of data and materials
All disease surveillance data used to reproduce the results can be obtained at http://doe.moph.go.th/surdata/index.php.

Declarations

Ethics approval and consent to participate
All analysis was conducted on data from publicly available surveillance databases. No ethics approval is required to conduct this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. [2]Department of Biological Sciences, National University of Singapore, Singapore, Singapore. [3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [4]Asian School of the Environment, Nanyang Technological University, Singapore, Singapore. [5]Earth Observatory of Singapore, Nanyang Technological University, Singapore, Singapore.

References
1. Thomas M, Bomar PA. Upper Respiratory Tract Infection. in *StatPearls* (StatPearls Publishing, 2022).
2. Jin X, et al. Global burden of upper respiratory infections in 204 countries and territories, from 1990 to 2019. eClinicalMedicine. 2021;37:100986.
3. Fendrick AM, Monto AS, Nightengale B, Sarnes M. The economic burden of non–influenza-related viral respiratory tract infection in the United States. Arch Intern Med. 2003;163:487–94.
4. West JV. Acute upper airway infections. Br Med Bull. 2002;61:215–30.
5. World Health Organization. Types of pollutants. https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants.
6. Chauhan AJ, Krishna MT, Frew AJ, Holgate ST. Exposure to nitrogen dioxide (NO2) and respiratory disease risk. Rev Environ Health. 1998;13:73–90.
7. Chen T-M, Kuschner WG, Gokhale J, Shofer S. Outdoor air pollution: nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects. Am J Med Sci. 2007;333:249–56.
8. Glencross DA, Ho T-R, Camiña N, Hawrylowicz CM, Pfeffer PE. Air pollution and its effects on the immune system. Free Radic Biol Med. 2020;151:56–68.
9. Wiegman CH, Li F, Ryffel B, Togbe D, Chung KF. Oxidative stress in ozone-induced chronic lung inflammation and emphysema: a facet of chronic obstructive pulmonary disease. Front Immunol. 2020;11:1957.
10. Wu Y-F, et al. Inactivation of MTOR promotes autophagy-mediated epithelial injury in particulate matter-induced airway inflammation. Autophagy. 2019;16:435–50.
11. Ghysels E, Sinko A, Valkanov R. MIDAS regressions: further results and new directions. Econom Rev. 2007;26:53–90.
12. Yu H-L, Chien L-C. Short-term population-based non-linear concentration–response associations between fine particulate matter and respiratory diseases in Taipei (Taiwan): a spatiotemporal analysis. J Expo Sci Environ Epidemiol. 2016;26:197–206.

Choo *et al. BMC Infectious Diseases*   (2023) 23:379

Page 13 of 13

13. Zhang Z, et al. Association of particulate matter air pollution and hospital visits for respiratory diseases: a time-series study from China. Environ Sci Pollut Res. 2019;26:12280–7.

14. Hastie TJ. Statistical Models in S. Routledge; 2017.

15. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. (Springer, 2002).

16. Kadane JB, Lazar NA. Methods and criteria for model selection. J Am Stat Assoc. 2004;99:279–90.

17. Thailand Ministry of Public Health Thailand. MOPH TH. Thailand Ministry of Public Health. 2022. https://www.moph.go.th/.

18. Thailand Department of Provincial Administration. Thailand Official Statistics Registration System. 2022. Thailand Official Statistics Registration System https://stat.bora.dopa.go.th/stat/statnew/statMenu/newStat/home.php.

19. Copernicus Climate Change Service. ERA5. *ERA5* https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. 2022.

20. Global Modeling And Assimilation Office. MERRA-2 tavg1_2d_aer_Nx: 2d,1-Hourly,Time-averaged,Single-Level,Assimilation,Aerosol Diagnostics V5.12.4. 2015. https://doi.org/10.5067/KLICLTZ8EM9D.

21. Global Modeling And Assimilation Office. MERRA-2 inst3_3d_aer_Nv: 3d,3-Hourly,Instantaneous,Model-Level,Assimilation,Aerosol Mixing Ratio V5.12.4. 2015. https://doi.org/10.5067/LTVB4GPCOTK2.

22. Gelaro R, et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J Clim. 2017;30:5419–54.

23. Ghysels E. Macroeconomics and the reality of mixed frequency data. J Econom. 2016;193:294–314.

24. Park T, Casella G. The Bayesian Lasso. J Am Stat Assoc. 2008;103:681–6.

25. Gelman, A. Bayesian data analysis. (CRC Press, 2014).

26. US EPA, O. Particulate Matter (PM) Basics. 2016. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics.

27. US EPA, O. Sulfur Dioxide Basics. 2016. https://www.epa.gov/so2-pollution/sulfur-dioxide-basics.

28. Meng Z, Geng H, Bai J, Yan G. Blood pressure of rats lowered by sulfur dioxide and its derivatives. Inhal Toxicol. 2003;15:951–9.

29. Meng Z, Qin G, Zhang B, Bai J. DNA damaging effects of sulfur dioxide derivatives in cells from various organs of mice. Mutagenesis. 2004;19:465–8.

30. Xing YF, Xu YH, Shi MH, Lian YX. The impact of PM2.5 on the human respiratory system. J Thorac Dis. 2016;8:E69-74.

31. Aguilera ER, Lenz LL. Inflammation as a modulator of host susceptibility to pulmonary influenza, pneumococcal, and co-infections. Front Immunol. 2020;11:105.

32. Mizgerd JP. inflammation and pneumonia: why are some more susceptible than others? Clin Chest Med. 2018;39:669–76.

33. Huang XL, Zhou JL, Zhou XH, Xian XH, Ding CH. Ameliorative effects of exogenous sulfur dioxide on lipopolysaccharide-induced acute lung injury in rats. Sheng Li Xue Bao. 2009;61:499–503.

34. Mitsuhashi H, et al. Increased levels of serum sulfite in patients with acute pneumonia. Shock Augusta Ga. 2004;21:99–102.

35. Wang XB, Du JB, Cui H. Sulfur dioxide, a double-faced molecule in mammals. Life Sci. 2014;98:63–7.

36. Ge E, et al. Ambient sulfur dioxide levels associated with reduced risk of initial outpatient visits for tuberculosis: a population based time series analysis. Environ Pollut. 2017;228:408–15.

37. Lebowitz MD, Fairchild GA. The effects of sulfur dioxide and A2 influenza virus on pneumonia and weight reduction in mice: an analysis of stimulus-response relationships. Chem Biol Interact. 1973;7:317–26.

38. Ryter SW, Ma KC, Choi AMK. Carbon monoxide in lung cell physiology and disease. Am J Physiol-Cell Physiol. 2018;314:C211–27.

39. Hampson NB, Piantadosi CA, Thom SR, Weaver LK. Practice Recommendations in the diagnosis, management, and prevention of carbon monoxide poisoning. Am J Respir Crit Care Med. 2012;186:1095–101.

40. Rose JJ, et al. Carbon monoxide poisoning: pathogenesis, management, and future directions of therapy. Am J Respir Crit Care Med. 2017;195:596–606.

41. Canova C, et al. Carbon monoxide pollution is associated with decreased lung function in asthmatic adults. Eur Respir J. 2010;35:266–72.

42. Russell DW, Blalock JE. The soot of all evil. eLife. 2015;4:e11709.

43. Bathoorn E, et al. Anti-inflammatory effects of inhaled carbon monoxide in patients with COPD: a pilot study. Eur Respir J. 2007;30:1131–7.

44. Yang L, Li C, Tang X. The Impact of PM2.5 on the Host Defense of Respiratory System. Front Cell Dev Biol. 2020;8:91.

45. Ziou M, et al. Outdoor particulate matter exposure and upper respiratory tract infections in children and adolescents: a systematic review and meta-analysis. Environ Res. 2022;210:112969.

46. Jedrychowski WA, et al. Intrauterine exposure to fine particulate matter as a risk factor for increased susceptibility to acute broncho-pulmonary infections in early childhood. Int J Hyg Environ Health. 2013;216:395–401.

47. Kampa M, Castanas E. Human health effects of air pollution. Environ Pollut. 2008;151:362–7.

48. US EPA, O. NAAQS Table. 2014. https://www.epa.gov/criteria-air-pollutants/naaqs-table.

49. Lim JT, Tan KB, Abisheganaden J, Dickens BL. Forecasting upper respiratory tract infection burden using high-dimensional time series data and forecast combinations. PLoS Comput Biol. 2023;19(2):e1010892. https://doi.org/10.1371/journal.pcbi.1010892.

50. Mäkinen TM, et al. Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. Respir Med. 2009;103:456–62.

51. Li YR, et al. Association between air pollution and upper respiratory tract infection in hospital outpatients aged 0–14 years in Hefei, China: a time series study. Public Health. 2018;156:92–100.

52. Liu Y, et al. Association between air pollution and emergency department visits for upper respiratory tract infection in Lanzhou. China Environ Sci Pollut Res. 2022;29:28816–28.

53. Wu Y, et al. Associations of fine particulate matter and constituents with pediatric emergency room visits for respiratory diseases in Shanghai. China Int J Hyg Environ Health. 2021;236:113805.

54. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. J R Stat Soc Ser B Stat. 2005;67:91–108.

55. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B Stat. 2006;68:49–67.

Publisher's Note