Particulate Air Pollution and Daily Mortality in Kathmandu Valley, Nepal: Associations and Distributed Lag

Srijan Lal Shrestha*

Central Department of Statistics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

Abstract: The distributed lag effect of ambient particulate air pollution that can be attributed to all cause mortality in Kathmandu valley, Nepal is estimated through generalized linear model (GLM) and generalized additive model (GAM) with autoregressive count dependent variable. Models are based upon daily time series data on mortality collected from the leading hospitals and exposure collected from the 6 six strategically dispersed fixed stations within the valley. The distributed lag effect is estimated by assigning appropriate weights governed by a mathematical model in which weights increased initially and decreased later forming a long tail. A comparative assessment revealed that autoregressive semi-parametric GAM is a better fit compared to autoregressive GLM. Model fitting with autoregressive semi-parametric GAM showed that a 10 μg m^{-3} rise in PM_{10} is associated with 2.57 % increase in all cause mortality accounted for 20 days lag effect which is about 2.3 times higher than observed for one day lag and demonstrates the existence of extended lag effect of ambient PM_{10} on all cause deaths. The confounding variables included in the model were parametric effects of seasonal differences measured by Fourier series terms, lag effect of mortality, and nonparametric effect of temperature represented by loess smoothing. The lag effects of ambient PM_{10} remained constant beyond 20 days.

Keywords: Ambient air pollution, autoregressive GAM, extended lag effect, Kathmandu valley, loess smoothing, mortality, statistical modeling.

1. INTRODUCTION

Particulate air pollution is a major environmental risk factor that can aggravate many health hazards to human population. This has been established in many studies conducted across the globe. Ambient particulate air pollution mainly in urban centers and industrial areas and indoor particulate air pollution mainly in rural areas of underdeveloped countries pose serious health threats to all those exposed. Various studies conducted at different parts of the world have demonstrated significant associations between different air pollutants mainly particulate matter (PM) and health effects such as mortality, lung cancer, hospitalization for respiratory and cardiovascular diseases, emergency room visits, asthma exacerbation, respiratory symptoms, restrictive activity days, loss of schooling, etc. [1].

Many studies have been published on the association between daily exposure to PM and mortality. In the study of 10 USA cities, Schwartz examined the daily effects of PM_{10} (particulate matter with diameter less than 10 micrometer) and reported that a 10 μg m^{-3} increase in the pollutant was associated with a 0.7% increase in daily mortality [2]. A study involving 29 European cities reported an association of 0.6 % increase in mortality per 10 μg m^{-3} increase in PM_{10} [3]. Combined results of 88 largest cities study of USA and 20 largest cities study of USA indicated an association between mortality and PM of approximately 0.5% change per 10 μg m^{-3} of PM_{10} [4]. More recent studies used an alternative statistical model and found an association of 0.27% per 10 μg m^{-3} of PM_{10} [5]. Some of the studies have also been conducted in cities outside of the US and European cities and in developing countries and reported the effect estimates similar to those found for US and European cities. Combined results of the studies conducted in Asia showed an association of 0.41% increase in all cause mortality per 10 μg m^{-3} increase in PM_{10} [6]. Similarly, a study on fine particulate pollution assessed by $PM_{2.5}$ (particulate matter with diameter less than 2.5 micrometer) and mortality in 9 California counties based upon time series data from 1999 till 2002 showed that a 10 μg m^{-3} increase (two day average) in $PM_{2.5}$ was associated with 0.6% increase in all cause mortality [7]. A more recent study on association between fine particulate pollution and mortality through extended follow up examination for 9 years in different cities of USA showed that increase in 10 μg m^{-3} of $PM_{2.5}$ was associated with 1.16 relative risk in overall mortality using Cox proportional hazards model after controlling for individual risk factors [8]. A cohort study in New Zealand urban areas for 3 years found the odds of all cause mortality in adults aged 30 to 74 years increased by 7% per 10 μg m^{-3} increase in average PM_{10} exposure using logistic regression model after controlling for age, sex, ethnicity, social deprivation, income, education, smoking history and ambient temperature [9]. A recent Health Effect Institute (HEI) research report (2010) on Public Health and Air Pollution in Asia (PAPA): Coordinated studies on short term exposure to air pollution exposure and daily mortality in four Asian cities showed that percent increase in mortality per 10 μg m^{-3} rise in PM_{10} was found to be 1.25 (0.8 – 3.01), 0.53 (0.26 – 0.81), 0.26 (0.14 – 0.37), and 0.43 (0.24 – 0.62) for Bangkok, Hong Kong, Shanghai, and Wuhan, respectively [10].

*Address correspondence to this author at the Central Department of Statistics, Tribhuvan University, Kirtipur, Kathmandu, Nepal; Tel: (977) 15539397; E-mail: srijan_shrestha@yahoo.com

1874-2823/12 2012 Bentham Open

Many studies have also been conducted where extended distributed lag effect of ambient particulate air pollution has been associated with health effects such as hospitalizations and mortality. In an analysis using data from 10 US cities, Schwartz has shown that if distributed lag effects are considered continued over several days, the relative risk of premature mortality that can be attributed to particulate pollution roughly doubles [11]. In a study by Goodman etal., showed that when 40 days lag effect was considered on total mortality due to black smoke, the effect was 2.75 times higher as compared to acute effect (3 day mean) [12]. A study conducted in Bangkok, Thailand and reported in 2008 demonstrated the effect due to extended lag from particulate air pollution on mortality. Effect on all cause mortality per 10 $\mu g\ m^{-3}$ increase in average PM_{10} was associated with increase in 1.2% for single day lag and 1.5% for 4 lagged days mean. Similarly, cardiovascular mortality increased from 0.5% to 1.9% and respiratory mortality increased from 1% to 1.9% [13].

Kathmandu valley's ambient air is also found to be polluted with particulate air pollution. Air quality monitoring of ambient air within Kathmandu valley in the past have shown this with majority of the days of a year exceeding the Nepal ambient air quality standard for 24 hour average PM_{10}. In the year 2004, altogether 193 days passed with 24 hour average concentrations exceeding the standard which is 120 $\mu g\ m^{-3}$ with most of the days falling in winter days. Monitoring of gaseous pollutants such as nitrogen dioxide, sulfur dioxide, carbon monoxide did not show such results with concentrations falling within national and WHO guidelines. Ambient air quality monitoring was done through 6 strategically fixed monitoring stations within the valley covering urban as well as rural areas installed by the then Ministry of Population and Environment (MOPE) of Nepal [14]. The major sources of particulate air pollution in the valley include dust re-suspension from vehicular movement and human activity, emissions from old vehicles, and cement and brick factories within the valley [15]. Several studies have also shown association between PM pollution and health effects in Nepal. A study conducted in Kathmandu valley has found that distributed lag effect of ambient particulate air pollution on respiratory morbidity is very high. Statistical analysis of the study showed that percent increase in chronic obstructive pulmonary disease (COPD) hospital admissions and respiratory admissions including COPD, asthma, pneumonia, and bronchitis per 10 $\mu g\ m^{-3}$ rise in PM_{10} are 4.85 % for 30 days lag effect, about 15.9 % higher than observed for same day lag effect and 3.52 % for 40 days lag effect, about 28.9% higher than observed for same day lag effect, respectively [16]. However, such studies conducted in Nepal have been very few. Moreover, most of the studies have extrapolated health effect coefficients derived from exposure response models of the studies conducted at other parts of the world [17].

The objective of this paper is to explore and model distributed lag effect of ambient particulate air pollution exposure in Kathmandu valley on all cause mortality using daily time series data. The extended exposure to PM_{10} is accounted by assigning weights to daily average PM_{10} based upon a suitable mathematical model. For statistical modeling, generalized linear model (GLM) and generalized additive model (GAM) are explored and applied. Data analysis for model building is carried out by SPLUS and Statistical Analysis System (SAS) software.

2. METHODOLOGY

2.1. Data

Analysis is based upon the data collected jointly in the Nepal Health Research Council (NHRC), Nepal study on 'Development of procedures and assessment of environmental burden of disease (EBD) of local levels due to major environmental risk factors', a World Health Organization (WHO) / Nepal funded project conducted in the year 2005 and the data compilation conducted by the author for individual research. Models could not be built from recent past data since daily monitoring of PM pollution has not been conducted through fixed monitoring stations in a regular basis.

2.1.1 Health Effect Data

Data on all cause mortality recorded as total daily deaths compiled from the leading hospitals in Kathmandu valley for one year during 2003/2004 is used. The hospitals are Bir Hospital (Kathmandu), TU Teaching hospital (Kathmandu), Patan hospital (Lalitpur) and Bhaktapur hospital (Bhaktapur). During the time of data compilation apart from these leading hospitals there were only small health centers and nursing homes / small hospitals which are excluded from the current analysis since major and serious cases which can lead to death of patients were ultimately referred to these hospitals for further treatment and almost all death cases were reported in these hospitals during that period of time in Kathmandu valley. Thus, exclusion of other health service providers from the current analysis can only have small impact on mortality coefficient which is ignored. All cause deaths include all deaths as mentioned in International Classification of Disease (ICD) codes (A00 – Z98) taken from the Department of Health Services, Nepal, 2003/2004 [18].

2.1.2 Exposure Data

Data compiled for PM_{10} on daily basis monitored from the 6 fixed stations installed within Kathmandu valley for the year 2003/2004 is used. For the same time period, daily average temperature data collected in Kathmandu valley monitored at the Tribhuvan International Airport are used. The six monitoring stations were set up at strategic locations to bring out the overall picture of the status of air quality in the valley. These comprise of one valley background station (Matsyagaon), two urban background stations (Bhaktapur and Kirtipur), two urban traffic area stations (Putalisadak and Patan) and one urban residential area station (Thamel). Daily PM_{10} was measured by medium volume sampler (MVS) through 24 hrs sampling which automatically measures PM_{10} continuously round the clock. The method of determination was gravimetric. It basically comprises of determination of the weight gained after a definite volume of ambient air has been sucked at a constant rate (2.3 m^3h^{-1}) through a pre-weighed filter paper. The filter papers were allowed to expose in a temperature and humidity controlled room before weighing and recorded before and after the sampling. The monitoring systems were calibrated once every month by a flow meter to check the flow rate. The flow meter itself was calibrated by a water flow meter [19].

2.2. Statistical Modeling

Statistical modeling is based upon autoregressive generalized linear model (GLM) and autoregressive semi-parametric generalized additive model (GAM) with log link function [20, 21]. In the models dependent variable is a count variable measuring daily hospital deaths and explanatory variables consist of a variable accounting for distributed lag effect of ambient particulate air pollution, a lagged variable and several confounding variables [22]. The semi-parametric GAM extends GLM by fitting both parametric terms as well as non-parametric functions f_i to estimate relationships between a response variable and predictor variables. Because f_i*'s* are generally unknown, they are estimated using some kind of scatter plot smoother [23]. Estimation of the additive terms in GAM is accomplished by replacing the weighted linear regression in GLM by the weighted back-fitting algorithm, known as the local scoring algorithm [24]. Two types of smoothers have been used namely, smoothing spline and locally weighted regression smoother (LOESS).

2.2.1. Model for Extended Lag Effect of Ambient Particulate Air Pollution

Under the initial screening of the lag effects on all cause mortality, it was detected that the value of the lag effect increased initially to a certain lag length and then decreased later. Consequently, the following mathematical model found suitable was taken for estimating weights for different lags.

$$W_t = c\left(t+1\right)e^{-\beta(t+1)} \tag{1}$$

where W_t is the weight assigned for t^{th} lag period, β is a constant and c is chosen such that $\sum_{t=0}^{k} W_t = 1$, k is a constant. W_k is the weight for maximum observed lag length.

2.2.2. Confounding Variables

Several confounding variables were considered for statistical modeling. These are weather, season, trend and day of week. Weather related variables such as average daily temperature and humidity are confounding variables in the study of air pollution epidemiology. In the present data analysis temperature is considered as one of the confounding variables. Humidity could not be considered as a confounding variable since its time series data was unavailable. Hospital admissions are also affected by seasonal changes. Consequently, Fourier series expansions were used to account for a seasonal effect. The daily time series data may also exhibit a long term trend. Therefore, a variable accounting for trend is also considered to see whether this is true or not. To distinguish between public holidays and working days, a dummy variable for holidays is additionally considered in the model.

2.2.3. De-Trended and De-Seasonalized Pollutant and Weather Variables

Though the seasonal effect and trend effect on the dependent variable are accounted for by inclusion of Fourier series terms and a trend variable, these variables can be correlated with the rest of the independent variables included in the model. This can result in multicollinearity between explanatory variables. PM_{10} and temperature are two such variables which contain seasonal / trend effects in themselves so that they could be correlated with seasonal variables included in the model. As a result, it becomes necessary to eliminate these effects which are accomplished by the following methodological procedure.

The effects of air pollution and temperature on mortality were separated from seasonal and trend effects by running linear regressions with the above variables as the dependent variables on seasonal variables and a trend variable as independent variables (trend variable was later excluded as it was not statistically significant). The resulting error components which could not be explained by regressions were indeed air pollution and temperature effects completely separated from seasonal effect. These separated effects representing air pollutant and weather effects on mortality were then included in the model as independent variables.

2.2.4. Model Adequacy Tests

Several measures have been considered for the test of the reliability of the models. These include overall goodness of fit, statistical significance of the estimated coefficients, accounting for overdispersion, residual analysis, and multicollinearity diagnostics.

The overall goodness of fit test is carried out by computation of deviance residual and Pearson generalized chi-square. The statistical significance of the estimated coefficients is done by Wald test. Similarly, presence of over-dispersion is assessed by estimating dispersion parameterϕ. If ϕ>1, then there is the problem of over-dispersion in the estimated model. Residual analysis is carried out through deviance and Pearson residuals. PP plots are used to assess normality of residuals. Autocorrelations are computed for an adequate large number of lags. Residual plots such as residuals in time sequence plots are also examined to detect model inadequacies. Multicollinearity is assessed through computation of variance inflation factors (VIF) [25].

2.2.5. Model Selection Criteria

Akaike's Information Criterion (AIC) is used to determine relevant explanatory variables that should be included in the final model. The model with minimum AIC was chosen.

3. RESULTS

3.1. Weights for Distributed Lag Effects of Ambient Particulate Air Pollution

The mathematical model expressed in Equation 1 is used to estimate weights for distributed lag effects of ambient particulate air pollution. For a predetermined lag length, a positive value of β is chosen such that the weights increase initially and then decrease resulting in a long tail. Thereafter, value of c is chosen such that the total weights sum up to unity. Several values are tested for β in between 0.1 to 0.4 since the curves showed an increase in weights initially and then decreased later. The Poisson model was fitted with other confounding variables and it was found that the deviance residual was minimum for β =0.3. The procedure is repeated for different lag lengths and similar results were obtained. Hence, values for β = 0.3, C = 0.091765 were chosen such that $\sum_{t=0}^{k} W_t = 1$.

The cumulative effect of ambient air pollution is examined for different lag periods in increasing order of lag lengths and corresponding pollutant coefficients were obtained. The procedure was repeated until the pollutant coefficient did not increase significantly. The corresponding distributed lag length was accepted for the final model which is 20.

The table (Table **1**) and corresponding figure (Fig. **1**) of weights for maximum lag length 20 is shown below.

Table 1. Weights for Distributed Lag Effects

Lag	Weight	Lag	Weight
0	0.067981	11	0.030088
1	0.100723	12	0.024147
2	0.111927	13	0.019265
3	0.110556	14	0.015291
4	0.102378	15	0.012083
5	0.091012	16	0.009511
6	0.078660	17	0.007460
7	0.066598	18	0.005834
8	0.055504	19	0.004549
9	0.045687	20	0.003539
10	0.037230		

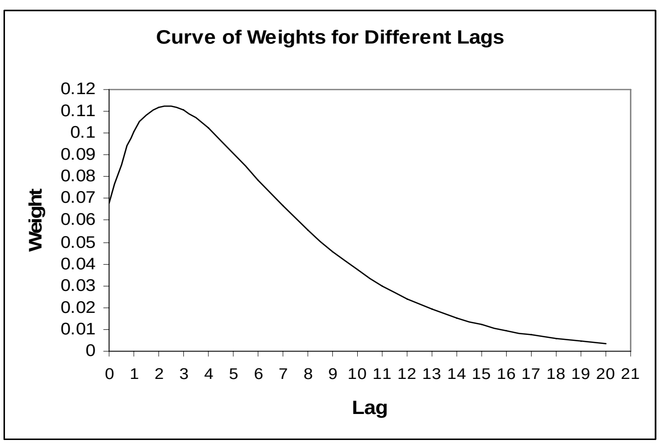

Fig. (1). Weights for Distributed Lag Effects.

3.2. De-Trended and De-Seasonalized Pollutant and Weather Variables

De-trended and de-seasonalized pollutant and weather variables are modeled through the following linear models since several model adequacy tests including residual analysis showed that linear models were more suitable than nonlinear models. Different sets of independent variables found statistically more significant for adjusted temperature series and adjusted PM_{10} series models were used to obtain de-trended and de-seasonalized pollutant and weather variables. Adjusted series was obtained as the difference between unadjusted and estimated values plus the average of the unadjusted series. Since mean values of unadjusted series were added to the deviation between unadjusted series and estimated series, the adjusted series is not just the deviation alone.

3.2.1. Model for Adjusted Temperature Series

Model for adjusted temperature series is:

$$t_{adj} = t_{unadj} - \hat{t}_{lm} + t_{mean} \tag{2}$$

where t_{unadj} is unadjusted temperature, $\hat{t}_{lm}$ is estimate of temperature from the fitted linear model and t_{mean} is the mean of unadjusted temperature series. $\hat{t}_{lm}$ is obtained from the following linear model:

$$\hat{t}_{lm} = \hat{\alpha}_0 + \hat{\alpha}_k Sin\left(\frac{2\pi kt}{m}\right) + \hat{\beta}_k Cos\left(\frac{2\pi kt}{m}\right) \tag{3}$$

where k is the number of oscillations in a year so that k=1,2,3,4 and t=1, 2, 3, ……, m; m is the total number of days in a given year. The fitted model produced significant estimates (p<0.1) as follows:

$$\hat{\alpha}_0 = 19.56;\ \hat{\alpha}_1 = 7.32;\ \hat{\beta}_1 = 0.21;\ \hat{\alpha}_2 = -0.21;$$
$$\hat{\beta}_2 = 2.02;\ \hat{\alpha}_3 = -0.44;\ \hat{\alpha}_4 = -0.3;\ \hat{\beta}_4 = -0.26$$

Here, t_{mean} =19.6 °C. It is to be noted that cos(6πt/365) is not included in the model since it is found to be statistically insignificant. For the fitted model, residual standard error is found to be 1.712 at 357 degrees of freedom with multiple R-Square: 0.9102, F-statistic: 517.1 at 7 and 357 degrees of freedom and p-value: ≈ 0.

3.2.2. Model for Adjusted PM_{10} Series

Model for adjusted PM_{10} series is:

$$PM_{adj} = PM_{unadj} - PM_{Estimate} + PM_{Mean} \tag{4}$$

where PM_{adj} is adjusted PM_{10}, $PM_{Estimate}$ is estimate of PM_{10} from the fitted linear model and PM_{Mean} is the mean of the unadjusted PM_{10} series. $PM_{Estimate}$ is obtained from the following linear model.

$$\begin{aligned} PM_{Estimate} &= \hat{\alpha}_0 + \hat{\alpha}_1 (Autumn) + \hat{\alpha}_2 (Winter) \\ &+ \hat{\alpha}_3 (Spring) + \hat{\alpha}_4 (Temperature) \\ &+ \hat{\alpha}_5 (Temperature^2) \end{aligned} \tag{5}$$

where α_i's are estimated coefficients. The fitted model produced significant estimates (p<0.0.01) as follows:

$$\hat{\alpha}_0 = 574.8;\ \hat{\alpha}_1 = -26.25;\ \hat{\alpha}_2 = 54.25;$$
$$\hat{\alpha}_3 = 88.50;\ \hat{\alpha}_4 = -49.15;\ \hat{\alpha}_5 = 1.29$$

Here, PM_{Mean} = 136.49 µg m^{-3}. For the fitted model, residual standard error is found to be 29.44 at 359 degrees of freedom with multiple R-Square: 0.7052, F-statistic: 171.8 at 5 and 359 degrees of freedom and p-value: ≈ 0. It is to be noted that seasonal variables are dichotomous contrast variables.

3.3. Distributed Lag Effects of Ambient Particulate Air Pollution

Analysis by autoregressive GLM showed that the effect of PM_{10} on all cause deaths increased as lag length increased from one day lag to 20 days lag. Thereafter, the effect remained approximately constant. If we examine Fig. (**2**), it is seen that mortality effect rose sharply till about 12 days lag effect and increased slowly and in small quantity up to 20 days. The difference of mortality effect between the two lags is very small. Even though statistical modeling can be done for 12 days lag effect, it is ultimately done for 20 days lag effect in the current analysis to maintain more precise estimate of extended lag effect of PM_{10}.

The percent increase in all cause deaths per 10 μg m^{-3} rise in PM_{10} is found to be 1.09 % for one day lag and 2.44 % for 20 days lag. The extended effect for 20 days lag is about 2.24 times higher than observed for one day lag which is a substantial increment and demonstrates the existence of extended and cumulative lag effect PM_{10} on all cause deaths. Estimation of all cause deaths and subsequent model building is therefore done for 20 days lag effect since distributed lag effect is effective up to this maximum lag and negligible for more extended lags i.e. more than 20 lagged days (Fig. **2**).

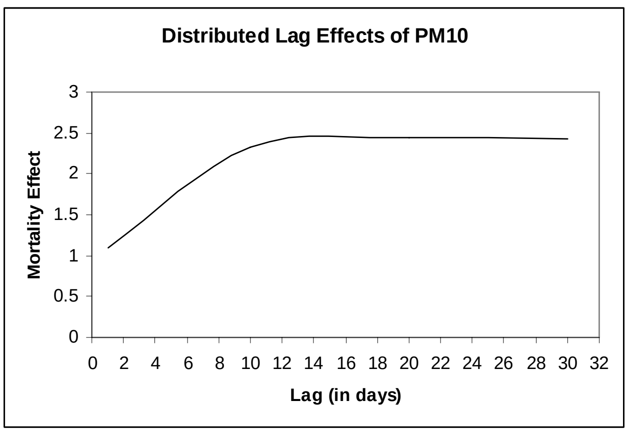

Fig. (2). Distributed Lag Effects of Ambient Particulate Air Pollution.

3.4. Autoregressive Models

Residual analysis of fitted GLM and GAM showed that deviance and Pearson residuals were slightly autocorrelated at 5^{th} lag which can be normally ignored since it cannot have a significant impact on model coefficients. The detailed analysis of GLM and GAM developed by excluding lagged term of the dependent variable has been shown in the author's earlier research work in this area [25]. However, the current analysis is carried out mainly to account errors due to ignoring marginally significant residual autocorrelations as observed in autocorrelation and partial autocorrelation plots. Thus, to maintain greater accuracy in the fitted models, the current model building process has developed more refined autoregressive GLM and autoregressive GAM with the inclusion of lagged parametric term of the dependent variable as an independent variable in the developed autoregressive models. The models, therefore, can also be viewed as modified forms or extensions of GLM and GAM without lagged terms.

3.5. Estimation of All Cause Deaths using Autoregressive GLM

3.5.1. Selection of Regressors with Minimum AIC

Among independent variables considered for modeling, a subset of the variables is chosen using Akaike's information criterion (AIC). The variables taken under consideration for modeling were seasonal variables, trend variable, day of week, temperature, air pollution, and the lagged term of the dependent variable. In the process of selection using AIC, trend, day of week and several sine and cosine terms are excluded from the model with minimum AIC = 1298.022. Since inclusion of the above variables as independent variables in the model generated higher AIC value, they were excluded from the final model.

3.5.2. Autoregressive GLM Estimates

The fitted model showed that all estimates of parameter coefficients are statistically significant with p values less than 0.05. Both PM_{10} and temperature are found to be positively associated with mortality. An increase of 2.6% of all cause mortality is estimated with 10 μg m^{-3} increase in ambient PM_{10} value with 95% confidence interval equal to 0.7% - 4.6%. The quadratic effect of temperature is also found to be statistically significant implying quadratic nonlinear association between the dependent variable and temperature. As far as seasonal and cyclic effects are considered, only sin(8πt/365) and cos(8πt/365) are included in the model. It implies that seasonal variations are significant with k = 4 meaning that cyclic variations with 4 complete oscillatory movements throughout a year with each cycle having only a quarter of a year as period are found to be statistically associated with mortality variations. The result signifies that cyclic patterns representative of seasonal variations are also statistically significant (Table **2**).

Table 2. Autoregressive GLM Parameter Estimates for All Cause Deaths

Parameter	Coefficient	Standard Error	*t* Value	*p* Value
Intercept	-6.3923	2.8136	5.1618	0.0231
Sin(8πt/365)	0.0834	0.0421	3.9241	0.0476
Cos(8πt/365)	-0.1097	0.0428	6.5832	0.0103
Temperature	0.7325	0.2891	6.4201	0.0113
$Temperature^2$	-0.0182	0.0074	6.9066	0.0145
Lag 5	-0.0489	0.0175	7.8191	0.0052
PM_{10}	0.0026	0.0010	2.427	0.0086

Model adequacy tests for the GLM model are provided in Appendix A.

3.6. Estimation of All Cause Deaths Using Autoregressive GAM

Two nonparametric smoothers are considered for generalized additive modeling namely smoothing spline and locally weighted regression smoother (LOESS). Since use of LOESS resulted in smaller residual deviance as well as more statistically significant nonparametric smoother, it was

Table 3. Parameter Estimates of All Cause Deaths Using Autoregressive GAM

Parameter	Estimate	Standard Error	*t* value	Pr > |*t*|
Intercept	0.94162	0.14785	6.38	< 0.0001
Sin (8πt/365)	0.08570	0.04217	2.02	0.0442
Cos (8πt/365)	-0.10570	0.04264	-2.50	0.0136
Lagged Term of Mortality	-0.04641	0.01749	-2.65	0.0084
PM_{10}	0.00257	0.00100	2.56	0.0108

preferred against smoothing spline in the current model building process. A semi-parametric GAM (with autoregressive dependent variable) is fitted by using a nonparametric smooth function for temperature and parametric terms for the other variables.

3.6.1. Model Parameter Estimates and Summary Statistics

The fitted autoregressive semi-parametric GAM showed statistically significant coefficient estimates for parametric as well as nonparametric effects. Sine and cosine terms are found to be statistically significant with tri-monthly oscillatory period. A 10 µg m^{-3} increase in PM_{10} is found to be associated with 2.57 % increase in all cause deaths (Table **3**). The value is approximately same as obtained in autoregressive GLM which is 2.60%. Moreover, a Loess smoother of temperature with 3.5 degrees of freedom is also found to be statistically significant with $\alpha = 0.01$ (Table **4**). This statistical significance of the nonparametric smoother justifies the application of GAM and demonstrates the existence of a nonlinear association for temperature (Table **5**).

Table 4. Fit Summary for Smoothing Component

Component	Smoothing Parameter	df
Loess (Temperature)	0.534722	3.50015

Model adequacy tests for the GAM model are provided in Appendix B.

4. DISCUSSION AND CONCLUSION

For estimating all cause deaths GLM and GAM with inclusion of lagged term of the dependent variable as independent variable are explored for their suitability as statistical models for associating mortality with ambient particulate air pollution. A comparative assessment revealed that autoregressive GAM is more suitable in modeling all cause deaths in Kathmandu valley compared to fully parametric autoregressive GLM. This is mainly because nonlinear effect of temperature assessed by Loess smoother is found to be statistically significant with $\alpha = 0.01$. Moreover, a semi-parametric autoregressive GAM is found to be more suitable instead of fully non-parametric autoregressive GAM since though temperature is found to have nonlinear effect on the dependent variable same is not found to be true for PM_{10}. Therefore, a semi-parametric model with parametric effects of PM_{10} and other confounding variables and a nonparametric smoother of temperature are included in the final GAM. However, the effect of PM_{10} is found to be only marginally different between GLM and GAM. The goodness of fit is marginally better in autoregressive GAM compared to autoregressive GLM and examination of residual autocorrelations and partial autocorrelations show marginally lower values as compared to GLM. Examination of standardized deviance residuals showed only a single significant outlier in both fitted models. Fitted models include the following characteristics.

- Fitted models contain trend and seasonally adjusted series for distributed lag effect of ambient particulate air pollution which verified that short term effect grossly underestimates the actual effect on all cause mortality that can be attributed ambient particulate air pollution as demonstrated in Kathmandu valley, Nepal.
- Elimination of trend and seasonality in daily time series data of PM_{10} greatly reduced the problem of multicollinearity. Several confounders such as trigonometric (sine and cosine) terms with k=4 for seasonal representation and temperature are also found to be statistically significant.
- The fitted GLM revealed that the percent increase in all cause deaths per 10 µg m^{-3} rise in PM_{10} increased up to 20 lagged days and remained constant thereafter. As estimated by autoregressive GAM, an increase in 2.57 % all cause deaths is estimated for 10 µg m^{-3} rise in PM_{10} which is marginally higher than observed for GAM without lagged variable (2.44%).

Developed models are based upon one year daily time series data on mortality and exposure. Mortality data was collected from records of the leading hospitals within Kathmandu valley. Some small scale nursing homes and hospitals were left out since major and serious cases which

Table 5. Analysis of Deviance

Source	df	Sum of Squares	Chi-Square	Pr > Chi-Square
Loess (Temperature)	3.50015	13.272890	13.5368	0.0058

may lead to deaths of patients were usually referred to these hospitals. Consequently, reported deaths were mostly from these hospitals. Under the assumption that this will not have significant bias on the mortality estimate only the leading four hospitals were taken for data compilation. However, the permanent residencies of died patients were not recorded as it was relatively difficult to retrieve information due to poor database system that prevailed at that time in the hospitals. As a result, misclassification of some died patients may have occurred which can be regarded as a limitation of the study.

Finally, the extent of effects on all cause mortality from exposure to ambient particulate air pollution is found to be substantial in Kathmandu valley. Estimate of all cause mortality is also higher compared to the findings of other studies at different parts of the world based upon only few days lag effect. However, similar to the findings of distributed lag effects studies at other parts of the world, the current analysis also showed that extended lag effect of air pollution on mortality is much higher (slightly higher than double) than single or few days lag effects. For instance, Schwartz has shown that if distributed lag effects are considered continued over several days, the relative risk of premature mortality that can be attributed to particulate pollution roughly doubles. The results, therefore, raise health concerns to all valley inhabitants caused by particulate air pollution. Even though efforts have been made in the direction of reducing the particulate levels in the valley, its urban air is still highly polluted. Therefore, this is a matter of serious concern and further steps are required to reduce pollutant levels in coming years.

ACKNOWLEDGEMENTS

The author is grateful to Nepal Health Research Council (NHRC), Kathmandu, Nepal for initiating the project entitled 'Development of procedures and assessment of environmental burden of disease of local levels from major environmental risk factors' and World Health Organization (WHO / Nepal) for providing fund and support for the project. Sincere thanks goes to Mr. Ram Hari Khanal, Coordinator, Sunil Babu Khatri, Research Assistant and Shivendra Thakur, Research Assistant of the project. Deep appreciation and many thanks to Dr. Mrigendra Lal Singh, Professor of Statistics and Dr. Iswori Lal Shrestha, Environmental Expert for their invaluable guidance and sharing knowledge and experiences in the author's research work. Many thanks to the reviewers of this manuscript for providing their valuable suggestions and pointing out some errors.

CONFLICT OF INTEREST

None declared.

APPENDIX A

Model Adequacy Tests for GLM

Overall Goodness of Fit

The overall goodness of fit of the fitted model is judged by deviance residual and Pearson chi-square. Deviance residual is found to be 356.35 at 353 degrees of freedom and Pearson chi-square is found to be 331.14 at 353 degrees of freedom. Both are statistically insignificant with p values 0.44 and 0.79, respectively. The statistical insignificance of the statistics suggests that the Poisson model fits well for the given data set.

Residual Analysis

Normality Tests of Residuals

Kolmogorov-Smirnov nonparametric test and the P-P plots of the residuals (Figs. **3**, **4**) show that the distribution of the deviance residual and Pearson residual may be regarded as normally distributed with p values greater than 0.05 (p=0.43 for Pearson residual and p=0.49 for deviance residual). It is to be noted that p values are much higher than 0.05 which is preferable.

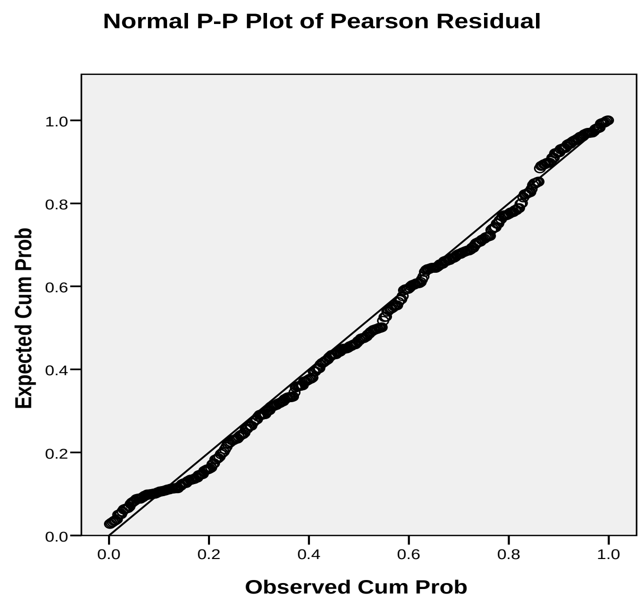

Fig. (3). Normal P-P Plot of Pearson Residuals.

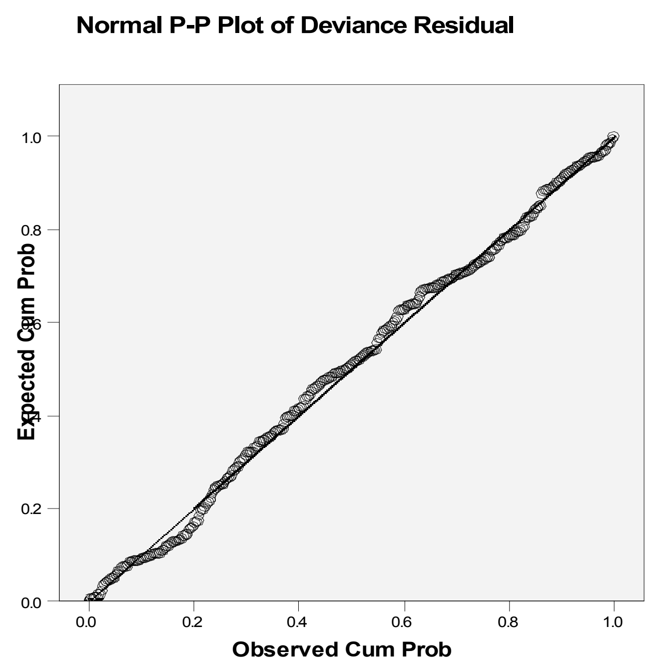

Fig. (4). Normal P-P Plot of Deviance Residuals.

Autocorrelations and Partial Autocorrelations of Residuals

In time series models, it is necessary to observe autocorrelation and partial autocorrelation plots of the residuals to examine if there are some statistically significant autocorrelations. Examination of the plots shows nonexistence such correlations up to a sufficiently large lag (15) for deviance and Pearson residuals.

Examination of Residual Plots

The partial residual plots show linear associations which includes quadratic transformation of temperature. This would imply nonlinear association between transformed dependent variable and temperature. The standardized deviance residual plot in time sequence does not show any pattern or trend and looks like errors are randomly distributed. This implies that variance of the residuals is fairly constant. In addition, the plot shows only one significant outlier (>3) (Fig. **5**).

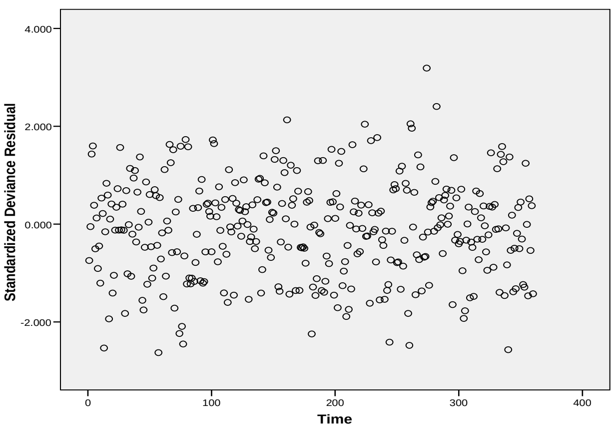

Fig. (5). Scatter Plot of Standardized Deviance Residual.

Variance Inflation Factor (VIF)

Examination of variance inflation factor (VIF) which is an important indicator of multicollinearity showed that the values are close to one except for temperature and its square term which are obviously high (Table **6**).

Table 6. Variance Inflation Factors

Variable	VIF
Sin(8πt/365)	1.03
Cos(8πt/365)	1.01
Temperature	208.4
Temperature^2	208.4
Lagged Term of Mortality	1.03
PM_{10}	1.02

APPENDIX B

Model Adequacy Tests for GAM

Goodness of Fit

Deviance residual is found to be 352.0 for 353.0 degrees of freedom which is statistically insignificant with 0.51 *p*-value. The value is higher than the corresponding *p* value in GLM (0.44) which implies the goodness of fit is marginally better in GAM compared to GLM.

Residual Analysis

Kolmogorov-Smirnov nonparametric test and the *P-P* plots of the residuals show that the distribution of the deviance residual and Pearson residual can be regarded as normally distributed with *p* values greater than 0.05 (*p*=0.32 for Pearson residual and *p*=0.51 for deviance residual). Examinations of autocorrelations and partial autocorrelations show insignificant correlations (<0.07) up to a sufficiently large lag (15) for deviance and Pearson residuals. This suggests that the errors are approximately independently distributed (Figs. **6**, **7**). Examination of partial residual plots shows nonlinear association between the dependent variable and temperature. Considering standardized deviance residuals for the detection of outliers, only one of them is found with value greater than 3 (Fig. **8**). Estimated coefficients remained approximately same after elimination of the outlier. Consequently, it is retained in the model. The partial residual plot of temperature is also shown (Fig. **9**). It shows the existence of nonlinear association between temperature and mortality.

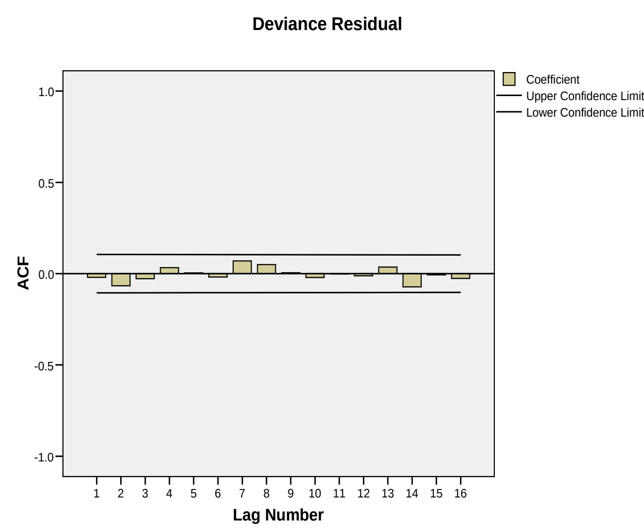

Fig. (6). Autocorrelation Plot of Deviance Residual.

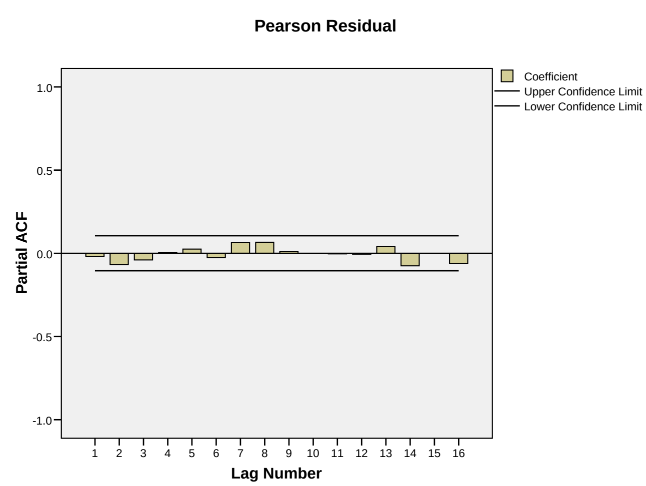

Fig. (7). Partial Autocorrelation Plot of Pearson Residual.

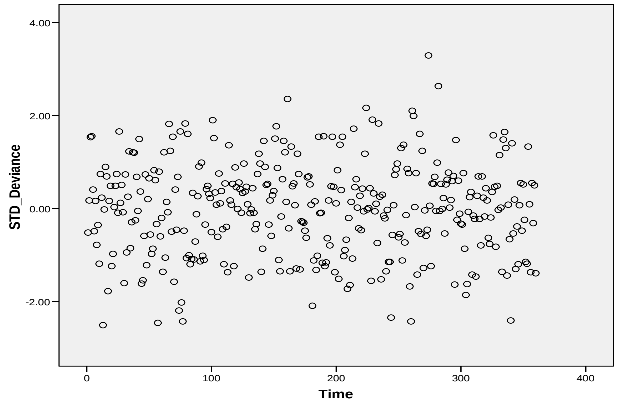

Fig. (8). Scatter Plot of Standardized Deviance Residual.

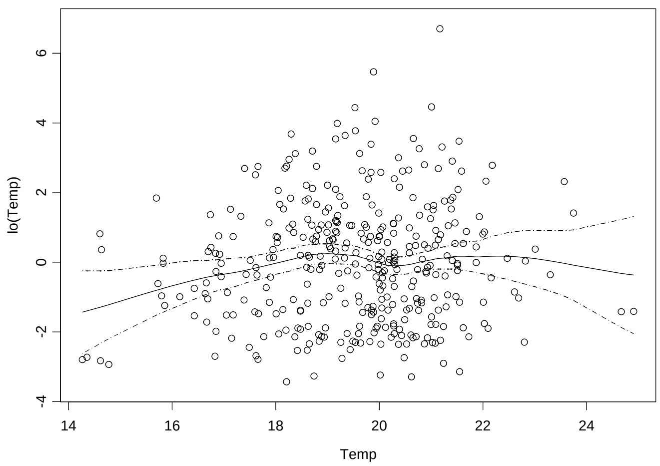

Fig. (9). Partial Residual Plot of Temperature.

REFERENCES

[1] Ostro B. Outdoor air pollution; environmental burden of disease series No. 5, World Health Organization, Geneva 2004.

[2] Schwartz J. Assessing confounding, effect modification, and threshold in association between ambient particles and daily deaths. Environ Health Perspect 2000a; 108: 563-8.

[3] Katsouyanni K, Touloumi G, Samoli E, *et al*. Confounding and effect modification in the short term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology 2001; 12: 521-31.

[4] Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. Fine particulate air pollution and mortality in 20 US cities, 1987-1994. N Engl J Med 2000b; 343: 1742-9.

[5] Dominici F, McDermot A, Zeger SL, Samet SM. On the use of generalized additive models in time series studies of air pollution and health. Am J Epidemiol 2002; 156: 193-203.

[6] HEI. Health effects of outdoor air pollution in developing countries of Asia: a literature review, Special report 15; Health Effects Institute, USA 2004.

[7] Ostro B, Broadwin R, Green S, Feng W, Lipsett M. Fine particulate air pollution and mortality in nine California counties: results from CALFINE. Environ Health Perspect 2006; 114 (1): 29-33.

[8] Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality. Am J Respir Crit Care Med 2006; 173: 667-72.

[9] Simon H, Blakely T, Woodward A. Air pollution and mortality in New Zealand: cohort study. Epidemiol Community Health, doi:10.1136/jech.2010.112490. 2010. [Online]. available: http://jech.bmj.com/content/early/2010/10/21/jech.2010.112490.full

[10] HEI. Public health and air pollution in Asia (PAPA): coordinated studies on short term exposure to air pollution and daily mortality in four Asian cities, research report 154, Health Effects Institute, USA, 2010.

[11] Schwartz J. The distributed lag between air pollution and daily deaths. Epidemiology 2000; 11: 320-6.

[12] Goodman PG, Dockery DW, Clancy L. Cause specific mortality and extended effects of particulate pollution and temperature exposure. Environ Health Perspect 2003; 112:179-85.

[13] Vichit-Vadakan N, Vajanapoom N, Ostro B. The public health and air pollution in Asia (PAPA) project: estimating the mortality effects of particulate matter in Bangkok, Thailand, Environ Health Perspect 2008; 116: 1179-82.

[14] Khanal RH, Shrestha SL. Development of procedures and assessment of environmental burden of disease of local levels due to major environmental risk factors; Nepal Health Research Council, Kathmandu, Nepal 2006.

[15] Sharma T, Rainey CM, Shrestha IL, *et al.* Roadside particulate levels at 30 locations in the Kathmandu valley, Nepal. Int J Environ Pollut 2002; 17: 293-305.

[16] Shrestha SL. Time series modeling of respiratory hospital admissions and geometrically weighted distributed lag effects from ambient particulate air pollution within Kathmandu valley, Nepal. Environ Model Assess 2007; 12(3): 239-51.

[17] CEN, ENPHO. Health impacts of Kathmandu's air pollution. Clean energy Nepal, Environment and Public Health Organization, Kathmandu, Nepal 2003.

[18] DoHS. Annual report. Department of health services, Kathmandu, Nepal 2005.

[19] ESPS. Ambient air quality monitoring in Kathamndu valley, yearly report, 2004, Ministry of population and environment, Kathmandu, Nepal 2005.

[20] Mc Cullagh P, Nelder JA. Generalized linear models, 2nd ed. Chapman and Hall, Inc. New York, USA 1989.

[21] Cameron AC, Trivedi PK. Regression analysis of count data; Cambridge University Press, UK 1998.

[22] Chow GC. Econometrics, Int. ed. Mc Graw-Hill, Inc, Singapore 1983.

[23] Hastie TJ, Tibsirani RJ. Generalized additive models; Chapman and Hall/CRC, USA 1990.

[24] Montgomery DC. Introduction to linear regression analysis, 3rd ed. John Wiley & Sons, Inc, Singapore 2003.

[25] Shrestha S. Statistical methods for linking health effects to air pollution. Lap Lambert Academic Publishing, Germany 2010.

Received: September 7, 2011 Revised: December 26, 2011 Accepted: March 12, 2012