

Stochastic modelling of air pollution impacts on respiratory infection risk

Sha He · Sanyi Tang · Yanni Xiao · Robert A. Cheke

This work is supported by the National Natural Science Foundation of China (NSFCs 61772017, 11471201 (ST) and 11631012 (YX)), and by the Fundamental Research Funds for the Central Universities GK201701001.

Sha He, Sanyi Tang
School of Mathematics and Information Science, Shaanxi Normal University, Xi'an, 710119, P.R. China
Tel.: +86-29-85310232
E-mail: heshas@snnu.edu.cn & sytang@snnu.edu.cn

Yanni Xiao
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710048, P.R. China
E-mail: yxiao@mail.xjtu.edu.cn

Robert A. Cheke
Natural Resources Institute, University of Greenwich at Medway, Central Avenue, Chatham Maritime,
Chatham, Kent, ME4 4TB, UK
E-mail: R.A.Cheke@greenwich.ac.uk

Abstract The impact of air pollution on people's health and daily activities in China has recently aroused much attention. By using stochastic differential equations, variation in a six year long time series of air quality index (AQI) data, gathered from air quality monitoring sites in Xi'an from 15 November 2010 to 14 November 2016 was studied. Every year the extent of air pollution shifts from being serious to not so serious due to alterations in heat production systems. The distribution of such changes can be predicted by a Bayesian approach and the Gibbs sampler algorithm. The intervals between changes in a sequence indicate when the air pollution becomes increasingly serious. Also, the inflow rate of pollutants during the main pollution periods each year has an increasing trend. This study used a stochastic SEIS model associated with the AQI to explore the impact of air pollution on respiratory infections. Good fits to both the AQI data and the numbers of influenza-like illness (ILI) cases were obtained by stochastic numerical simulation of the model. Based on the model's dynamics, the AQI time series and the daily number of respiratory infection cases under various government intervention measures and human protection strategies were forecasted. The AQI data in the last 15 months verified that government interventions on vehicles is effective in controlling air pollution, thus providing numerical support for policy formulation to address the haze crisis.

Keywords Air pollution · Stochastic differential equation · Change point · Respiratory infection · Intervention measures

Mathematics Subject Classification (2000)

1 Introduction

As China's economy has developed there has been increasing demand for, and exploitation of, its resources [16, 28, 33]. Consequently, an intractable problem has arisen as more and more industrial fumes are released into the air, seriously damaging the aerial environment. The problem is especially severe in winter, when fossil fuels are harnessed at higher rates than in other seasons to maintain warm indoor temperatures in northern China [5, 47, 48]. For instance, approximately 168 million tonnes of coal were used in 2010 for central heating in China. Taking Beijing as an example, the mass per cubic metre of air of particles with a size (diameter) generally less than 2.5 micrometres (m), known as the PM2.5 air quality index, reached $22.7g/m^3$ as a consequence of burning too much coal in January 2010. In contrast, the index was only $0.7g/m^3$ in July in 2000 [44]. The influence of air pollution is profound not only as it disrupts people's daily lives but also because it causes physical problems [9, 26]. Shrouded in haze, people will discontinue vigorous outdoor activities since exercising in environments with high levels of air pollution can increase the risk of health problems. When cities suffer extremely severe haze, primary schools have to be suspended to protect students health [10, 36].

There is no doubt that air pollution has hindered the improvement of people's living standards, caused a lot of inconvenience and, worst of all, is severely hazardous to people's health [12, 14]. Studies have shown the impact of high concentrations of ambient particles and gaseous pollutants on health and that different kinds of pollutants have different adverse health effects [4, 7, 17, 25]. Both short-term and long-term exposures to pollutants will reduce respiratory functions, leading to increases in hospital respiratory admissions, in medication use by asthmatic subjects, and in deaths, particularly for sensitive groups such as children, the elderly or those with chronic respiratory illnesses [15, 18, 29].

Much research has been conducted to better understand seasonal variations in air pollution and to investigate its effects on human health [3, 6, 13, 34, 40]. By using data from a roadside air quality monitoring station during the 1999-2003 period, Wang et al. [41] analysed the varying trends of the Hong Kong API (the air pollution index in Hong Kong) and the levels of related air pollutants. They demonstrated by statistical analysis that the daily mean API in a seasonal period is a stationary time series, and that the API time series could be predicted in different seasonal specifications by the autoregressive moving average (ARMA) method. Samoli et al. [31] presented a more rigorous and sophisticated model, a Poisson regression in a generalized additive model (GAM), to analyse an epidemiological time series associated with meteorological variables. Yu et al. [46] applied chaos theory to analysed two chaotic indicators, the correlation dimension and the Lyapunov exponent, of the API time series during the past 10 years in Lanzhou. The above studies focused mainly on environmental pollution trends, but some other studies have investigated the relationship between human health and air pollution [1, 11, 22, 45]. To estimate the time-varying effects of pollution on mortality a multi-site time series, Peng et al. [27] developed Bayesian semiparametric hierarchical models and found that different geographic regions present different seasonal characteristics. Chauhan et al. [8] revealed that air pollution and infections act synergistically and both of them could cause respirato-

ry illnesses, especially by exacerbating symptoms in individuals with pre-existing respiratory conditions such as asthma and chronic obstructive pulmonary disease. A generalized additive mixed model with an autoregressive component was used to assess the association of air pollution with the preceding week's pollutant levels and severity of respiratory symptoms and infections measured weekly in healthy infants [35]. Based on a Bayesian approach to meta-analysis, Mehta et al. [24] provided a summary estimates that the risk of acute lower respiratory infections occurrence increased 1.12 (1.03, 1.30) times with an annual average PM_{2.5} concentration increase per 10 g/m³ by studies of longer-term (subchronic and chronic) effects.

The main winter period when heating is needed in Xi'an, the capital city of Shaanxi Province in China, is from 15 November to 15 March of the subsequent year. For the convenience of research, the year here is defined as 15 November to 14 November of the subsequent year instead of using a calendar year. The data on influenza-like illness (ILI) cases and the **air quality index** (AQI) in Fig. 1 (A) and (B) are divided into two parts with the heating period shown in magenta and the non-heating period shown in blue. Both the number of ILI cases and the value of the AQI are generally higher in heating periods than that in non-heating periods, as heating leads to increased emissions of pollutants. Not only that, the ILI cases and AQI data have increasing trends from year to year. There are six levels of air pollution, excellent, good, lightly polluted, moderately polluted, heavily polluted and severely polluted. The corresponding value intervals of the AQI for each level are (0, 50), (51, 100), (101, 150), (151, 200), (201,300) and more than 300. The histogram in Fig. 1 (C) also illustrates how the air pollution becomes even more serious with time. To analyse the AQI dynamics, the rate of inflow of pollutants was modelled with a piecewise function that takes different values during heating periods and non-heating periods by Tang et al. [37]. In this way the severity of air pollution in a year was divided into two levels of serious and not serious, according to the heating or non-heating periods. As can be seen from Fig. 1 (A) and (B), the length of the intervals between serious and not serious air pollution levels are different every year and in general there is an overall increasing trend year on year. Thus, the rate of inflow of pollutants should be expressed in a more general piecewise function.

In order to quantify the dynamics of air pollution and its impact on the risk of respiratory infection, Tang et al. [37] carried out an integrated data analysis to reveal an association between the AQI, meteorological variables and respiratory infection risk in Xi'an city in for the period of 15 November 2010 to 14 November 2016. An ordinary differential equation model with AQI-dependent incidence and AQI-based behaviour change interventions was developed to depict the AQI trend and respiratory infection dynamics. The main results showed that in terms of respiratory infection risk reduction, the persistent control of emissions in the China's blue-sky programme is much more effective than substantial social-economic interventions implemented only during the smoggy days [37].

Based on the data shown in Fig.1, we can see that the annual haze season does not strictly depend on the heating period, and it can be significantly affected by the randomness of environmental and meteorological factors, which brings challenges on how to determine the timings of switches (i.e. change points) between the not serious to the serious air pollution states based on the AQI data. However, the change

points play a key role in how to describe the AQI dynamics (i.e. the rate of inflow of pollutants) using a piecewise function [37]. Moreover, the number of ILI cases will also be random because of the randomness of demographic statistics.

Therefore, in order to quantify the effects of these random factors on AQI and ILI cases, we have developed stochastic differential models for AQI evolution and respiratory infection dynamics. Based on the characteristics of the AQI data, we estimated the series of change points in the AQI data between 2010 and 2016 and determined the parameters in each interval using a Bayesian method described by Tang et al. [38]. Further, by employing the methods developed by Allen et al. [2] we have proposed a stochastic differential equation model (i.e. stochastic SEIR model) for describing the respiratory infection dynamics, and MCMC method was used to identify the unknown parameters. Based on all estimated parameter values, the final purpose is to quantify the predictability of government policies and people's own initiatives, such as different traffic restriction measures and the intensity of self-protection actions [21]. We provide predictions of the air pollution condition and numbers of respiratory infections in 2017 and 2018 under various measures by stochastic simulations. The effectiveness of government intervention is verified by comparing the results of the actual AQI data and our predictions for the AQI data of 2017.

2 The AQI model

2.1 Stochastic differential equation of AQI

Human activities are the main sources of air pollution and lead to atmospheric contamination with a wide variety of pollutants such as carbon monoxide (CO), nitrogen oxide (NO_x), hydrocarbons, sulfur oxide and particulate matter (PM) [30]. However, some pollutants may be cleared by vegetation or dispersed by wind. The aggregative index AQI is used to describe the degree of air pollution in China. In order to describe the dynamics of the AQI, we establish the following simple ordinary differential equation (ODE):

$$\frac{dF(t)}{dt} = c' - \mu' F(t), \quad (2.1)$$

where F represents the AQI, which is an inflow-clearance process, and c' denotes the rate of inflow of pollutants into the air and μ' gives the rate of clearance of pollutants. It is apparent that meteorological and environmental conditions are affected by noise, so we introduce a stochastic differential equation (SDE) to better describe the variation in the AQI. An intuitive way to derive continuous diffusion models for model (2.1) is to consider the limit of the discrete model. If we regard the $dF(t)$ as a random variable, the model (2.1) becomes a simple birth and death process with linear transition rates in the general sense. The transition probabilities of this process define the probability of an inflow or a clearance in a short time dt . We note that in the discrete model the jump size of the linear process is Δh and the discrete distribution

of it satisfies

$$dF(t) = \begin{cases} \Delta h & \text{with probability } c' dt + o(dt), \\ -\Delta h & \text{with probability } \mu' F(t) dt + o(dt), \\ 0 & \text{otherwise,} \end{cases}$$

which indicates that the expectation of $dF(t)$ is $[c' - \mu' F(t)]\Delta h dt + o(dt)$ and the variance of $dF(t)$ is $[c' + \mu' F(t)]\Delta h^2 dt + o(dt)$.

Letting $\Delta h \rightarrow 0$, one has [20]

$$-\mu' \Delta h \rightarrow -\bar{\mu} \sigma^2 < 0, c' \Delta h \rightarrow \bar{c} \sigma^2 > 0, \mu' \Delta h^2 \rightarrow \sigma^2, c' \Delta h^2 \rightarrow 0.$$

where $\bar{\mu}$ and \bar{c} are two positive constants, and the derivations of the above formula were given in literature [20]. The drift function and the volatility of this birth and death progress are defined as $m(F)$ and $\sigma^2(F)$. Thus, the expressions of expectation and variance can be written as follows:

$$\begin{aligned} E(dF(t)) &= m(F) = [\bar{c} \sigma^2 - \bar{\mu} \sigma^2 F(t)] dt + o(dt), \\ \text{Var}(dF(t)) &= \sigma^2(F) = \sigma^2 F(t) dt + o(dt). \end{aligned}$$

Therefore, the approximation of the SDE yields:

$$dF(t) = [\bar{c} \sigma^2 - \bar{\mu} \sigma^2 F(t)] dt + \sigma \sqrt{F(t)} dW(t).$$

Further, denoting $c = \bar{c} \sigma^2$, $\mu = \bar{\mu} \sigma^2$, then the SDE model for AQI becomes as follows:

$$dF(t) = [c - \mu F(t)] dt + \sigma \sqrt{F(t)} dW(t), \quad (2.2)$$

where W is the Wiener process.

2.2 Change points for AQI

In model (2.2), it is assumed that the rates of both inflow and clearance of pollutants are assumed to be constant. However, in reality, AQI has distinct seasonal variations with lowest values in summer and higher values in other seasons, especially during winter. This phenomenon is mainly affected by seasonal variations in climate and vegetation, such as different rainfall patterns in the four seasons leading to variation in the pollutant clearance rate. However, the AQI data shown in Fig. 1 illustrate that in winter the air pollution is getting worse year by year, with associated decreases in the rate of clearance. Based on the above and without loss of generality, we assume that the clearance rate of pollutants $\mu(t)$ is a time-dependent and non-periodic function which can be given as follows [37]:

$$\mu(t) = \mu_0 + (\mu_1 + \mu_2 t) \sin(\omega t + \phi_0), \quad (2.3)$$

where \sin is the sin function, and $\omega = 2\pi/365$ is a periodic parameter assuming that there are 365 days in a year. ϕ and μ_0 are the phase parameter and offset parameter, respectively. The amplitude is controlled by μ_1 and a small amplitude increasing rate μ_2 .

For the rate of inflow of pollutants, we know that pollutants are not permanent and that their concentrations change with seasonal factors. Burning fuels for heating systems leads to higher influxes of pollutants than usual. $T_{n,1}$ and $T_{n,2}$ are two change points in a year, as defined in the Introduction. In this paper, Xi'an is taken as an example, where the heating systems are operational from 15 November to 15 March. **Mathematically, the two times $T_{n,1}$ and $T_{n,2}$ represent the end day in spring and next start day of heating in winter, that is to say, the time interval $[T_{n,1}, T_{n,2}]$ is ordinary days with a lower pollutant inflow rate and the corresponding interval $[n-1, T_{n,1}) \cup (T_{n,2}, n]$ is the heating period with a higher inflow rate, and n denotes the n th year.** $T_{n,1}$ and $T_{n,2}$ are the change points, such that the parameter $c(t)$ jumps from one value to another. Here the heating periods in each year provide good prior information when identifying the change points. Therefore, the inflow rate $c(t)$ could be defined by a piecewise function with an annual periodic function as follows:

$$c(t) = \begin{cases} c_{01}, & t \in [T_{n,1}, T_{n,2}], \\ c_{02}, & t \in [n-1, T_{n,1}) \cup (T_{n,2}, n] \end{cases} \quad (2.4)$$

for all $n = 1, 2, \dots$.

In addition to pollution from heating systems, there are other potentially time-varying factors, such as the increase in the number of cars, which contribute to the worsening climate and which, together with government control measures, can introduce randomness into the timings of serious air pollution occurrences. Thus, the change points in the AQI time series are not constants but random variables, which need to be estimated using statistical methods. In addition, the period of high inflow rates can vary. Therefore, the function $c(t)$ can be re-written as follows:

$$c(t) = \begin{cases} c_1, & t \in [T_0, T_1), \\ c_2, & t \in [T_1, T_2), \\ \vdots & \vdots \\ c_S, & t \in [T_{S-1}, T_S), \\ c_{S+1} & t \in [T_S, T_{S+1}] \end{cases} \quad (2.5)$$

with $T_0 = 0, T_{S+1} = T$, $c_i, (i = 1, \dots, S+1)$ defined as constants. $T_j (j = 1, \dots, S)$ denote the j -th change point, and S represents the number of change points during the interval $[0, T]$.

The dynamics of the AQI can now be depicted by the following SDE equation:

$$dF(t) = [c(t) - \mu(t)F(t)]dt + \sigma\sqrt{F(t)}dw(t), \quad (2.6)$$

where $c(t)$ and $\mu(t)$ are defined by equations (2.3), (2.4) and (2.5).

2.3 Identification of change points and parameter determination

With the exception of ω all of the parameters in model (2.6) are unknown and can be denoted as $\Theta = (c_1, \dots, c_{S+1}, \mu_0, \mu_1, \mu_2, \phi_0, \sigma)^T$, which is the vector of unknown parameters. Since an analytical solution of this SDE equation is hard to obtain, the

Euler- Maruyama algorithm could be used to discretize the SDE equation. The approximate solution satisfies the following discrete equation:

$$F_{k+1} = F_k + \alpha(F_k, \Theta)\Delta t + \hat{\sigma}(F_k, \Theta)\varepsilon_k, \quad k = 1, \dots, N', \quad (2.7)$$

where N' denotes the length of the data and $\alpha = [c(t) - \mu(t)F(t)]$, $\hat{\sigma} = \sigma\sqrt{F(t)}$, $\varepsilon_k \sim N(0, \Delta t)$. ε_k represents the increment of the Wiener process and obeys the normal distribution with mean zero and variance Δt . The time step Δt depends on the interval between data observations. Here AQI data were collected every day and so the time step $\Delta t = 1$ is small enough to guarantee that the approximation of the SDE equation is good. The estimation of the unknown parameter vector is based on the time series of the observed AQI data, which are represented by $F = (F_1, F_2, \dots, F_N)'$.

2.3.1 Posterior distribution of parameter vector Θ

According to the discrete equation (2.7), F_{k+1} can be treated as a random variable obeying the normal distribution:

$$F_{k+1} - F_k \sim N(\alpha(F_k, \Theta)\Delta t, \hat{\sigma}^2(F_k, \Theta)\Delta t).$$

Therefore, the likelihood function of F is:

$$p(F|\Theta) = \prod_{k=1}^{N-1} \frac{1}{(2\pi\sigma^2 F_k h)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} \frac{(F_{k+1} - F_k - (c(t) - \mu(t)F_k)h)^2}{F_k h}\right), \quad (2.8)$$

where $\Delta t = h$. For convenience, the piecewise function $c(t)$ can be represented by an indicator function. After expanding $\mu(t)$ and further manipulations we have

$$\begin{aligned} & \frac{(F_{k+1} - F_k - (c(t) - \mu(t)F_k)h)^2}{F_k h} \\ &= \frac{\left[(F_{k+1} - F_k - \left(\sum_{j=1}^{S+1} c_j I_a(T_{j-1} \leq t < T_j) \right) - (\mu_0 + (\mu_1 + \mu_2 t) \sin(\omega t + \phi_0))F_k h) \right]^2}{F_k h} \\ &= \left[\frac{F_{k+1} - F_k}{(F_k h)^{1/2}} - \left(\sum_{j=1}^{S+1} c_j I_a(T_{j-1} \leq t < T_j) \right) F_k^{-1/2} \sqrt{h} + (\mu_0 + (\mu_1 + \mu_2 t) \sin(\omega t + \phi_0)) F_k^{1/2} \sqrt{h} \right]^2 \\ &= \left[\frac{F_{k+1} - F_k}{(F_k h)^{1/2}} - \left(\sum_{j=1}^{S+1} c_j I_a(T_{j-1} \leq t < T_j) \right) F_k^{-1/2} \sqrt{h} + (\mu_0 + \mu_1 (\sin(\omega t) \cos \phi_0 + \sin \phi_0 \cos(\omega t)) \right. \\ & \quad \left. + \mu_2 t (\sin(\omega t) \cos \phi_0 + \sin \phi_0 \cos(\omega t)) F_k^{1/2} \sqrt{h} \right]^2, \end{aligned}$$

where the indicator function $I_a(\cdot)$ is a function defined on a set, having the value one for all elements in the set and the value zero for all elements not in the set.

Denoting Y as the vector obtained by the elements $(F_{k+1} - F_k)/(F_k h)^{1/2}$ for all k , and B as the matrix obtained by stacking

$$\begin{aligned} & \left[I_a(T_0 \leq t < T_1) F_k^{-1/2} \sqrt{h}, \dots, I_a(T_S \leq t < T_{S+1}) F_k^{-1/2} \sqrt{h}, -F^{1/2} \sqrt{h}, -\sin(\omega t) F^{1/2} \sqrt{h}, \right. \\ & \quad \left. -\cos(\omega t) F^{1/2} \sqrt{h}, -t \sin(\omega t) F^{1/2} \sqrt{h}, -t \cos(\omega t) F^{1/2} \sqrt{h}, \right], \end{aligned}$$

for all k , and $\mu_1 \cos \phi_0$, $\mu_1 \sin \phi_0$, $\mu_2 \cos \phi_0$, $\mu_2 \sin \phi_0$ as μ_1' , μ_1'' , μ_2' , μ_2'' , respectively. This function has the following simple form:

$$\left(\frac{F_{k+1} - F_k}{(F_k h)^{1/2}} - c(t) F_k^{-1/2} \sqrt{h} + \mu(t) F_k^{1/2} \sqrt{h} \right)^2 = (Y_k - B_k \beta)^2 = (Y_k - B_k \beta)' (Y_k - B_k \beta)$$

and

$$\sum_{k=1}^n (Y_k - B_k \beta)^2 = (Y - B \beta)' (Y - B \beta),$$

and hence the likelihood function can be rewritten as:

$$p(F|\Theta) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (Y - B \beta)' (Y - B \beta)\right), \quad (2.9)$$

with $\beta = (c_1, \dots, c_{S+1}, \mu_0, \mu_1', \mu_1'', \mu_2', \mu_2'')$. We have simplified the complex likelihood function into the same form as the Bayes' linear regression model of Y on B , so the estimation of parameter vector (β, σ) can be easily solved by using the Bayes' linear regression model:

$$Y = B \beta + \varepsilon, \quad \varepsilon_k \sim N(0, \sigma^2). \quad (2.10)$$

The next task is to calculate the posterior distribution of β and σ^2 . For model (2.10), the conjugate (joint) prior for β and σ^2 is the normal inverse-gamma distribution (NIG), which allows $\sigma \sim \text{IG}(n_0, s_0)$ and $\beta | \sigma \sim N(m, V)$, where n_0 and s_0 are the prior parameters of the gamma distribution for $1/\sigma^2$, and m and V are the prior mean vector and prior covariance matrix of β , respectively. The posterior distribution can then be derived from the likelihood function and the prior distribution. The expression for the latter and details of its derivation are given in Appendix A.

2.3.2 Detection of change-points

The identification of change points is important for delimiting the serious or not serious air pollution segments, and the parameter vector (β, σ) corresponding to the segments can be estimated when the change points are detected. The change point timings can be determined by a Bayesian approach and the change points T_1, T_2, \dots, T_S are considered as additional parameters. Combining the actual background data with a preliminary analysis of the AQI data, the heating systems are clearly the main cause of sudden changes in air pollution levels. The timing of such changes is different every year because of climate change and random factors, but the change points are nevertheless always near the start and end of the period when the winter heating systems are active. Based on this prior information on change points, if we denote the time series of the beginning and end of winter heating days as $Tin = (T_{01}, \dots, T_{0S})$, each change point T_i is assumed to be in the interval $[T_{0i} - M, T_{0i} + M]$, where M is a constant in the model and the value of M is given when the change points are estimated. Thus, an obvious choice for the prior distribution of each change point is a uniform distribution with the time interval $2M$, i.e.,

$$p(T_i) \sim U(T_{0i} - M, T_{0i} + M).$$

Futhermore, the estimation of each change point T_i is based on the AQI data in the discrete points of this interval $(T_{0i} - M, \dots, T_{0i} + M)$, which can be denoted as $F_{[T_{0i}-M, T_{0i}+M]}$. And the likelihood of each change point T_i is given by

$$L_i(T_i|F_{[T_{0i}-M, T_{0i}+M]}) = \prod_{k=T_{0i}-M}^{T_{0i}+M} \frac{1}{(2\pi\sigma^2 F_k h)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} \frac{(F_{k+1}-F_k-F_k \frac{(c_k-\mu(t)F_k)h}{F_k})^2}{F_k h}\right), \quad (2.11)$$

for $i = 1, \dots, S$. The parameter c_k in this function switch from c_i to c_{i+1} at change points T_i , which can be expressed as:

$$c_k = \begin{cases} c_i, & k \in [T_{0i} - M, T_i), \\ c_{i+1}, & k \in [T_i, T_{0i} + M]. \end{cases} \quad (2.12)$$

Finally, the posterior distribution of $T_i|F_{[T_{0i}-M, T_{0i}+M]}$ becomes

$$p(T_i|F_{[T_{0i}-M, T_{0i}+M]}) = \frac{L_i(T_i|F_{[T_{0i}-M, T_{0i}+M]})}{\sum_{i=T_{0i}-M}^{T_{0i}+M} L_i(T_i|F_{[T_{0i}-M, T_{0i}+M]})}. \quad (2.13)$$

Sampling from the posterior distribution (2.13) is based on the premise that the parameter vector (β, σ) is known. When the change points T_1, T_2, \dots, T_S and parameter vector (β, σ) are unknown, Chin Choy and Broemeling (1980) have shown a posterior distribution expression of a single change point for a general switching linear model. For a discrete time series of data \hat{F} , the length of \hat{F} is N , so the possible value of a single change point in this series is the corresponding time for each data point, which we denote as $\tau_1, \tau_2, \dots, \tau_{N-1}$. The posterior distribution of a change point is as follows [23]:

$$p(\hat{T}|\hat{F}) \begin{cases} \propto D(\hat{T})^{-n_1} |M_1|^{-1/2}, & \hat{T} = \tau_1, \tau_2, \dots, \tau_{N-1}, \\ = 0, & \hat{T} = T_N, \end{cases} \quad (2.14)$$

where

$$\begin{aligned} n_1 &= n_0 + \frac{N}{2}, \\ D(\hat{T}) &= s_0 + \frac{[(Y - B\beta_1)'Y + (M - \beta_1)'Q_0M]}{2}, \\ \beta_1 &= Q_1^{-1}[Q_0M + B'Y], \quad Q_1 = B'B + Q_0, \quad Q_0 = V^{-1} \frac{s_0}{n_0}, \end{aligned}$$

and all the symbols including Y, B, n_0, s_0, m, V correspond to the previous ones in subsection 2.3.1.

2.3.3 Algorithm and results

The Gibbs-sampling algorithm can be applied for generating the parameters sampled from the posterior distribution. Specifically, each change point is generated from the posterior distribution (2.14) based on the data $F_{[T_{0i}-M, T_{0i}+M]}$. After the change point

has been detected, the parameter vector β can be sampled from the multivariate normal distribution $N(m^*, \sigma^2 V^*)$ where m^* is a $(S+6)$ dimensional vector and V^* is an $(S+6) \times (S+6)$ matrix, and σ^2 can be drawn from the inverse gamma distribution $IG(n^*, s^*)$. For details of the expression about m^*, V^*, n^*, s^* refer to Appendix A. In this paper, the hyperparameters n_0, s_0 are chosen to be $n_0 = 300$ and $s_0 = 0.2$. A non-informative prior for the parameter vector is a $(S+6)$ dimensional zero vector for β and a diagonal matrix with elements 10^6 for V . The algorithm implementation process can be summarized as follows:

Step 1: Generate the sample of first change point T_1 on the space $[T_{01} - M, T_{01} + M]$ from the posterior (2.14), the one that maximizes the distribution values is the first optimum change point. Successively, get other change points in the same way.

Step 2: Give rational initial iteration values for parameter $\beta^0 = (c_1^0, \dots, c_{s+1}^0, \mu_0^0, \mu_1^0, \mu_1^{\prime 0}, \mu_2^0, \mu_2^{\prime 0})$ and σ^0 , and set the iteration number $i = 1$.

Step 3: Generate a draw of $\beta^i = (c_1^i, \dots, c_{s+1}^i, \mu_0^i, \mu_1^i, \mu_1^{\prime i}, \mu_2^i, \mu_2^{\prime i})$ from the conditional posterior $N(m_i^*, (\sigma^{i-1})^2 V_i^*)$ given in (5.5).

Step 4: Generate a draw of $(\sigma^i)^2$ from the conditional posterior $IG(n^*, s_i^*)$ given in (5.5).

Step 5: Increase i to $i+1$ and return to Step 2, until we have obtained a large enough sample for each parameter vector Θ_0 , where $\Theta = (\beta, \sigma)$.

For the AQI data, the length of the time series is 2192 days from 15 November 2010 to 14 November 2016, and the number of change points, i.e. the value of S was determined to be 11. To estimate the change points, we choose the value of M as 35. The theoretical basis for this proposal is that the period of serious air pollution during the heating times generally increases with increasing of the years, such that the beginning of the period may be ahead of the start of heating and the end of the period may be after the heating end time. For the convenience of estimation, it can be reasonably assumed that the period extends across both starts and ends of the heating time for five weeks, i.e. 35 days. The first step of the algorithm provided estimations of the change points to be (107, 375, 462, 766, 855, 1063, 1252, 1469, 1619, 1862, 1984), as listed in Table 1. This illustrates that a period of serious air pollution in those years could be marked by the series of change points. In addition, we have calculated the period of serious air pollution in each year in Table 1, denoted as m . The larger the value of m , the longer the period of serious air pollution. The results showed that the most serious air pollution period during the span of the existing data was from the beginning of heating in 2013 to its ending in 2014, and the value of m was 189. The period with the least air pollution was from the beginning of heating in 2013 to its ending in 2014, for which the value of m is 87. It follows from the Fig. 2 that the values of m is larger in the three most recent years than it was before, which indicates that the severity of air pollution is increasing.

Once the change points have been detected, we can then estimate the parameter vector Θ by using the Gibbs sampling algorithm outlined in this section to generate a sample draw from the joint posterior, with 30000 iterations used to obtain estimates after a burn-in period of 10000 iterations. If we assume that the rate of inflow of pollutants into the air is a periodic change, i.e. $c(t)$ satisfies function (2.4), the rate takes a value of c_{01} in seriously polluted times and c_{02} at other times. $T_{n,1}$ and $T_{n,2}$

are not the start and end days of heating but are determined by the estimated change points. Then the parameter vector is $\Theta_1 = (c_{01}, c_{02}, \mu_0, \mu_1, \mu_2, \phi, \delta)$. The estimation of the posterior mean and standard deviation of parameter vector Θ_1 is shown in Table 2. Meanwhile, the Monte Carlo standard errors (MCSE) of the estimates are quite small, meaning that the autocorrelations within the chains are weak.

With the baseline estimation, we are able to incorporate the air pollution trend. Fitting model (2.6) with (2.4) to the AQI data, we obtain the best fit estimation, shown in Fig. 2. The green points are results from a single fitting, and the red points are the mean values of 500 fitting iterations. From Fig. 2, it can be seen that the single fitting stochastic path shows the trend of increasing AQI year by year. In addition, the range of the oscillations is consistent with the actual AQI data. The goodness of fit results reveal that the stochastic model depicts the dynamics of the AQI well.

Further, if we assume that the rate of inflow of pollutants into the air is different in each interval between every two change points, i.e. $c(t)$ satisfies function (2.5), then the parameter vector is $\Theta_2 = (c_1, \dots, c_{S+1}, \mu_0, \mu_1, \mu_2, \phi, \delta)$. The estimation results of Θ_2 are shown in Table 3. The most seriously air polluted period in winter has the largest inflow rate with a value of 37.0314, and the least air polluted winter period has a lower inflow rate with a value of 26.5031. Under this assumption, we calculate the mean inflow rate in seriously and not seriously polluted period, respectively. These are denoted as $c_{\bar{0}1}$ and $c_{\bar{0}2}$, and their values are also listed in Table 3. Based on the parameter vector Θ_2 , the corresponding fitting results were obtained, which are shown in Fig. 3. One of the stochastic paths represented by the green points in Fig. 3 fit better than the path in Fig. 2.

Since the assumption of the inflow rate varying in each period is reasonable, the clearance rate $c(t)$ should choose the function (2.5) when fitting the AQI data, which can provide a better fit to the data and determine the parameters including the rates of inflow and clearance of pollutants for each year but cannot be used for prediction. However, the function (2.4) for $c(t)$ can effectively predict future changes in air pollution. This method can help us to analyze the effectiveness of control measures on air pollution, the details of which will be discussed later.

3 SEIS stochastic model

3.1 ODE model of respiratory infection dynamics

Air pollution can lead to adverse respiratory effects. This has been shown not only by associated data provided for epidemiological studies of the effects of air pollution on human health, but also by data which now clearly delineate the role of pollutant-mediated adverse interactions in allergic disease of human air-ways. To address this, we would like to propose a stochastic SEIS model to depict the respiratory infection development process, which is shown in Fig. 4. First we briefly introduce the ODE

model [37]:

$$\begin{cases} \frac{dS}{dt} = -\lambda(t) + \gamma_s I_s + \gamma_a I_a, \\ \frac{dE}{dt} = \lambda(t) - \sigma E, \\ \frac{dI_s}{dt} = \delta \sigma E - \gamma_s I_s, \\ \frac{dI_a}{dt} = (1 - \delta) \sigma E - \gamma_a I_a, \end{cases} \quad (3.1)$$

where $\lambda(t) = \beta(F(t))(1 - g(F(t)))S \frac{I_s + \theta I_a}{N}$ is the AQI-dependent incidence rate, which assumes that the transmission probabilities are related to the AQI and to human behavioral changes. The total population is $N = S + R + I_s + I_a$, and the incidence function is $\beta(F(t)) = \beta_1 F(t)$. The maximum protection rate on account of behavioral changes has the form $g(F(t)) = bF/(k_2 + F)$. The meanings of the state variables and the parameters are given in Table 4 or [37] for more details of the above model.

3.2 SDE model

The stochastic SEIS model corresponding to (3.1) is a system of Itô stochastic differential equations. Let S, E, I_s, I_a denote continuous random variables for densities of susceptible, exposed, infectious with symptoms and infectious without symptoms. In addition, the infinitesimal increment in a short time Δt is denoted as $\Delta X = \Delta X(t) = X(t + \Delta t) - X(t)$, where ΔX is the random vector $(\Delta S, \Delta E, \Delta I_s, \Delta I_a)$. Then

$$\Delta X(t) = \mathbf{E}(\Delta X(t)) + \sqrt{C \Delta t}, \quad (3.2)$$

here, the $\mathbf{E}(\Delta X(t))$ is the expectation and has the same form as the right-hand side of model (3.1). C is the covariance matrix of $\Delta X(t)$, which has the following form:

$$C = \begin{pmatrix} \lambda(t) + \gamma_s I_s + \gamma_a I_a & -\lambda(t) & -\gamma_s I_s & -\gamma_a I_a \\ -\lambda(t) & \lambda(t) + \sigma E & -\delta \sigma E & -(1 - \delta) \sigma E \\ -\gamma_s I_s & -\delta \sigma E & \delta \sigma E + \gamma_s I_s & 0 \\ -\gamma_a I_a & -(1 - \delta) \sigma E & 0 & (1 - \delta) \sigma E + \gamma_a I_a \end{pmatrix}$$

which is a symmetric matrix, and the square root of C is a unique positive definite matrix. We denote \sqrt{C} as a 4×4 matrix $A = (a_{ij})$. Taking the limitation of (3.2) as $\Delta t \rightarrow 0$, the following stochastic SEIS model is obtained[2]:

$$\begin{cases} dS = [-\lambda(t) + \gamma_s I_s + \gamma_a I_a]dt + \sum_{j=1}^4 a_{1j} dW_j, \\ dE = [\lambda(t) - \sigma E]dt + \sum_{j=1}^4 a_{2j} dW_j, \\ dI_s = [\delta \sigma E - \gamma_s I_s]dt + \sum_{j=1}^4 a_{3j} dW_j, \\ dI_a = [(1 - \delta) \sigma E - \gamma_a I_a]dt + \sum_{j=1}^4 a_{4j} dW_j, \end{cases} \quad (3.3)$$

where $W_j, j = 1, 2, 3, 4$ are four independent Wiener processes. In this model, the inflow rate for the dynamics of F is assumed to satisfy the function (2.5). The values of the parameters in model (3.3) were estimated by the MCMC method, except θ and σ which are from [37]. Parameter estimation for this stochastic SEIS model is described in the next subsection.

3.3 Results of parameter estimation and model fitting

In order to run the model, we acquired the ILI case data from the Shaanxi Center for Disease control and Prevention for the same period as the AQI data, i.e. from 15 November 2010 to 14 November 2016. For the SEIS stochastic differential equations (3.3), an adaptive Metropolis-Hastings (M-H) algorithm was used to estimate parameters in the model by carrying out the Markov Chain Monte Carlo (MCMC) procedure. Based on the AQI data and the ILI case numbers for the above period, we estimated the parameters, from the mean values of 10000 samples after a burn-in period of 5000 iterations. The meanings and estimated values of all of parameters relevant to the epidemiology of the respiratory infections are shown in Table 4.

With the baseline estimation, the numerical sample path of the SEIS stochastic equations can be obtained. Fig. 6 shows the best fit to the ILI cases data, in which the magenta points represent the ILI cases and the blue points show a simulation result with the estimated parameters. From Fig. 6, we can see that the correspondence between the simulation and the data is good. The characteristics of the ILI data including seasonal changes, progressive increase, and peaks have been reproduced by the simulations results, indicating that stochastic model (3.3) can describe the dynamics of respiratory infection well.

4 The effect of intervention measures

Investigating the short-term effects of intervention measures on the degree of air pollution is challenging since factoring in the uncertainty associated with changes in the climate and environmental change conditions is difficult. We now focus predictions of the AQI time series and the daily number of respiratory infection cases for two years from the stochastic model with the estimated parameters. Forecasts of the AQI could act as warning signals to the public before episodes of serious air pollution and help government departments make decisions about taking precautionary actions such as issuing orders restricting the number of vehicle trips. Thus, we considered six traffic restriction measures in Xi'an listed in Table 5. The type I is when the air is moderately or more seriously polluted ($AQI > 150$) and 20% of cars are banned from travelling for a week according to the last two digits of their licence plate numbers and type II is for the same pollution levels but when 50% of cars are banned from travelling for a week according to whether their licence plate numbers are odd or even. The type III and type IV measures are invoked when the air is lightly or more seriously polluted ($AQI > 100$) and 20% and 50% of cars, respectively, are banned from travelling for a week according to the same licence plate criteria as for types I and II above. The type V and type VI measures involve 20% and 50%, respectively, of cars being banned from travelling all of the time, irrespective of the pollution level, according to the above licence plate number criteria. To predict the respiratory infection cases under different levels of restrictions, the period between two change points in 2017 and 2018 is selected as the mean value for such a period m , which is 122. In addition, the parameters for the clearance rates are selected as the estimated results in Table 3, the inflow rates during seriously and not seriously polluted periods,

respectively, are selected as the values of c_{01} and c_{02} . While the inflow rate decrease to 90% and 80% of its initial value with the last two digits of the license plate rule and odd-and-even license plate rule, respectively.

AQI forecast reports could help the public, especially susceptible groups such as infants and old people, who could take measures to protect themselves by measures such as wearing masks and minimizing outdoor activities to reduce the risk of respiratory infections. Here we assumed that the rate of protection could increase to 1.1*b* and 1.2*b* because of human behavioural changes, which are regarded as another two intervening measures denoted by the type VII and type VIII. The efficacy of air pollution control policies from 15 November 2016 to 15 November 2018 in Xi'an are predicted in a random environment, using the mean simulated stochastic path of the model.

Simulated implementations of the government's restrictive measures, the mean of 500 stochastic path simulations and one good stochastic path of respiratory infection dynamics under six different traffic restriction measure types, represented by six different colors (red, yellow, lilac, bottle green, light green, cyan successively) are shown in Fig. 5 and Fig. 6. Based on the mean values of 500 simulations, the predicted numbers of polluted days ($AQI > 100$) and the total number of ILI cases are shown in Table 6. The blue line in Fig. 5 is a plot of the model predictions of the number of ILI cases in Xi'an from 15 November 2010 to 15 November 2016 and the prediction of the number of ILI cases in the next two years without any action, which can serve as a reference. The restricted strength increase in turn from traffic restriction measure type I to type VI, and the predicted results indicate the peak reduction of corresponding ILI cases. The results in Table 6 also illustrate that the stricter the control measures, the fewer the cumulative number of cases. It is worth noting that the cumulative number of ILI cases with VI is much lower than for the other cases and the number of polluted days is less than for the others, which demonstrates that regular and high strength traffic restriction measures are relatively efficient.

The validity and accuracy of our model are tested by comparing the predicted AQI data with the observed data from 15 November 2016 to 31 January 2018. The actual data for this period are represented by a steel gray color in Fig. 8 and the mean stochastic paths of AQI shown in Fig. 8 are represented by the six colors in accordance with Fig. 5. The polluted days for observed data and prediction during this period are counted and listed in Table 7, where the predicted results are the mean of 500 simulations under six of the government's restriction measures. By contrast, the actual data are closest to the simulation result of traffic restriction measure type IV, that is 50% of cars banned from travelling for one week based on whether their license plate numbers are odd or even when the air is lightly or more seriously polluted ($AQI > 100$). In fact, the government took actions on vehicle control in the winters of the 2016 and 2017 by invoking the type III measure or type IV measures in the winter of 2016, and the type V measure in the winter of 2017. Furthermore, the actual data in Fig. 8 illustrate that the most restrictive regular measures, type V and type VI, are more effective than only taking measures in seriously polluted periods.

Similarly, the efficacy of people's self-protection strategies can also be verified by the simulated data shown in Fig. 7 and Table 6. Through a horizontal comparison between type VII and type VIII, the greater the self-protection measures, the less

people suffer from respiratory diseases. However, the longitudinal comparison shows that the simulated cumulative numbers of ILI cases with a self-protection strategy is less than the case with traffic restriction measures. This indicates that people's self protection is more effective in reducing the incidence of respiratory diseases, probably because it can directly reduce the impact of pollutants on the respiratory system.

5 Discussion

This study is a prospective analysis of AQI data in relation to respiratory infections in Xi'an during 2010 to 2016. Firstly, we analysed how the AQI changes in one year and the features of its variations from one year to another and how they can be modeled with a stochastic model with change points. Secondly, considering the effects of air pollution on people's health, a stochastic SEIS model was established to describe the dynamics of respiratory disease transmission. Based on the collected data, the models were parameterised and their goodness of fit demonstrated. Finally, we forecasted the effects of several government intervention measures on the air pollution condition and the number of respiratory cases two years ahead.

The inflow-clearance process of air pollution was modeled by an SDE equation (2.6), which was derived from an ODE equation (2.1). Actual environmental changes, human activities, the inflow rate and the clearance rate were described by a piecewise function and a non-periodic time dependent function. The change points when the inflow rate switches between a serious and a not so serious pollution condition were estimated by using a Bayesian framework. The estimated results of change points shown in Table 1 indicate that air pollution has a worsening trend because the period between switches is increasing. This period m , which is the numbers of serious pollution days, can serve as an important indicator. In the years 2010, 2011 and 2012, there were relatively light air pollution conditions in winter, which was shown by the values of m being shorter than during the period when there was active heating. However, in the years 2013, 2014 and 2015, the situations were more serious than in the previous three years, since m was longer than the heating period. The data characteristics from the analysis of the change points for the AQI data suggest that the air pollution regime has a buffer time when the winter heating began several years ago, but that it became worse before a large influx of air pollutants resulting from winter heating in the three most recent years. Thus, not only the state of the air but also the self-regulation of the environment are getting worse and worse. In addition to the index m , we estimated the inflow rate at each period under two assumptions, and the results are shown in Table 3, which also indicates a growing trend for the inflow rate. For the dynamics of the AQI, the estimation of the values of the parameters in model (2.6) and the fitting of data shown in Fig. 3 are good interpretations.

Stochastic differential epidemic models were formulated for the dynamic of respiratory infections, assuming a constant population size. A function that reflects the impact of both the AQI and human behavior on the disease incidence rate was included. Most of the parameters in the stochastic SEIS model were estimated by MCMC methods and the results including mean values and standard deviations of parameters

are shown in Table 4. The solution of the stochastic epidemic model was numerically approximated by the Euler method, since analytical solutions of complicated stochastic differential equation are difficult to obtain. By using numerical simulation, the stochastic path fit to the true ILI cases data was good, which is shown by the blue line in Fig. 5.

To verify the effectiveness of government interventions, we first summarised six different restrictions on highway vehicles and two strengths of people's self protective measures. We then simulated the impact of these interventions by changing the values of parameters in the model as well as adding judgment and control statements to the numerical simulation program. Consequently, each intervention has a set of corresponding prediction results. The predicted numbers of polluted day ($AQI > 100$) and the total numbers of ILI cases with eight intervening measure types were obtained and are given in Fig. 5 and Fig. 7. Predictions, based on the mean values of 500 simulations shown in Table 6, revealed that air pollution can be effectively controlled by normalizing the restriction on traffic. In addition, we found suggestive evidence that minimizing exposure to outdoor air pollution can reduce the risk of developing respiratory diseases. Secondly, the actual data and predicted simulation results for the period from 15 November 2016 to 31 January 2018 were compared. The numerical simulation results showed that the government control measures are beneficial for improving air quality. The AQI data of nearly 15 months showed that government interventions have worked well, and that the more vehicles that are restricted, the better the air quality. Thus, we should choose "green" travel in our lives, by opting to take buses, or riding bicycles or suchlike rather than using motor vehicles.

A previous study established a deterministic model to describe the dynamics of air pollutants and respiratory infection [37], and this paper builds on that work with an analogous stochastic model, which is more accurate for describing the relations between air pollutants and respiratory infections. **Note that comparing with the main results shown in literature [37], one of interesting results in the present paper is that the change points have been estimated by real data rather than fixed them, which can more accurately evaluate and predict the situations of air pollution and the difference in air pollution per year, besides, the time interval of change point sequence can depict the tendency of air pollution to become serious. Moreover, the stochastic models can fit the AQI and ILI cases well, which could more actually depict the effects of randomness of environmental and meteorological factors. In terms of related parameters of the clearance rate such as μ_0 , μ_1 and μ_2 , the differences of their estimated values in deterministic model and stochastic model are obvious. This is embodied in that the clearance rate is large than that in deterministic model, in addition to the estimated values of inflow rate $c(t)$, the annual average estimated value in the stochastic model is significantly higher than that in the deterministic model, which reveals that the AQI could be underestimated in deterministic model.** Further work is still needed on the factors affecting air quality, on air pollution control strategies and related topics. In general, confirming the beneficial effects of current air pollution control in China is a challenge for environmental scientists, mathematicians, epidemiologists and clinical scientists, and more effort is needed to formulate public policies as well as reconciling different environmental objectives.

Appendix A: The prior distribution of β and σ^2 in Bayes' linear model

The probability density function of the prior distribution of β and σ^2 is given by

$$\begin{aligned} p(\beta, \sigma^2) &= p(\beta|\sigma^2)p(\sigma^2) = \mathbf{N}(m, \sigma^2 V) \text{IG}(n_0, s_0) \\ &= \frac{s_0^{n_0}}{(2\pi)^{|V|^{1/2}} \Gamma(n_0)} (\sigma^2)^{-(n_0+2)} \exp \left[- \left((\beta - m)' V^{-1} (\beta - m) + 2s_0 \right) / (2\sigma^2) \right], \end{aligned} \quad (5.4)$$

where $\Gamma(\cdot)$ represents the standard gamma function and IG is the inverse-gamma distribution. With m as the prior mean of the coefficient vector β , its prior variance is given by $\sigma^2 V$.

Using the Bayes rule we can then form the posterior distribution of the parameters, the posterior is determined by multiplying together the expressions in (2.9) and (5.4), the constant term $p(Y)$ is a regularization factor that does not depend on either β or σ^2 . The posterior distribution of β and σ^2 is

$$\begin{aligned} p(\beta, \sigma^2 | Y) &\propto (\sigma^2)^{-(n^*+2)} \exp \left[- \frac{(\beta - m^*)' (V^*)^{-1} (\beta - m^*) + 2s^*}{2\sigma^2} \right] \\ &= (\sigma^2)^{-1} \exp \left[- \frac{(\beta - m^*)' (V^*)^{-1} (\beta - m^*)}{2\sigma^2} \right] \times (\sigma^2)^{-(n^*+1)} \exp \left(- \frac{s^*}{\sigma^2} \right) \quad (5.5) \\ &\propto \mathbf{N}(m^*, \sigma^2 V^*) \text{IG}(n^*, s^*), \end{aligned}$$

where

$$\begin{aligned} m^* &= (V^{-1} + B'B)^{-1} (V^{-1}m + B'Y), \\ V^* &= (V^{-1} + B'B)^{-1}, \\ n^* &= n_0 + (N_1 - 1)/2, \\ s^* &= s_0 + [m'V^{-1}m + Y'Y - (m^*)'(V^*)^{-1}m^*]/2. \end{aligned}$$

References

1. Aguilera I, Pedersen M, Garcia-Esteban R, Ballester F, Basterrechea M, Esplugues A, Fernández-Somoano A, Lertxundi A, Tardón A, Sunyer J (2013) Early-life exposure to outdoor air pollution and respiratory health, ear infections, and eczema in infants from the INMA study. *Environ Health Persp* 121(3):387-392.
2. Allen LJS, Allen EJ, Jonsson CB (2006) The impact of environmental variation on hantavirus infection in rodents. In: Gumel AB, Castillo-Chavez C, Mickens RE, Clemence DP (eds) *Contemporary mathematics series. Proceedings of the joint summer research conference on modeling the dynamics of human diseases: emerging paradigms and challenges*, vol 410. AMS, Providence, pp 1-15.
3. Becker S, Dailey LA, Soukup JM, Grambow S, Devlin RB, Huang YT (2005) Seasonal variations in air pollution particle-induced inflammatory mediator release and oxidative stress. *Environ Health Persp* 113(8):1032-1038.
4. Bernstein JA, Alexis N, Barnes C, Bernstein IL, Nel A, Peden D, Diaz-Sanchez D, Tarlo SM, Williams PB (2004) Health effects of air pollution. *J Allergy Asthma Clin Immunol* 114:116-1123.
5. Bing Z, Wilson E, Jun B (2011) Controlling air pollution from coal power plants in China: incremental change or a great leap forward. *Environ Sci Technol* 45(24):10294-5.
6. Botchev MA, Verwer JG (2003) A new approximate matrix factorization for implicit time integration in air pollution modeling. *J Comput Appl Math* 157: 309-327.
7. Chan-Yeung MNW (2000) Air pollution and health. *Hong Kong Med J* 6:390-8.
8. Chauhan AJ, Johnston SL (2003) Air pollution and infection in respiratory illness. *Br Med Bull* 68:95-112.
9. Chen BH, Kan HD, Chen RJ, Jiang SH, Hong CJ (2011) Air pollution and health studies in China-policy implications. *J Air Waste Manage Assoc* 61(11):1292-1299.
10. Chen RJ, Zhang YH, Yang CX, Zhao ZH, Xu XH, Kan HD (2013) Acute effect of ambient air pollution on stroke mortality in the China air pollution and health effects study. *Stroke* 44:954-960.
11. Chow JC, Watson JG, Mauderly JL, Costa DL, Wyzga RE, Vedal S, Hidy GM, Altshuler Sam L, Marrack D, Heuss JM, Wolff GT, Pope CA, Dockery DW (2006) Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manage* 56(10):1368-1378.
12. Cropper M 2010. *What are the health effects of air pollution in China? Is Economic Growth Sustainable?* (Palgrave-Macmillan, London).

13. Genç DD, Yesilyurt C, Tuncel G (2010) Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. *Environ Monit Assess* 166:11-27.
14. HEI-International Scientific Oversight Committee (2010) Outdoor air pollution and health in the developing countries of Asia: A comprehensive review. Special Report 18. Health Effects Institute, Boston, MA.
15. Huang L, Zhou L, Chen J, Chen K, Liu Y, Chen XD, Tang FY (2016) Acute effects of air pollution on influenza-like illness in Nanjing, China: A population-based study. *Chemosphere* 147:180-187.
16. Ji Y (2014) Economic, growth, urbanization and air pollution in China: An empirical research based on panel data. *Energies* 7(7):4202-4220.
17. Kampa M, Castanas E (2008) Human health effects of air pollution. *Environ Pollut* 151:362-367.
18. Katsouyanni K (2003) Ambient air pollution and health. *Br Med Bull* 68:143-156.
19. Kim J, Cheo S (2010) Bayesian multiple change-point estimation with annealing stochastic approximation Monte Carlo. *Comput Stat* 25:215-239.
20. Kou SC. and Kou SG, (2004) A diffusion model for growth stocks. *Math Method Oper Res* 29:191-212.
21. Li FZ, Liu Y, Lü JJ, Liang LC, Harmer P (2015) Ambient air pollution in China poses a multifaceted health threat to outdoor physical activity. *J Epidemiol Commun* 69(3):201-204.
22. Low RB, Bielory L, Qureshi AI, Dunn V, Stuhlmiller DFE, Dickey DA (2006) The relation of stroke admissions to recent weather, airborne allergens, air pollution, seasons, upper respiratory infections, and asthma incidence, September 11, 2001, and Day of the Week. *Stroke*. 37:951-957.
23. Lyons TJ, Zheng WA (1990) On conditional diffusion processes. *Proc Roy Soc Edinburgh Sect A* 115:243-255.
24. Mehta S, Shin H, Burnett R, North T, Cohen AJ (2013) Ambient particulate air pollution and acute lower respiratory infections: a systematic review and implications for estimating the global burden of disease. *Air Qual Atmos Health* 6:69-83.
25. Mostofsky E, Schwartz J, Coull BA, Koutrakis P, Wellenius GA, Suh HH, Gold DR, Mittleman MA (2012) Modeling the association between particle constituents of air pollution and health outcomes. *Am J Epidemiol* 176(4):317C326.
26. Nielsen CP, Ho MS, (2007). Air pollution and health damages in China: An introduction and review. In: Ho MS, Nielsen CP (Eds), *Clearing the air: The health and economic damages of air pollution in China*. MIT Press, Cambridge MA.

27. Peng RD, Dominici F, Pastor-Barriuso R, Zeger SL, Samet JM (2005) Seasonal analyses of air pollution and mortality in 100 US cities. *Am J Epidemiol* 161:585-594.
28. Poon JPH, Casas I, He CF (2006) The impact of energy, transport, and trade on air pollution in China. *Eurasian geography and economics. Eurasian Geogr Econ* 47(5):568-584.
29. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol* 14(9):813.
30. Rohde RA, Muller RA (2015) Air pollution in China: Mapping of concentrations and sources. *PLoS ONE* 10(8): e0135749.
31. Samoli E, Schwartz J, Wojtyniak B, Touloumi G, Spix C, Balducci F, Medina S, Rossi G, Sunyer J, Bacharova L, Anderson HR, and Katsouyanni K (2001) Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA Project: A sensitivity analysis for controlling long-term trends and seasonality. *Stoch Environ Res Risk Assess* 109(4):349-353.
32. Shi K, Liu CQ, Ai NS, Zhang XH (2008) Using three methods to investigate time-scaling properties in air pollution indexes time series. *Nonlinear Anal Real World Appl* 9:693-707.
33. Shostya A(2016) Ambient Air pollution in China: Predicting a turning point. *Int Adv Econ Res* 22:295-307.
34. Song CB, Wu L, Xie YC, He JJ, Chen X, Wang T, Lin YC, Jin TS, Wang AX, Liu Y, Dai QL, Liu BS, Wang YN, Mao HJ (2017) Air pollution in China: Status and spatiotemporal variations. *Environ Pollut* 227: 334-347.
35. Stern G, Latzin P, Rösli M, Fuchs O, Proietti E, Kuehni C, Frey U (2013) A prospective study of the impact of air pollution on respiratory symptoms and infections in infants. *Am J Respir Crit Care Med* 187(12): 1341-1348.
36. Sun CW, Yuan X, Yao X (2016) Social acceptance towards the air pollution in China: Evidence from public's willingness to pay for smog mitigation. *Energy Policy* 92:313-324.
37. Tang SY, Yan QL, Shi W, Wang X, Sun XD, Yu PB, Wu JH, Xiao YN (2018) Measuring the impact of air pollution on respiratory infection risk in China. *Environ Pollut* 232:477-486.
38. Tang SY, Heron EA (2008) Bayesian inference for a stochastic logistic model with switching points. *Ecol Model* 219: 153-169.
39. Tang SY, Xiao YN, Yang YP, Zhou YC, Wu JH, Ma ZE (2010) Community-based measures for mitigating the 2009 H1N1 pandemic in China. *PLoS One* 5: e10911.

40. Touloumi G, Samoli E, Pipikou M, Tertre AL, Atkinson R, Katsouyanni K (2006) Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. *Statist Med*. 25:4164-4178.
41. Wang XK, Lu WZ (2006) Seasonal variation of air pollution index: Hong Kong case study. *Chemosphere* 63:1261-1272.
42. Wang QX, Liu PW, Zhang ZY (2013) A Characteristic finite volume element method for the air pollution model. *Numer Func Anal Opt* 34(6):664-694.
43. Wang P, Zhang ZY (2010) Quadratic finite volume element method for the air pollution model. *Int J Comput Math* 87(13):2925-2944.
44. Xiao QY, Ma ZW, Li SS, Liu Y (2015) The impact of winter heating on air pollution in China. *PLoS ONE* 10(1): e0117311.
45. Xia XL, Zhang A, Liang S, Qi QW, Jiang L, Ye YJ (2017) The association between air pollution and population health risk for respiratory infection: A case study of Shenzhen, China. *Int J Environ Res Public Health* 14(9):950.
46. Yu Bo, Huang CM, Liu ZH, Wang HP, Wang LL (2011) A chaotic analysis on air pollution index change over past 10 years in Lanzhou, northwest China. *Stoch Environ Res Risk Assess* 25:643-653.
47. Zhao Y, Wang SX, Duan L, Lei Y, Cao PF, Hao JM (2008) Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction. *Atmos Environ* 42:8442-8452.
48. Zhang JF, Smith KR (2007) Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. *Environ Health Persp* 115(6):848-855.

Table 1 The change point values and the periods between each two change points.

Change points	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}
Values	0	107	375	462	766	855	1063	1252	1469	1619	1862	1984
Period m		107	87		89		189		150		122	

Table 2 The estimated results for parameters in (2.4).

	c_{01}	c_{02}	μ_0	μ_1	μ_2	ϕ	δ
Mean	29.9888	27.3045	0.3047	0.0641	$1.5021 * 10^{-5}$	3.4439	2.3522
St.dev	1.7451	1.4855	0.0159	0.0188	$1.6369 * 10^{-5}$	0.2474	0.0315
MCSE	0.0119	0.0101	0.0001	0.0001	0.0000	0.0020	0.0003

Table 3 The estimated results for parameters in (2.5) and the mean inflow rates.

	c_1	c_3	c_5	c_7	c_9	c_{11}	c_{01}
Mean	29.1718	26.5031	35.0392	37.0314	28.3588	32.8716	31.4960
St.dev	3.0980	1.8897	3.2804	2.4893	1.7393	3.0036	
MCSE	0.0220	0.0146	0.0234	0.0178	0.0133	0.0226	
	c_2	c_4	c_6	c_8	c_{10}	c_{12}	c_{02}
Mean	24.0904	24.8243	26.7563	29.7524	27.1135	29.4371	26.9957
St.dev	1.7244	1.5923	2.0197	1.9954	1.6615	2.0563	
MCSE	0.0123	0.0097	0.0149	0.0150	0.0121	0.0155	
	μ_0	μ_1	μ_2	ϕ	δ		
Mean	0.3076	0.0620	$1.9100 * 10^{-5}$	3.2968	2.3434		
St.dev	0.0156	0.0206	0.0003	0.3246	0.0315		
MCSE	0.0001	0.0002	0.0000	0.0026	0.0002		

Table 4 The estimated values for the parameters and initial values of the SEIS model

Parameters	Definition	Mean value	(estimated)std	Source
β_1	baseline transmission coefficient	0.0035	$7.8987 * 10^{-5}$	MCMC
b	maximum protection rate due to interventions	0.666	0.0151	MCMC
σ	rate of progression to infection	1/3	–	[39]
γ_s	recovery rate for infectives with symptoms	0.1364	0.0005	MCMC
γ_a	recovery rate for infectives without symptoms	0.0214	0.0002	MCMC
δ	proportion of infected individuals who have developed flu symptoms	0.528	0.0028	MCMC
θ	modification factor in transmission coefficient of the asymptomatic infectious individuals	0.4	–	[37]
k_2	the AQI at which the maximum protection rate is of its half	2.673	0.0536	MCMC
$S(0)$	initial number of susceptibles	732.2	7.1205	MCMC
$E(0)$	initial number of exposed individuals	112.3	1.1684	MCMC
$I_s(0)$	initial number of infectives with symptom	68	–	[37]
$I_a(0)$	initial number of infectives without symptom	448.1	1.4209	MCMC

Table 5 The six traffic restriction measures considered.

Type I	$AQI > 150$ & 20% of cars banned from travelling for one week based on the last two digits of their license plate numbers
Type II	$AQI > 150$ & 50% of cars banned from travelling for one week based on whether their license plate numbers are odd or even
Type III	$AQI > 100$ & 20% of cars banned from travelling for one week based on the last two digits of their license plate numbers
Type IV	$AQI > 100$ & 50% of cars banned from travelling for one week based on whether their license plate numbers are odd or even
Type V	20% of cars banned from travelling at any time based on the last two digits of their license plate numbers
Type VI	50% of cars banned from travelling at any time based on whether their license plate numbers are odd or even

Table 6 The prediction results of polluted days and total ILI cases in 2017 and 2018.

		predicted polluted days in 2017	predicted ILI cases in 2017	predicted polluted days in 2018	predicted ILI cases in 2018
no measures	$c(t)$	153	155	25056	25509
Type I	$0.9c(t)$	150	151	24544	24279
Type II	$0.8c(t)$	148	149	23707	23600
Type III	$0.9c(t)$	137	138	23261	23347
Type IV	$0.8c(t)$	121	122	21965	21987
Type V	$0.9c(t)$	124	125	21887	21467
Type VI	$0.8c(t)$	94	96	18097	16420
Type VII	1.1b	148	157	15463	17383
Type VIII	1.2b	164	178	5207	6979

Table 7 The actual and predicted numbers of polluted days from 15 Nov 2016 to 31 Jan 2018

	Actual data	No measures	Type I	Type II	Type III	Type IV	Type V	Type VI
Actual polluted days / Predicted polluted days	176	232	230	226	208	184	192	147

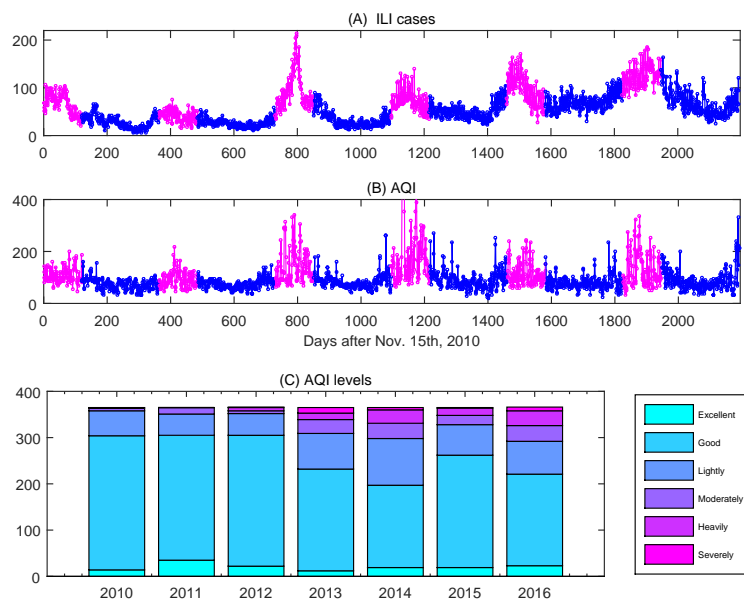


Fig. 1 Time series of the numbers of reported influenza-like illness (ILI) cases and the AQI from 15 Nov. 2010 to 14 Nov. 2016. (A): The numbers of new ILI cases reported during sentinel surveillance at seven hospitals; (B): The AQI for Xi'an, Shaanxi province; (C): The numbers of days at each air pollution level during 2010-2016. The pink curves in (A) and (B) correspond to winter periods from 15 Nov. to 15 Mar. of the following year.

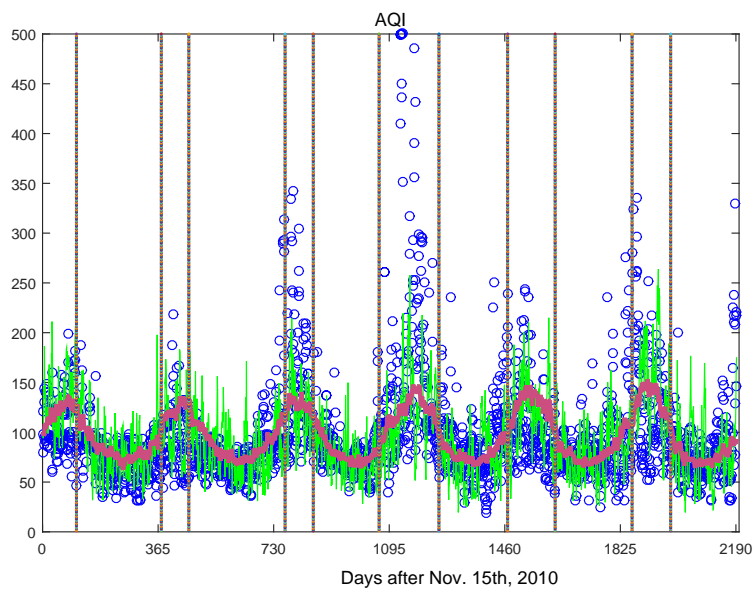


Fig. 2 Goodness fit of the AQI data at Xi'an based on model (2.2) with (2.3). Blue circles represent the AQI data during 15 November 2010 to 14 November 2016; the green line shows results of a single simulation with the estimated parameters and the red line shows the mean result from 500 simulations.

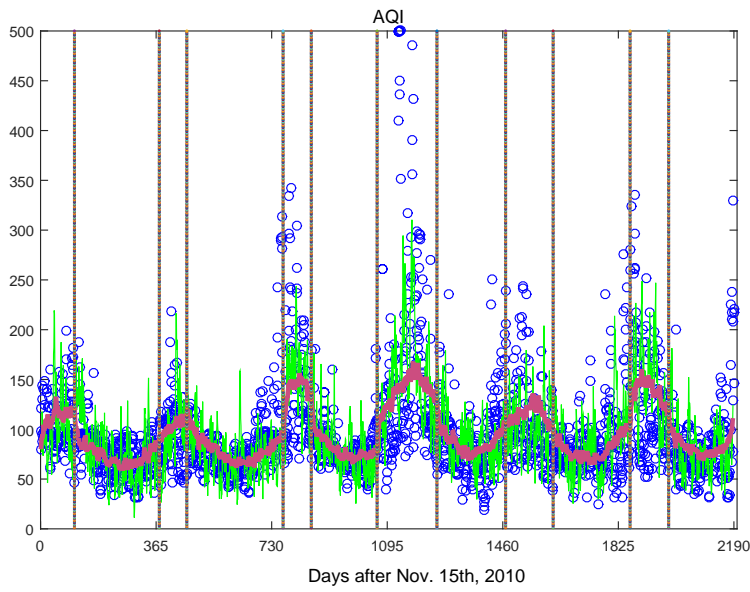


Fig. 3 Goodness fit of the AQI data at Xi'an based on model (2.2) with (2.4). Blue circle represent the AQI data during 15 November 2010 to 14 November 2016; the green line shows results of a single simulation with the estimated parameters and the red line shows the mean result from 500 simulations.

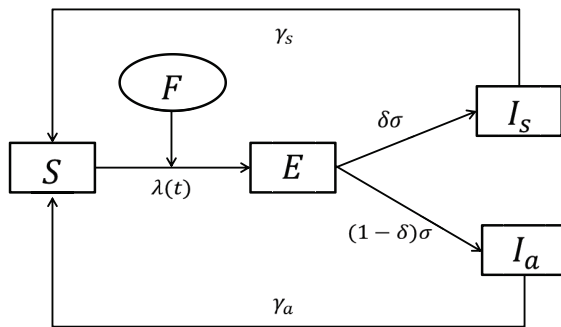


Fig. 4 Flow diagram representing respiratory infection disease development routes.

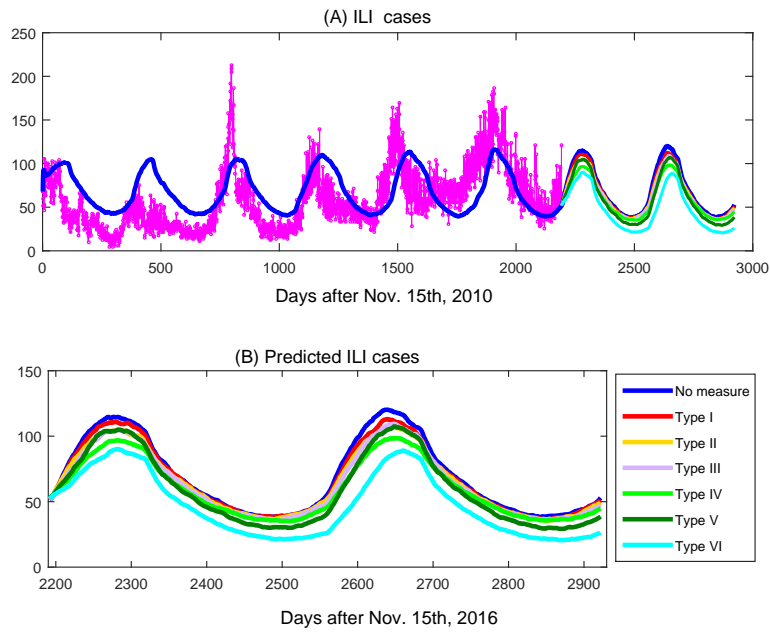


Fig. 5 The time series of respiratory infection from 15 November 2010 to 15 November 2016 and the predicted results under six different traffic restriction measures from 15 November 2016 to 15 November 2018, which are the means of 500 stochastic simulations. (A): The ILI cases data are shown by the magenta curve and the fitted data are represented by the blue curve. The red, yellow, lilac, bottle green, light green and cyan curves represent the predicted respiratory infection cases of Type I to VI; (B): The magnified version of seven predicted curves in Fig. 5(A).

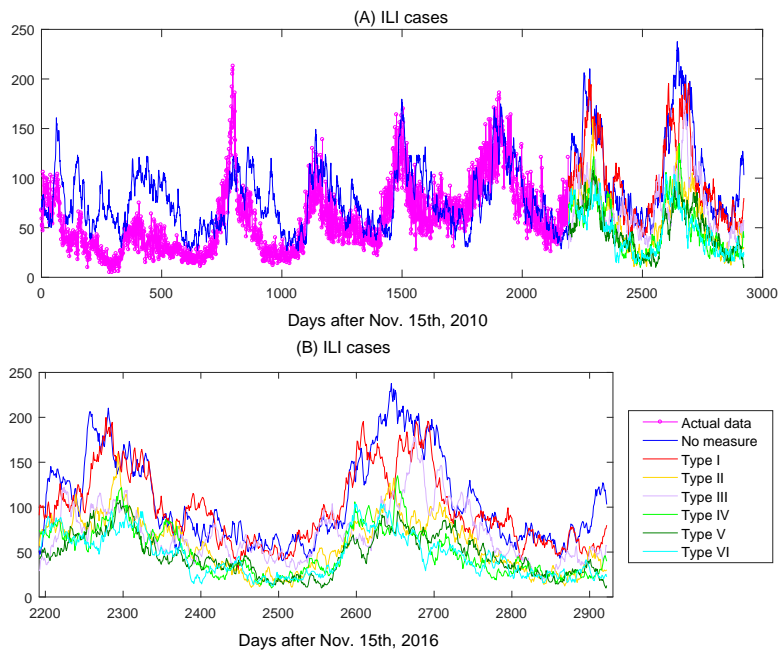


Fig. 6 One stochastic simulation from those illustrated in Fig. 5.

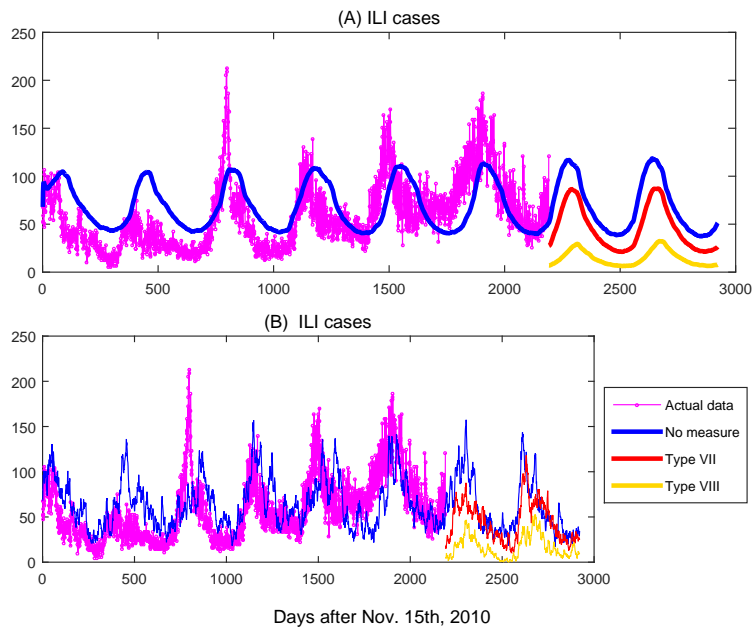


Fig. 7 Predicted results of respiratory infection cases under two kinds of self-protection measures from 15 November 2016 to 15 November 2018. The red curve and yellow curve are with protection measures of 1.1*b* and 1.2*b*, respectively. (A): They are means of 500 stochastic simulations. (B): The result of one stochastic simulation for good fitting.

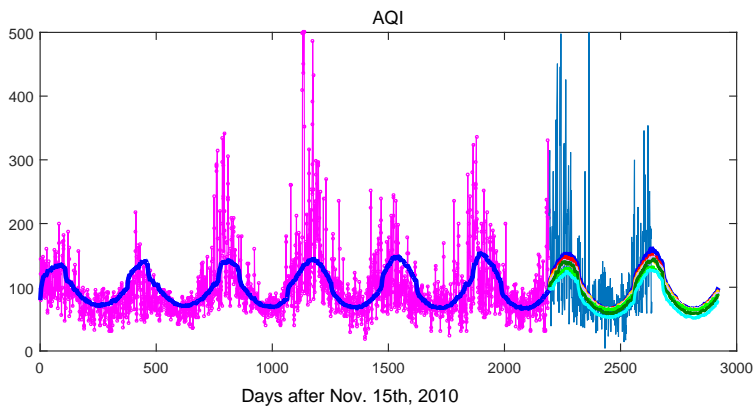


Fig. 8 Good fits for the time series of AQI from 15 November 2010 to 15 November 2016 and prediction results under six different traffic restriction measures from November 15th 2016 to November 15th 2018. The magenta curve represents the actual data that were used for parameter estimation and fitting data and the steel gray curve represents the actual data that were used for verify the validity and accuracy of model (2.6). The other six colors correspond to the six control measures as in with Fig. 5.