American Journal of Epidemiology and Infectious Disease, 2016, Vol. 4, No. 2, 22-33
Available online at http://pubs.sciepub.com/ajeid/4/2/2

DOI:10.12691/ajeid-4-2-2

Forecasting Based On a SARIMA Model of Urban Malaria for Kolkata

Krishnendra S. Ganguly[1], Soumita Modak[2], Asis K. Chattopadhyay[2], Krishna S. Ganguly[3,*], Tapan K. Mukherjee[3], Ambar Dutta[1], Debashis Biswas[3]

[1]Computer Science Department, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
[2]Department of Statistics, University of Calcutta, Kolkata, West Bengal 700019, India
[3]Health Department, The Kolkata Municipal Corporation, Kolkata, West Bengal 700013, India
*Corresponding author: dr.k.s.ganguly@gmail.com

Abstract In India Urban Malaria is considered to be the one of the most widespread vector-borne diseases taking lives of many people including children. Kolkata is one of the Metropolitan cities in India where the seasonal effect of malaria is very marked. In the present work attempts have been made to study temporal variation of urban malaria incidences using time series model on the basis of a large population survey conducted by the Kolkata Municipal Corporation. It is found that the proposed time series (SARIMA) model can be used very successfully for prediction purposes.

Keywords: *spatio-temporal variation, time series model, sarima model, urban malaria*

Cite This Article: Krishnendra S. Ganguly, Soumita Modak, Asis K. Chattopadhyay, Krishna S. Ganguly, Tapan K. Mukherjee, Ambar Dutta, and Debashis Biswas, "Forecasting Based On a SARIMA Model of Urban Malaria for Kolkata." *American Journal of Epidemiology and Infectious Disease*, vol. 4, no. 2 (2016): 22-33. doi: 10.12691/ajeid-4-2-2.

1. Introduction

According to WHO, in the South-East Asian Region, amongst around 1.4 billion people living in 11 countries (land area 8466600 sq. km., i.e., 6% of global area), 1.2 billion are exposed to the risk of malaria, most of whom live in India. [1] However, South-East Asia contributed to only 2.5 million cases to the global burden of malaria. Of this, India alone contributed 76% of the total cases. Taking into account clinical episodes, it has now been estimated with the help of epidemiologic models and geographical & demographic data that Plasmodium falciparum [PF] estimates outside Africa, especially in South-East Asia, are 200% higher than reported by the World Health Organization (i.e., 118.94 million of global estimates of 515 million cases). [2] In addition to this, burden of P. vivax [PV] malaria in the world has been calculated at 71-80 million cases, of which Southeast Asia and Western pacific countries contributed 42 million cases [3].

Urban Malaria is a relatively new term associated with the Malaria cases in urban settings. They have several important characteristics in comparison to Rural Malaria cases, particularly when those are considered in respect of their control programme. In India by 1970s, incidence of Rural malaria came down drastically i.e. 0.1 to 0.15 million cases per year but the urban malaria is found to have ascending trend. Malaria in urban areas is contributed by large scale rural-urban migrations triggered by urban "push" (for earning livelihood in suburban and rural areas) and urban "pull" (for availing both Medicare/ educational opportunities in urban areas) phenomenon. Demographic and societal changes, unplanned urbanization, completion of various projects in total disregard of health impact assessment and incorporation of non-eco-friendly technologies, etc., all contributed to increased potentials for breeding of vectors of malaria and other diseases. Insufficient capacities of the civic bodies to deal with pure water-supply to every household, regular and hygienic disposal of sewage &/or solid-wastes, etc. led to an all-round disruptions. Intermittent water-supply led to increased water-storage practices, which resulted in extensive breeding of An. stephensi, the main vector of urban malaria in the study-area. The control of Urban malaria was thus considered as an important strategy for National Vector Borne Disease Control Programme (NVBDCP) in India.

Kolkata is one of the multi-racial, multi-lingual, metropolitan cities in India where malaria is an age-old public health problem [4]. The onus of preventing transmission of a mosquito-borne disease in Kolkata rests solely with the Health Department of The Kolkata Municipal Corporation (KMC). Since the end of 2010 the KMC authorities launched multi-faceted poly-pronged control of urban-malaria through multidisciplinary approach. Entomological studies show how breeding habitats are being modified over time-span [5,6,7]. Societal studies show how students are involved in the city of Kolkata to combat this vector-borne disease [4]. The present study is only a tiny off-shoot of the mainframe work-galore, where data have been utilized to have some idea about the pattern of occurrence of the disease of malaria in a geographical region of the city of Kolkata.

A trend-analysis, based on the seasonal effects of malaria incidences, keeping in consideration of factors of Epidemiological Triad (i.e., different co-variates), e.g., Economic, Geo-Climatic, Politico-Legal, Socio-Cultural, etc. as shown in Figure 1 [8,9,10,11,12], using time-series model, has been done to study temporal variation of urban-malaria incidences on the basis of a large survey conducted by the KMC, which guides the management aspect of cost-effectiveness & resource-allocation for disease-intervention. Surveillance through time-series data analysis can be used [13,14,15] for forecasting various warning signals of malaria and thus institution of intervention to control malaria will be much better before any epidemic takes place [16,17] Similar efforts were earlier done in Ethiopia, [18] Kenya, [19,20] Southern Asia [21] and China [22].

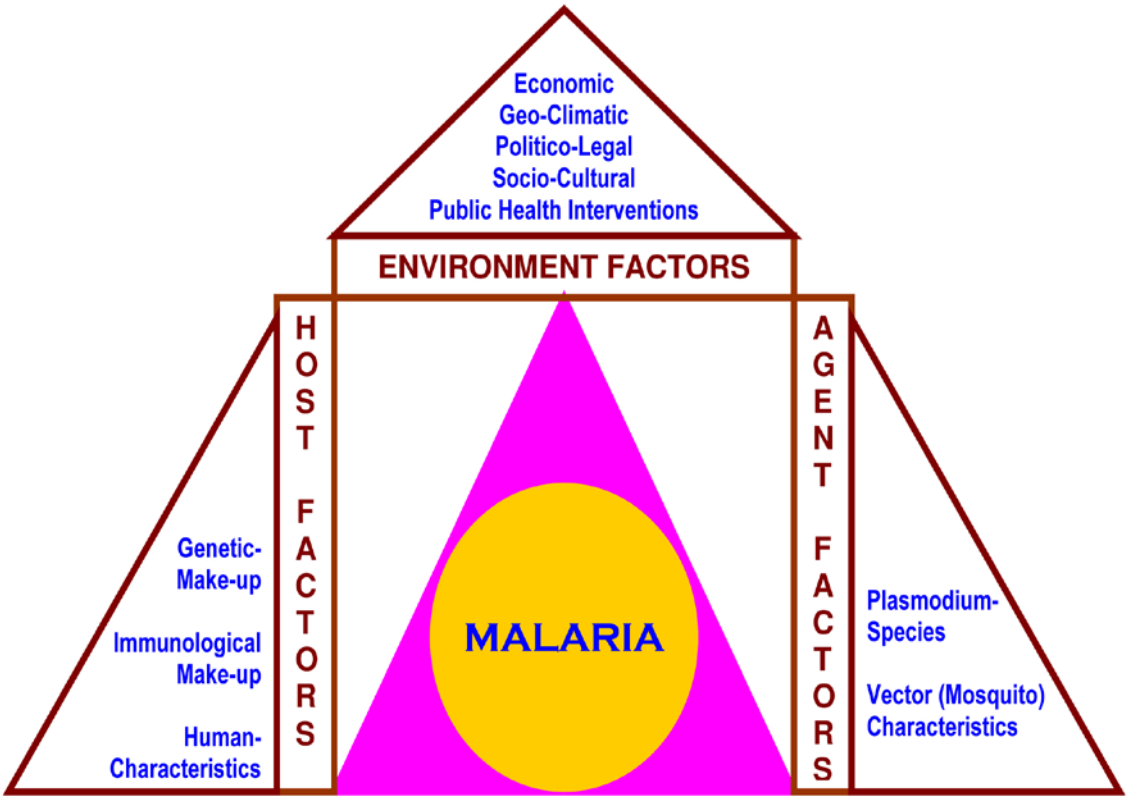

Figure 1. Factors of Epidemiological Triad (i.e., different co-variates), e.g., Economic, Geo-Climatic, Politico-Legal, Socio-Cultural, etc

2. Materials and Methods

Kolkata consists of 141 Wards (Figure 2). Out of 15 Boroughs of Kolkata, the administrative area of Borough-V (Figure 3) is an agglomeration of eleven (11) Wards of the city. Within the geographical boundary of Borough-V, this study has been designed on the population of Borough-V. More elaborately, it is a study based on the population of Ward 36, 37, 40, 41, 42, 43, 44, 45, 48, 49 and 50 of Kolkata. But specifically and more accurately, the study-design is based on the patients, through active and/or passive surveillance, who are being tested plus treated for febrile illnesses of malaria at malaria Clinics-cum-Treatment-Centres of 11 Wards of Borough-V of the KMC between January 2008 and December 2014. The study is planned based on secondary data over the first six-year period. It is worth mentioning here that Borough-V is chosen, because Borough-V is most infamous for its malaria-prone nature for years together.

Blood-slides are collected from all febrile patients or suspected fever patients according to their age-group and sex. These blood-slides are labeled and stained with Giemsa stain and examined by laboratory-microscopist, for the presence of malarial parasite and the species type. Medical Officers of Wards and Laboratory Coordinator with Senior Microscopists monitor and supervise the works of microscopist and reliability of prepared slides and positive-cases. Blood Slides Examined (BSE) and Blood Slides Positive (BSP) Slides are cross-checked and reconfirmed by random selection of 10% of slides which were labeled negative. All data are recorded and tabulated according to the months and Wards in Kolkata, which includes Annual Blood-slide Examination Rate (ABER =all slides x 100 / population), Slide Positivity Rate (SPR =total positive slides x 100 / total slides), species differentiation and Plasmodium falciparum Ratio (PFR% =total falciparum-malaria x 100 / total positive slides) and Annual Parasite Incidence (API =total positive cases x 100 /population).

WHO defines Slide Positivity Rate (SPR) as a Ratio between the number of laboratory-confirmed malaria cases per 100 clinically suspected cases, which indicates how good is detection of patients with malaria (i.e., increasing Rate indicating beginning of a malaria epidemic or a seasonal outbreak). Proportion of fever caused by malaria is almost a similar Ratio where number of Fever cases with confirmed parasitaemia divided by total number of fever, suggests how important a health priority is malaria. SPR is considered as a substitute to estimate temporal changes in malaria-incidence. SPR, an easily-available and inexpensive way to know malaria-burden in a

population utilizing health-care facilities, is more accurate as it only considers laboratory-confirmed malaria-cases. SPR is used in cross-sectional studies to know malaria endemicity [23,24] and malaria-control interventions. [25,26] But there is little work to know quantitative relationship between the metrics of temporal SPR-changes and malaria-incidence changes using empiric data. In a study conducted by Trevor P Jensen, et al in Africa [27] compared the results and showed observed changes in the incidence of malaria with changes in incidence of malaria estimated from the SPR and concluded SPR as a surrogate measure of malaria- burden. After interventions to control malaria, one should follow secular trends in malaria incidence based on surveillance data like SPR.

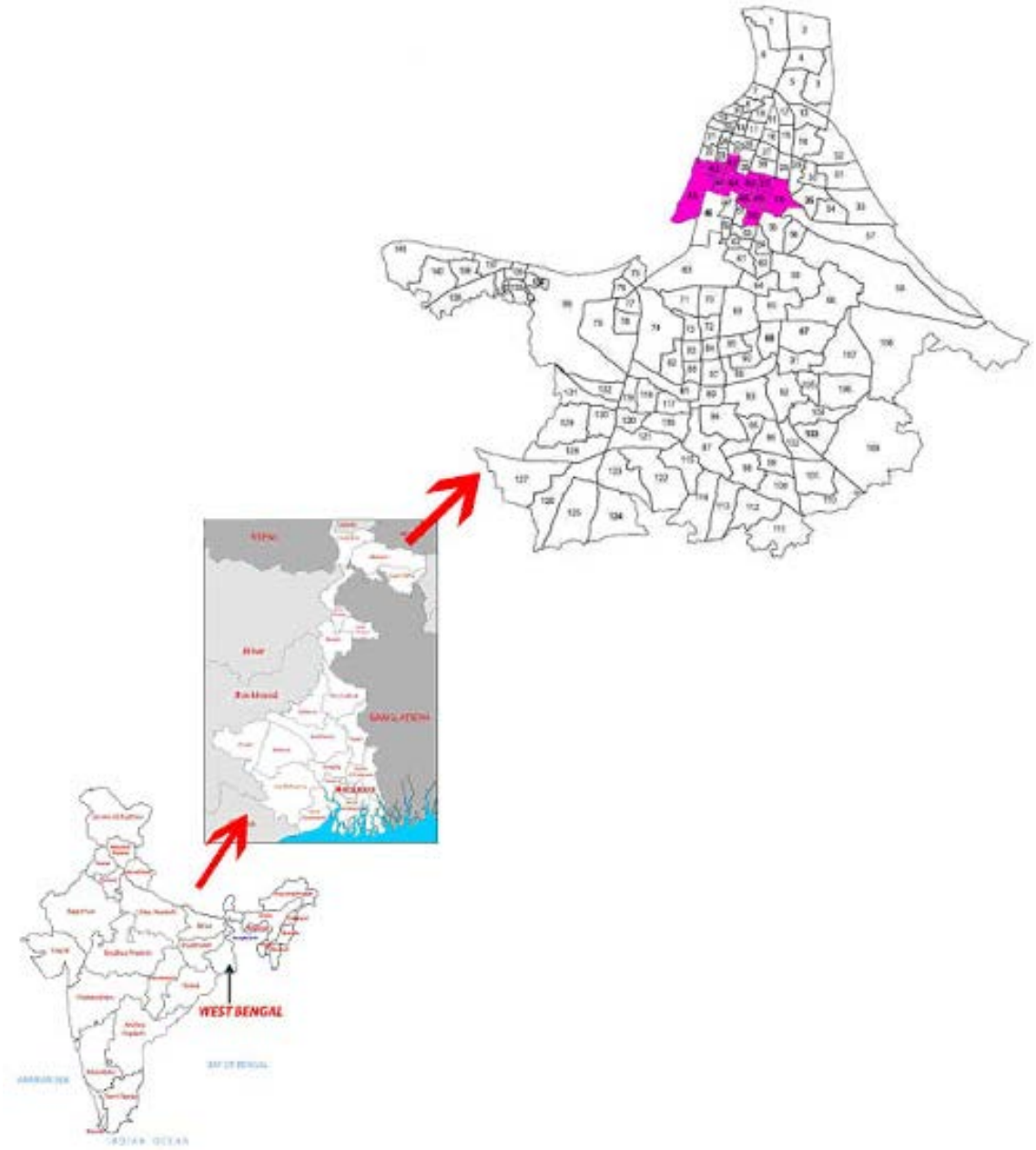

Figure 2. Kolkata in West Bengal in India

SPR has some advantages as a surrogate measure to know changes in malaria-incidence, as it uses normal HMIS data, collected at a nominal cost of measuring actual malaria-incidence. Though SPR is a useful measure of malaria-control interventions, any alteration in SPR is not corresponding to a proportional or linear change in the incidence of malaria, because as fever is used as the criterion for laboratory testing for SPR, a change in the incidence of non-malaria fevers can result in a change in SPR that does not reflect the true change in the incidence of malaria. Thus, when SPR is used to estimate relative changes in the malaria-incidence over time, it cannot estimate the actual malaria- incidence in a target population.

Additional factors, such as correct and quality laboratory diagnosis, age and sex of patient, seasonality of climatic variables can affect of malaria-incidence, which may maximize or suppress the impact of interventions. In real world, these factors affect the ability of SPR to estimate surveillance-effects on modifying actual malaria-incidences. On the contrary, most malaria surveillance systems gather only a subset of fever-cases in a target population, or are limited by incomplete and/or inaccurate laboratory testing.

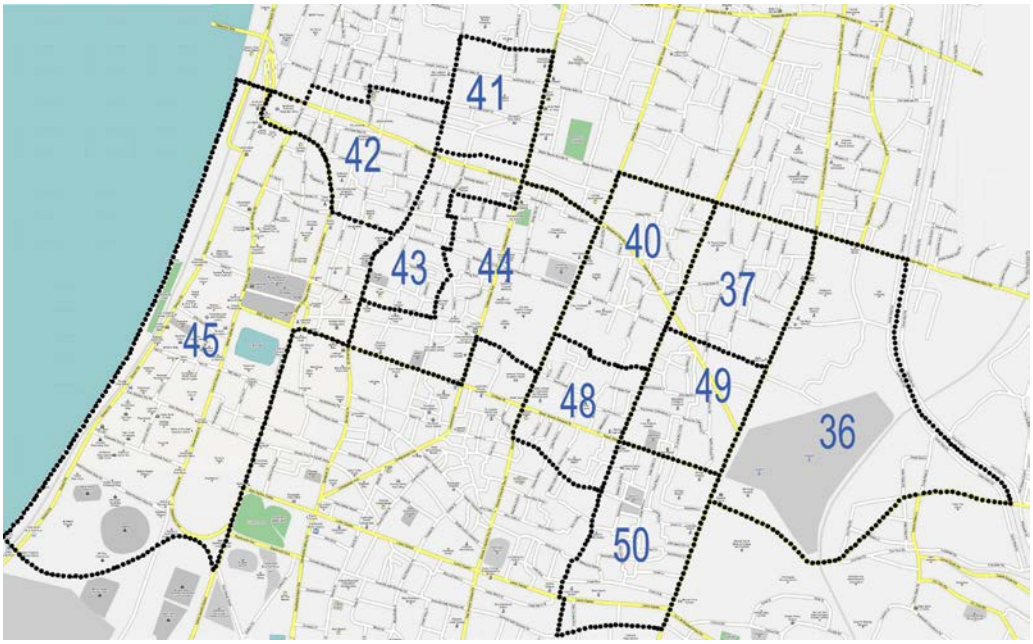

Figure 3. Borough-V of Kolkata

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. Your paper is one part of the entire proceedings, not an independent document. Please do not revise any of the current designations.

3. Results

3.1. Exploratory Data Analysis

At the initial stage of this longitudinal study, a descriptive analysis of malaria Positivity study has been conducted Year-wise and Ward-wise. Crude malaria Slide Positivity Rates (SPR) for each Ward are calculated by dividing the observed number of slide-positive malaria cases across 2008-2013 by the number of BSE for malaria. Table 1 depicts Microscopic Blood-Slide Test Results in Borough-V from 2008-2013, while Table 2 gives comparison of 2008-2013 parameters of ABER, SPR, PFR and API. Table 1 shows there is gradual rise of PV and PF amongst BSP slides from 2008 to 2010, but after several interventions from he part of the KMC there is gradual fall in BSP from 2011 to 2013. Simultaneously there is similar rise and fall of both PV and PF, but PF-incidences remarkably got reduced. Pair-wise Pearson cross-correlations for each month (January-December) are calculated to investigate temporal correlation in the malaria positive data, separately for P. vivax or P. falciparum cases, over the temporal span of years for 2008-2013, in cases of Borough V.

Table 1. Microscopic Blood-Slide Test Results in Borough–V from 2008–2013

Years	BSE	BSP	PV	PF
2008	35305	6830	5764	1054
2009	41248	12672	10139	2515
2010	42878	14532	11722	2756
2011	25606	6622	5677	934
2012	32420	5915	5058	849
2013	19307	3072	2769	299

Table 2. Comparison of six years parameters

Years	ABER (%)	SPR (%)	PFR (%)	API (%)
2008	13.37	19.35	15.43	2.59
2009	15.62	30.72	19.85	4.80
2010	16.23	33.89	18.97	5.50
2011	9.69	25.86	14.10	2.51
2012	12.27	18.24	14.35	2.24
2013	7.31	15.91	9.73	1.16
6 Year-Average	12.41	23.99	15.40	3.13

Table 3. Monthly Slide Positivity Rate (X_t), of Borough–V Malaria Cases in Kolkata from January 2008 to December 2013

YEAR	January	February	March	April	May	June	July	August	September	October	November	December
2008	0.076449	0.069351	0.08642	0.094378	0.140588	0.140891	0.2	0.235177	0.196334	0.229087	0.212941	0.167568
2009	0.113285	0.079918	0.121279	0.237934	0.155544	0.302343	0.33255	0.358866	0.332957	0.333333	0.335996	0.2715
2010	0.18481	0.145095	0.241769	0.309524	0.30504	0.281837	0.331989	0.423395	0.375956	0.347832	0.320859	0.288827
2011	0.19709	0.191242	0.252895	0.365642	0.379242	0.263343	0.317151	0.339789	0.236282	0.222278	0.183855	0.178826
2012	0.086735	0.095395	0.149502	0.162592	0.157895	0.104704	0.099198	0.196131	0.217778	0.185358	0.214412	0.143013
2013	0.104294	0.079572	0.096216	0.132444	0.180907	0.196667	0.194006	0.186756	0.19569	0.156463	0.134586	0.070892

In Borough-V of Kolkata, in six years (2008-2013) total slides examined are 196764 in all Wards where total positive malarial parasites are 49643, out of which P. vivax is found in 41129 and P. falciparum in 8333 cases, as depicted in Table 1. Table 3 shows Monthly Slide Positivity Rate (X_t), of Borough-V malaria Cases in Kolkata from January 2008 to December 2013. Based on total slides examined and positive cases for malarial parasites, data are subjected to determine Annual Blood Examination Rate (ABER), Slide Positivity Rate (SPR),

Plasmodium Falciparum Ratio (PFR%) and Annual Parasite Incidence (API), which are shown in Table 2. In six years, ABER has an average of 12.41%. During this period, SPR is 23.99%. Ulike SPR, PFR increased from 2008 to 2011, but there is gradual fall from 2012 to 2013. PFR ranges from 9.73% to 19.85% with an average of 15.40%. Annual Parasite Incidence was being 1.16 to 5.50 with an average of 3.13 during this period. The month wise data revealed that SPR, PFR and API increased in post-monsoon season maximally between August-September and November-December.

GIS Maps of crude SPRs for the 11 Wards of Borough-V for 2008-2013 are presented in Figure 4. Maps show that low SPR Ward are mainly located in the central region of Borough-V, i.e., Ward 44. It is observed from the GIS maps that there is a noticeable change of SPR over the period from 2008 to 2013 for each Ward of Borough-V. Here, the R-G-B Color Gradient Scale is used where Blue indicates the lowest range of SPR values and Red indicates the Highest SPR values in Borough-V. Now from the GIS maps over the period, we can tell nature of changes of SPR-values is different for each Ward. Some Wards, e.g., 36, 41, 43, 44 & 48 do not ever cross the values of 35% (Green), whereas Ward 50 only goes beyond the value 65% (Red). It can also be seen from the GIS Maps that for every Ward there is an increase of SPR from year 2008 to 2010 but from 2011 to 2013 the SPR values decrease; SPR for some wards like 36, 37, 40, 44, 48, 49 and 50 come below 15% (Blue zone) during this period 2011-2013. It is also noticeable that the SPR values for Ward 50 both increases and also decreases sharply within 2008-2013.

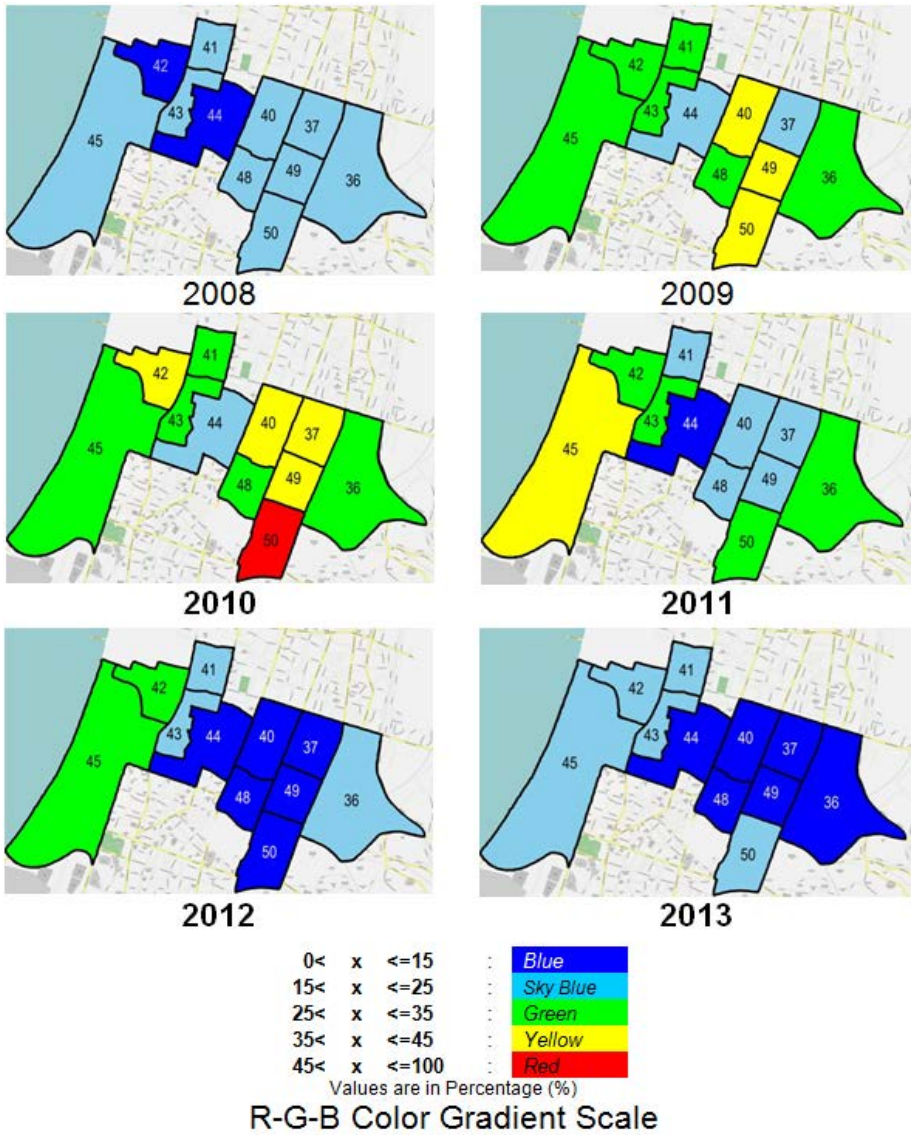

Figure 4. GIS Maps of crude SPRs for the 11wards of Borough-V for 2008-2013

3.2. Time Series Analysis

A time series analysis is done considering the temporal data of SPR series 2.1.1. Front Matter

3.2.1. Time Series Plot

The first and foremost tool to analyse a time series is to look at the graph of the observed values of study variable(SPR) over time points. Time series under study(SPR series) is a monthly series over six years. Time series plot (Figure 5) indicates a trend, upward from the beginning to late 2010 and then downward towards the end. Figure 5 also shows up a seasonality pattern of period one year (characterized by a peak in summer and rains (July-September, peak month August) and a trough in winter (December-January), a smaller peak preceding the

summer peak is also apparent, occurring in November of most years) and the magnitude of the seasonal variation increases at the same sort of rate as the yearly mean levels, indicating a multiplicative seasonal model is plausible.

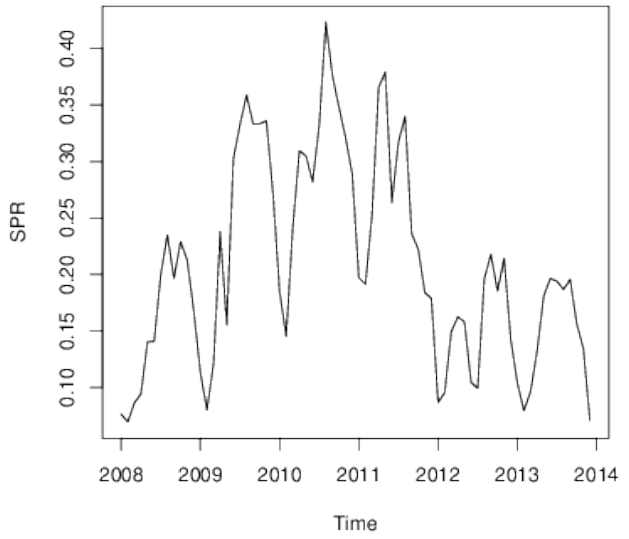

Figure 5. Time Series Plot

3.2.2. Time Series Plot

Time series under study is decomposed into seasonal, trend and irregular components using moving averages (Kendall and Stuart, 1983). The multiplicative model used is:

$$Y_t = T_t * S_t * e_t$$

where,

Y_t : The original series at time t.

T_t : The trend component at time t.

S_t : The seasonal component at time t.

e_t : The remainder part at time t.

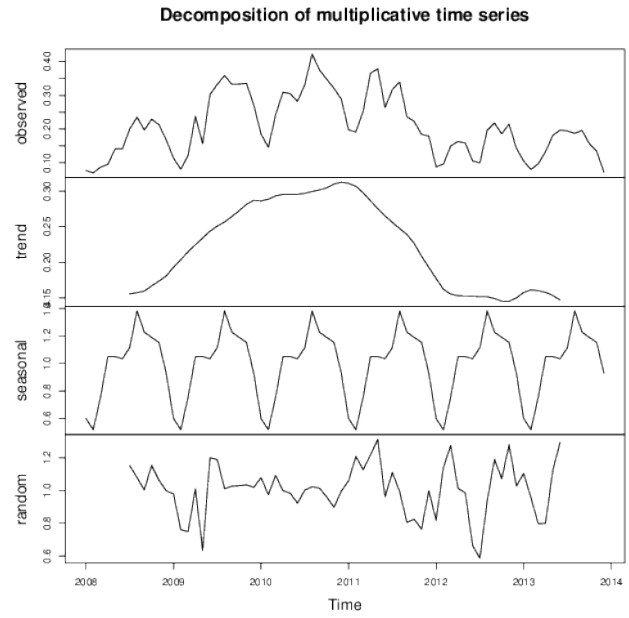

Figure 6. Components of Time Series

The trend component is first determined using a moving average (a symmetric window with equal weights is used), and is removed from the time series. Then, the seasonal figure is computed by averaging, for each time unit, over all periods. The seasonal figure is then centered. Finally, the error component is determined by removing trend and seasonal figure (recycled as needed) from the original time series.

Decomposition of SPR series (Figure 6) clearly shows up a trend, upward from the beginning of 2008 up to late 2010 and then downward towards the end of 2013, and seasonality of period 12 months.

3.2.3. Correlogram

For a given time series of length N, $Y_1, Y_2, \ldots, Y_N$ the sample version of the autocorrelation coefficient (ac.f.) at lag k is given by,

$$k = \frac{\sum_{t=1}^{N-k} (Y_k - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N} (Y_t - \bar{Y})^2} \quad (1)$$

for $k = 1, 2, \ldots, M$, where $M < N$. The plot of sample autocorrelation coefficients r_k against the lag k for $k = 0, 1, \ldots, M$, where M is usually much less than N is called a correlogram. The correlogram is an useful tool for revealing the structure of time series by interpreting a set of autocorrelation coefficients.

Figure 7 shows the correlogram of the original series along with the dotted lines at $\pm\frac{2}{\sqrt{N}} = 0.236$ (as in our sample $N = 72$). Values outside these lines are said to be significantly different from zero [28].

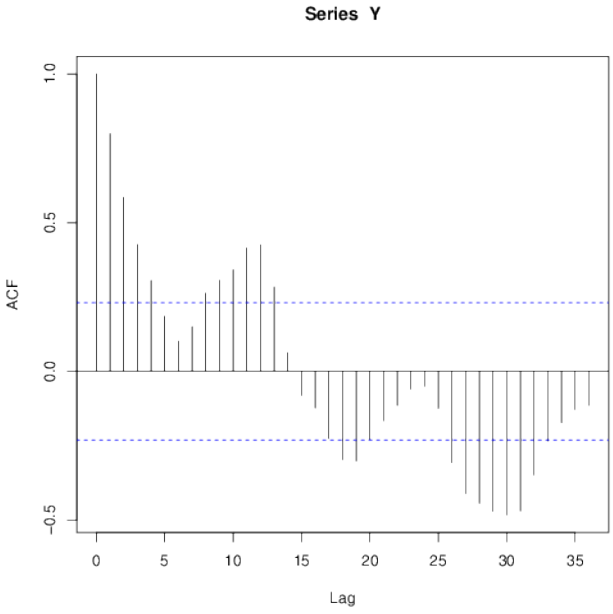

Figure 7. The correlogram of of Y_t

Correlogram (Figure 7) clearly shows the existing seasonality pattern in the SPR series. The values of r_k did not come down to zero except for large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because of the trend.

3.3. Model

A model to describe a time series consisting trend and seasonality is required. Our model is a general multiplicative seasonal ARIMA (SARIMA) model[29] as

$$\varphi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t$$

where B denotes the backward shift operator, $\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p, P, q, Q, respectively, Z_t denotes a purely random process and

$$W_t = \nabla^d \nabla_s^D Y_t$$

denotes the differenced series, formed from the original series Y_t by appropriate differencing to remove non-stationary terms. The above model is called a SARIMA model of order $(p,d,q)\times(P,D,Q)s$, where

p: Autoregressive (AR) order
d: difference order
q: Moving-Average (MA) order
P: Seasonal Autoregressive (SAR) order
D: seasonal difference
Q: Seasonal Moving-Average (SMA) order
s: seasonal period

3.3.1. Model Identification

To choose the values of parameters in SARIMA model for time series under study, first the series has to made stationary. Non-stationary components trend and seasonality are to be removed. Here $s = 12$, to remove seasonality first we take s^{th} order difference on the SPR series Y_t and obtained $\nabla_{12} Y_t = \tilde{Y}_t$.

Figure 8 shows seasonality is removed and strong trend is present in Y_t series. To remove trend we take first order difference on $\tilde{Y}_t$ and obtained

$$W_t = \nabla\nabla_{12} Y_t = \nabla_{12} Y_t - \nabla_{12} Y_{t-1}$$
$$= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}).$$

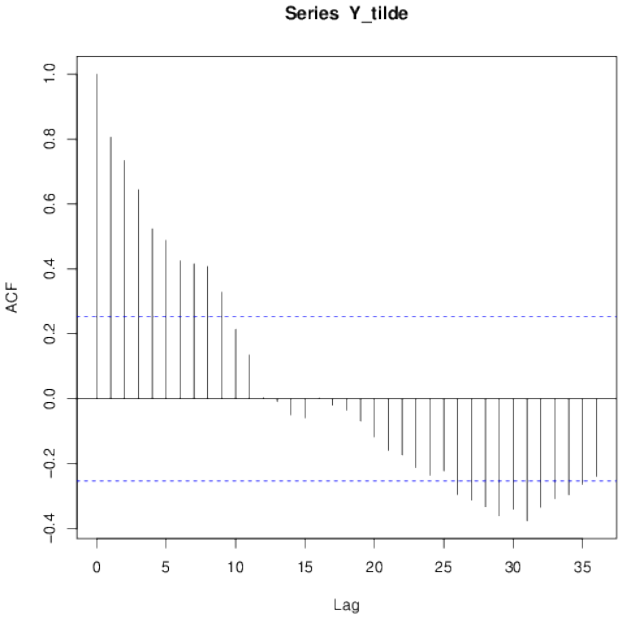

Figure 8. The correlogram of $\tilde{Y}_t$

The difference parameters $d = 1$ and $D = 1$ have already been chosen by the choice of differencing operator so as to make the differenced series stationary.

The seasonal parameters P and Q are determined by the values of the sample autocorrelation function and partial autocorrelation function of W_t at lags 12,24,36,… ,etc. For our data consideration up to lag 36 is enough. The ac.f. plot (Figure 9) and the partial ac.f. plot (Figure 10) show that the values are 'significant' at lag 12 but 'insignificant' at lags 24 and 36. It indicates presence of one seasonal AR term and one seasonal MA term in our model. Thus $P = 1$ and $Q = 1$ are chosen. The values of the non-seasonal parameters p and q are determined by the first few values of the sample ac.f. and partial ac.f. Figure 9 and Figure 10 show that the only 'significant' values are at lag 1. It indicates presence of one AR term and one MA term in our model. Thus $p = 1$ and $q = 1$ are taken.

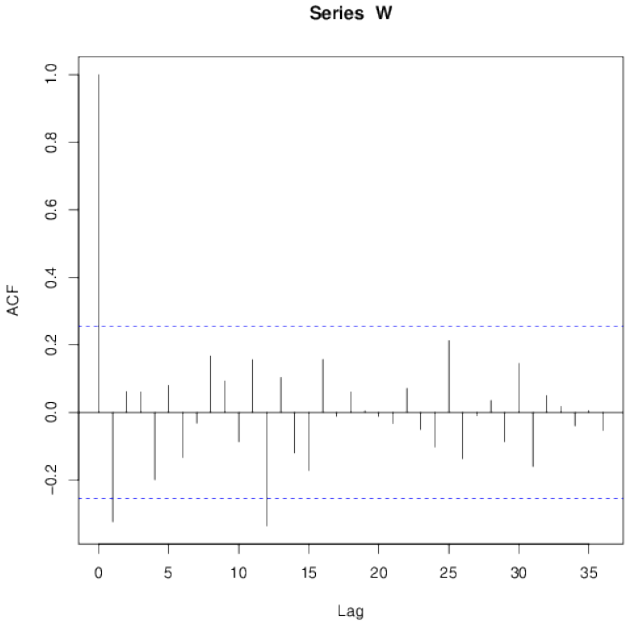

Figure 9. The correlogram of W_t

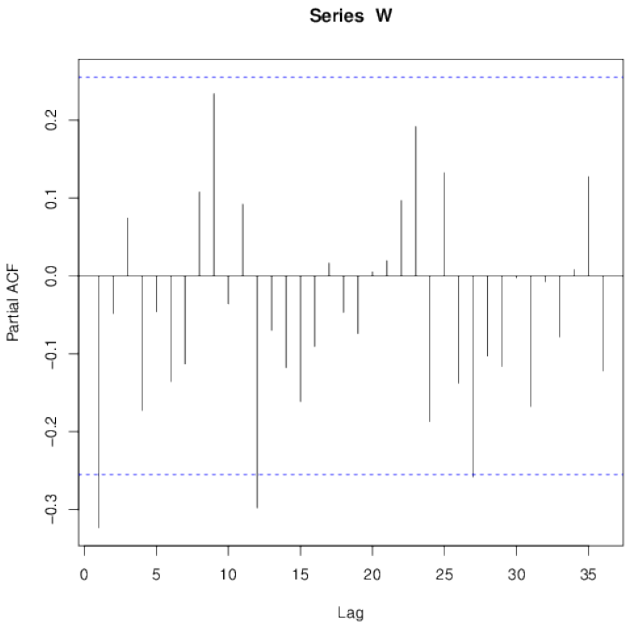

Figure 10. The partial ac.f. plot of W_t

So, the chosen seasonal ARIMA model with $p=1,=d=1,q=1$ and $P=1,D=1,Q=1$ is given by

$$(1+\varphi B)(1+\Phi B^{12})W_t=(1+\theta B)(1+\Theta B^{12})Z_t$$

where $W_t=\nabla\nabla_{12}Y_t$ and $\varphi,\Phi,\theta,\Theta$ are constant parameters.

The equation (2) boils down to

$$\begin{aligned}X_t = {} & (1-\varphi)X_{t-1}+\varphi X_{t-2}+(1-\Phi)X_{t-12}\\ & +(\Phi-\varphi\Phi+\varphi-1)X_{t-13}+\varphi(\Phi-1)X_{t-14}\\ & +\Phi X_{t-24}+\Phi(\varphi-1)X_{t-25}-\varphi\Phi X_{t-26}\\ & +Z_t+\theta Z_{t-1}+\Theta Z_{t-12}+\theta\Theta Z_{t-13}.\end{aligned}$$

3.3.2. Model Fitting

Parameters in the chosen SARIMA model are estimated using maximum likelihood method, where conditional-sum-of-squares is used to find starting values of the parameters. Fitted SPR series model is obtained by placing the fitted parameter values in equation 3. Fitted parameter values for equation 3 are shown in Table 4.

Table 4. Maximum Likelihood Estimates of Model Parameters

	ϕ	θ	Φ	Θ
estimate	0.1364	-0.4300	0.1409	-1.0000
Standard error	0.5217	0.4865	0.1567	0.2654

3.3.3. Diagnostic Checking

A time series is said to be completely random if it consists of a series of independent observations having the same distribution. Then, for large N, we expect to find that $r_k \cong 0$ for all non-zero values of k, where r_k is as given in 1. In fact, for a random time series, r_k is approximately N(0, $\frac{1}{N}$). If the chosen model adequately describes the data and the model is well estimated, then we expect the residuals to be random time series, where,

$$residual=observation-fitted.$$

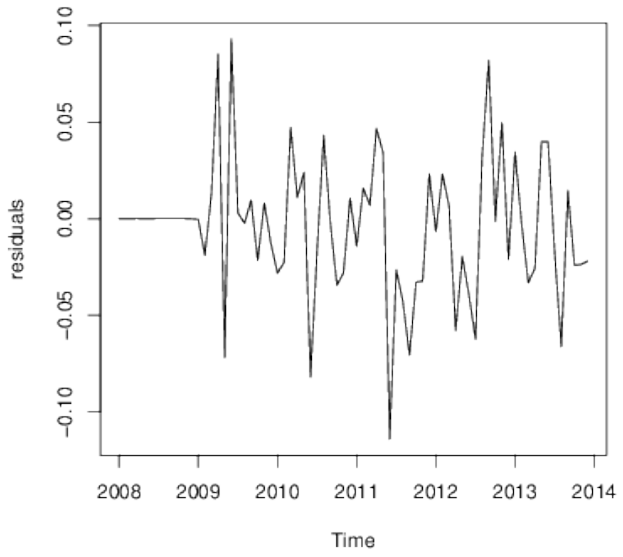

Figure 11. Time Series plot of Residuals

Figure 11 shows randomness of residuals. To get a closer look at the residuals, their correlogram (Figure 12) is drawn. Figure 12 shows insignificancy of all the sample autocorrelation functions of residuals (except one slightly significant at lag 15, a physically non-interpretable lag, and hence ignorable), indicating that the residuals are random time series i.e. the fitted model describes the observed SPR series well. Further a non-parametric test, Kolmogorov-Smirnov (K-S) test is performed to test the null hypothesis that $r_k \sim N(0,\frac{1}{N})$ for the residual series. Table 5 shows p-value of K-S test is greater than 0.05 (level of significance) and hence at 5% level of significance the null hypothesis is accepted. Test based on the modified Ljung-Box-Pierce statistic [30] for examining the null hypothesis that the fitted model is appropriate, is performed. Table 6 shows all the p-values are greater than 0.05 (level of significance). Hence, at 5% level of significance the null hypothesis is accepted and adequacy of the fitted model is concluded.

Table 5. Test for normality of sample ac.f. of residuals

K-S test statistic	p-value
0.1218	0.6452

Table 6. Test for adequacy of fitted model

Lag	p-value for Ljung-Box test
1	0.933
2	0.903
3	0.974
4	0.990
5	0.998
6	0.714
7	0.788
8	0.652
9	0.505
10	0.594
11	0.674
12	0.741
13	0.792
14	0.731
15	0.531
16	0.583
17	0.627
18	0.692
19	0.724
20	0.770
21	0.763
22	0.793
23	0.820
24	0.825
25	0.690
26	0.691
27	0.590
28	0.641
29	0.691
30	0.649
31	0.550
32	0.599
33	0.637
34	0.679
35	0.661
36	0.090

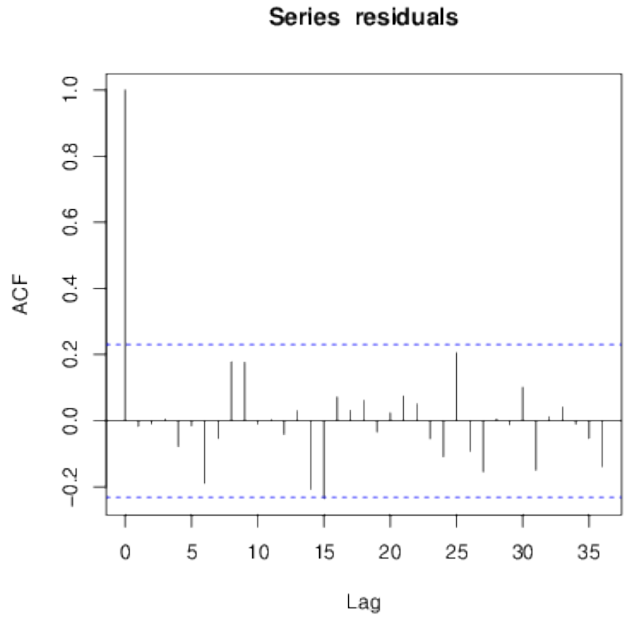

Figure 12. The correlogram of residuals

Diagnostic checking suggested our fitted model is good enough to describe the SPR series, so further improvement on the chosen model is not required.

3.4. Forecasting

Prediction of the SPR series for the year 2014 is carried out (Table 7) using equation 3 and the fitted values of model parameters from Table 4. Figure 13 shows a graphical presentation of forecast for the year 2014, based on our fitted SARIMA model.

Figure 14 shows the observed SPR series for the year 2014 exactly lies within 80% confidence intervals and is close enough to our point forecast for the year 2014. It indicates the success of our proposed forecast model for the underlying time series.

Figure 15 shows the correlogram of departure of observed SPR series from our forecast SPR series for the year 2014. Clearly, insignificance of sample autocorrelation functions at all lags is indicating that the departures are random time series i.e. the forecast model describes the observed SPR series well.

Table 7. Forecast for the year 2014

Time	Point Forecast	80% Confidence Interval for Forecast	95% Confidence Interval for Forecast
Jan 2014	0.012948	(-0.043790, 0.069686)	(-0.073826, 0.099721)
Feb 2014	-0.004012	(-0.073218, 0.065194)	(-0.109853, 0.101829)
Mar 2014	0.038264	(-0.040391, 0.116919)	(-0.082028, 0.158556)
Apr 2014	0.092490	(0.005537, 0.179444)	(-0.040494, 0.225474)
May 2014	0.103698	(0.009188, 0.198208)	(-0.040843, 0.248239)
Jun 2014	0.102282	(0.000778, 0.203786)	(-0.052955, 0.257518)
Jul 2014	0.128294	(0.020249, 0.236339)	(-0.036947, 0.293535)
Aug 2014	0.163900	(0.049687, 0.278112)	(-0.010774, 0.338573)
Sep 2014	0.139362	(0.019298, 0.259426)	(-0.044260, 0.322983)
Oct 2014	0.122742	(-0.002900, 0.248385)	(-0.069412, 0.314897)
Nov 2014	0.109085	(-0.021902, 0.240071)	(-0.091242, 0.309411)
Dec 2014	0.059436	(-0.076697, 0.195570)	(-0.148762, 0.267635)

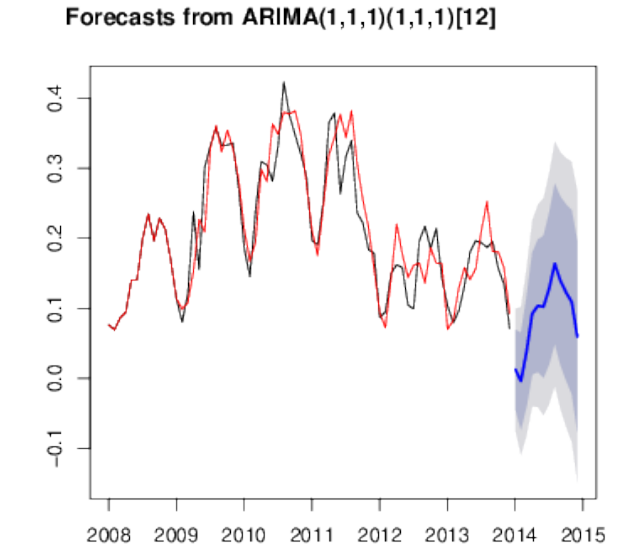

Figure 13. Black line: Observed Time Series, Red line: Fitted Time Series, Blue line: Forecast Time Series, Dark Grey region: 80% Confidence Interval for Forecast, Light Grey region: 95% Confidence Interval for Forecast.

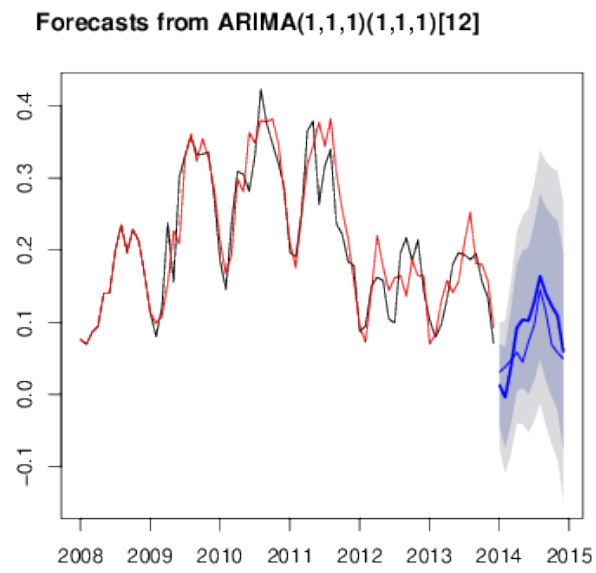

Figure 14. A thin blue line is added to Fig.12, showing monthly observed time series data for the year 2014

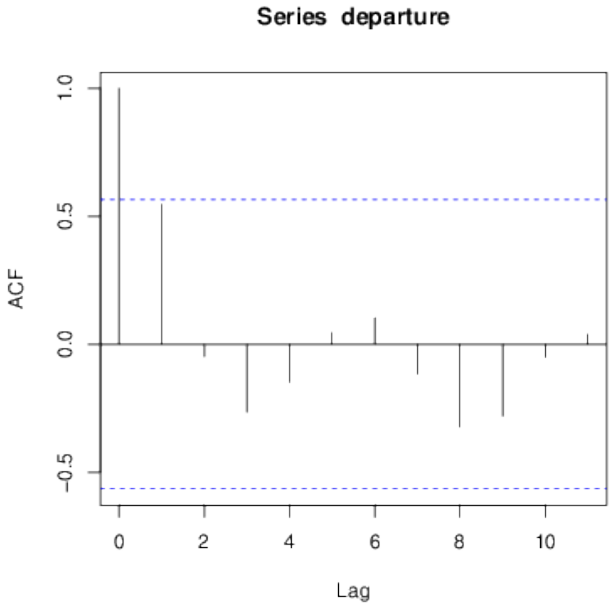

Figure 15. The correlogram of departure of observed time series from our forecast time series for the year 2014

Kolmogorov-Smirnov (K-S) test is performed to test that the sample ac.f. of the departure series for the year 2014 approximately follow N(0, $\frac{1}{N}$). Table 8 shows the p-value > 0.05, indicating the forecast model describes the observed SPR series well.

Table 8. Test for normality of sample ac.f. of departures of Observed time-series from Forecast time-series in 2014

K-S test statistic	p-value
0.2515	0.4346

P-values for the modified Ljung-Box-Pierce statistic are computed (Table 9) on the departure of observed SPR series from our forecast SPR series for the year 2014. All the p-values are greater than 0.05, indicating adequacy of our forecast model for the year 2014.

Table 9. Test for adequacy of forecast model

Lag	p-value for Ljung-Box test
1	0.076
2	0.188
3	0.218
4	0.344
5	0.422
6	0.527
7	0.521
8	0.240
9	0.155
10	0.199
11	0.264

4. Discussions & Conclusions

For effective malaria-control, knowledge of disease-burden and trends, plus measuring intervention-effects [31] is an effective tool. Here we fit a SARIMA time series model on malaria SPR data collected at the KMC Ward-level malaria Clinics (cum-Treatment-Centres) over a six-year period from January 2008 to December 2013. On the basis of our fitted model we forecasted malaria SPR for the year 2014 for Borough-V of Kolkata-population. Before discussions on this issue, we need highlighting another unique feature of this model, which is based on SPR, rather than malaria Incidence Rate, as is generally used in many concurrent studies.

WHO defined malaria Incidence Rate as Number of New malaria Cases out of total population per week in each area, which points out whether malaria is increasing and necessary actions to be taken when malaria increases. As incidence of malaria is defined as the number of confirmed malaria cases per person-time, [32] to monitor and to evaluate a general working definition of a malaria-case may be taken as “fever with malarial parasitaemia”, that normally demands all patients who require anti-malarial treatment. [32] To estimate malaria incidence, longitudinal studies in defined populations of a defined geographical area, involving all suspected malaria-cases (from all fever cases) and subjecting them to a diagnostic test (having high sensitivity and specificity) are necessary, which in reality demands considerable resources. It is rarely followed in routine malaria surveillance programme.

Malaria incidence is usually estimated based on the number of reported malaria-cases recorded in HMIS (Health Management Information System) of a country or its part. To determine relationships between malaria-incidence and control-interventions, using HMIS-data, in reality, entail different risk-factors leading to high bias and confounding factors, e.g., a significant lag-time before their availability, incomplete reporting, temporal variations in reporting, varied utilization of health-care services, lack of a proper denominator (for variation in population-size), and genuineness and accuracy of laboratory-confirmation of these data. [33] But statistical procedures (in spite of limitations of HMIS data and of diagnostic testing) can better accuracy-factors of derived estimates of malaria-incidence.

The intensity of malaria transmission can be estimated using different indicators such as ABER, API, SPR and Incidence of malaria [34,35,36,37,38]. Annual malaria incidence includes numbers of laboratory-confirmed malaria cases as numerator and local population as denominator. But as local population size may be under or overestimated because census is only carried out once 10 years in India. Huge population movement is common due to economic reasons in Kolkata, particularly in last decade. Thus, malaria incidence might be inaccurate due to limited health care resources [35] or wrong population-size [39]. It is important to estimate the burden of malaria accurately to plan public health interventions. SPR is used as a surrogate to measure Incidence of malaria [35,37,40,41], to define the level of malaria endemicity in any region [39], and to identify high malaria-risk areas [42]. This is one of the principal monitoring indicator in malaria control programme in Kolkata for several years [43,44], through the malaria annual reporting system. The changes in malaria incidences can be estimated also from the SPR trends. [35] Some studies have demonstrated that SPR has steadily decreased with the decline in malaria incidence, [36,40] while others found that the annual parasite index (API) increased, but SPR-increase remained somewhat slow at the same level.[36] In our study SPR-trend has been subjected to statistical analysis.

Now coming to the statistical analysis, we find that the magnitude of the seasonal variation increases at the same rate as the yearly mean-values, which indicates that a multiplicative seasonal model is appropriate. From the analyzed result, we find that between projected Model and fitted data in that Model, there are few apparent dissimilarities, viz, the usual bi-modal rise of malaria in 2008-2013 plot is modified to a uni-modal peak in 2014 (somewhat like 2010 data, but of much lowered SPR-value). The plateau-area of highest malaria-incidence, is constricted to a single peak of incidence. Instead of phased monthly rise with phased monthly fall of malaria-incidences, there is sharp rise and sharp fall in malaria-incidences in the study-population (somewhat like 2010 data). There is a high-degree seasonal pattern (mainly, climatological) in the data all throughout the period, but clearly there is a upward trend from 2008-2010 followed by a downward trend gradually since 2011 up to 2014, which may be as a result of different anti-malaria and anti-mosquito interventions adopted by the K.M.C., which influenced the Model. It may noted that

(a) Climatological factors are more or less same in a city but there is still intra-Ward variation of controlling of malaria over the study-period.
(b) Socio-economic factors are major differences amongst the Wards, which might be one prime reason for these differences, and so also the Politico-Legal Framework of the Ward.
(c) Minor Intra-Ward Spatial Variations are not taken into consideration in this study.
(d) Study-population, as was considered, did not change during these years from 2008 to 2014.
(e) Societal & political changes and vector-control interventions during these years have impacts on the result.

This study is an important step to attain a better forecasting of urban malaria through multivariate analysis as considered ahead.

Statement of Competing Interests

The authors have no competing interest.

Acknowledgements

The authors sincerely acknowledge the assistance, cooperation and encouragement of Sri Atin Ghosh, Hon'ble Member, Mayor-in-Council (Health), The Kolkata Municipal Corporation. The authors are indebted to Sri Khalil Ahmed, Municipal Commissioner and Dr. Manirul Islam Molla, Chief Municipal Health Officer, The K.M.C., for extending their help in all possible ways in pursuing the research and preparation of this paper.

References

[1] Kondrachine AV, Malaria in WHO Southeast Asia Region, *Indian J Malariol* 29:129-160., 1992.

[2] Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI, The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* 434: 214217. 2005.

[3] Mendis K, Sina BJ, Marchesini P, Carter R The neglected burden of Plasmodium Vivax malaria. *Am J Trop Med* Hyg 64:97-106, 2001.

[4] Hati AK, Mukherjee H, Chandra G, Bhattacharyya J, Chatterjee KK, Banerjee A, Biswas D, Halder S. Vector-borne diseases in urban community. *Your Health*, 40, 157-158. 1991

[5] Biswas D, Mandal B, Biswas B, Banerjee A, Mukherjee TK. Plying of speedboats along canals in the city of Kolkata, India, to prevent mosquito breeding. *Transactions of the Royal Society of Tropical Medicine and Hygiene*., 107, 147-151. 2013.

[6] Biswas D, Biswas B, Mandal B, Banerjee A, Mukherjee TK, Nandi J. Evaluating school students' perception about mosquitoes and mosquito-borne diseases in the city of Kolkata, India. *Dengue Bulletin*, 35, 223-230. 2011

[7] Biswas D, Biswas B, Mandal B, Banerjee A. A Note on Distribution of Breeding Sources of Aedes aegypti (Linnaeus) in the City of Kolkata, India, Following an Outbreak of Dengue during 2012. *Current Urban Studies*. 2, 57-61. 2014.

[8] Nihei N, Hashida Y, Kobayashi M, Ishii A: Analysis of malaria endemic areas on the Indochina Peninsula using remote sensing. *Jpn J Infect Dis*, 55:160-166. 2002.

[9] Klinkenberg E, Hoek W, Amerasinghe FP: A malaria risk analysis in an irrigated area in Sri Lanka. *Acta Trop*, 89:215-225. 2004

[10] Leonardo LR, Rivera PT, Crisostomo BA, Sarol JN, Bantayan NC, Tiu WU, Bergquist NR: A study of the environmental determinants of malaria and schistosomiasis in the Philippines using Remote Sensing and Geographic Information Systems. *Parassitologia*, 47:105-114. 2005

[11] Zhou G, Sirichaisinthop J, Sattabongkot J, Jones J, Bjornstad ON, Yan G, Cui L: Spatio-temporal distribution of Plasmodium falciparum and P. vivax malaria in Thailand. *Am J Trop Med Hyg*, 72:256-262. 2005.

[12] Childs DZ, Cattadori IM, Suwonkerd W, Prajakwong S, Boots M: Spatiotemporal patterns of malaria incidence in northern Thailand. *Trans R Soc Trop Med Hyg*, 100:623-631. 2006.

[13] Ceccato P, Connor SJ, Jeanne I, Thomson MC: Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. *Parassitologia*, 47:81-96. 2005.

[14] Rogers DJ, Randolph SE, Snow RW, Hay SI: Satellite imagery in the study and forecast of malaria. *Nature*, 415:710-715. 2002.

[15] Abeku TA, Hay SI, Ochola S, Langi P, Beard B, de Vlas SJ, Cox J: Malaria epidemic early warning and detection in African highlands. *Trends Parasitol*, 20:400-405. 2004.

[16] Thomson MC, Connor SJ: The development of Malaria Early Warning Systems for Africa. Trends Parasitol 2001, 17:438-445.

[17] Cox J, Abeku TA: Early warning systems for malaria in Africa: from blueprint to practice. *Trends Parasitol*, 23:243-246. 2007.

[18] Abeku TA, de Vlas SJ, Borsboom G, Teklehaimanot A, Kebede A, Olana D, van Oortmarssen GJ, Habbema JD: Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. *Trop Med Int Health*, 7:851-857. 2002.

[19] Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi I, Alipui N, Snow RW: Forecasting, warning, and detection of malaria epidemics: a case study. *Lancet*, 361:1705-1706. 2003.

[20] Thomson M, Indeje M, Connor S, Dilley M,Ward N: Malaria early warning in Kenya and seasonal climate forecasts. *Lancet*, 362:580. 2003.

[21] Bouma MJ, Kaay HJ: The El Nino Southern Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system for future epidemics? *Trop Med Int Health*, 1:8696. 1996.

[22] Archie CA Clements, Adrian G Barnett, Zhang W Cheng, Robert W Snow and Hom N Zhou: Space-time variation of malaria incidence in Yunnan province, China , *Malaria Journal*, 8:180. 2009.

[23] Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis*, 4(6):327- 336. 2004.

[24] Joshi PL, Chandra R, Bhattacharya M, Vaish HC: Validity of using slide positivity rate (SPR) in identification of high risk malarious segments in rural areas. *The Journal of communicable diseases*, 29(1):41-45. 1997.

[25] Ceesay SJ, Casals-Pascual C, Erskine J, Anya SE, Duah NO, Fulford AJ, Sesay SS, Abubakar I, Dunyo S, Sey O, et al.: Changes in malaria indices between 1999 and 2007 in The

Gambia: a retrospective analysis. *Lancet*, 372(9649):1545-1554. 2008.

[26] O'Meara WP, Bejon P, Mwangi TW, Okiro EA, Peshu N, Snow RW, Newton CR, Marsh K: Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet*, 372(9649): 1555-1562. 2008.

[27] Trevor P Jensen, Hasifa Bukirwa, Denise Njama-Meya, Damon Francis, Moses R Kamya, Philip J Rosenthal and Grant Dorsey : Use of the slide positivity rate to estimate changes in malaria incidence in a cohort of Ugandan children, *Malaria Journal*, 8:213. 2009.

[28] Chatfield C. The Analysis of Time Series: An Introduction (Fourth Edition ed.). New York, NY: *Chapman & Hall*, 1989.

[29] Box G.E.P., Jenkins G.M., Reinsel G. C. Time Series Analysis: Forecasting and Control, 4th Edition, 2008.

[30] Ljung G. M. and Box G. E. P. On a measure of lack of fit in time series models. *Biometrika*, 65: 297-303. 1978.

[31] Breman JG, Holloway CN: Malaria surveillance counts. *Am J Trop Med Hyg*, 77(6 Suppl):36-47. 2007.

[32] World Health Organization: World malaria report. *Technical document*. WHO/HTM/GMP/2008.1. 2008.

[33] Cibulskis RE, Bell D, Christophel EM, Hii J, Delacollette C, BakyaitaN, Aregawi MW: Estimating trends in the burden of malaria at country level. *Am J Trop Med Hyg*, 77(6 Suppl): 133-137. 2007.

[34] Roberts DR, Laughlin LL, Hsheih P, Legters LJ: DDT, Global strategies, and a malaria control crisis in South America. *Emerg Infect Dis*, 3:295. 1997.

[35] Jensen TP, Bukirwa H, Njama-Meya D, Francis D, Kamya MR, Rosenthal P, Dorsey G: Use of the slide positivity rate to estimate changes in malaria incidence in a cohort of Ugandan children. *Malar J*, 8:213. 2009.

[36] Metzger W, Groin A, Vivas-Martnez S, Gonzlez J, Charrasco A, Mordmller B, Magris M: A rapid malaria appraisal in the Venezuelan Amazon. *Malar J*, 8:291. 2009.

[37] Subbarao SK, Vasantha K, Raghavendra K, Sharma V, Sharma G: Anopheles culicifacies: siblings species composition and its relationship to malaria incidence. *J Am Mosq Control Assoc*, 4:29. 1988.

[38] Montanari R, Bangali A, Talukder K, Baqui A, Maheswary N, Gosh A, Rahman M, Mahmood A: Three case definitions of malaria and their effect on diagnosis, treatment and surveillance in Cox's Bazar district, Bangladesh. *Bull World Health Organ*, 79:648-656. 2001.

[39] Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis*, 4:327-336. 2004.

[40] Lee PW, Liu CT, Rampao H, do Rosario V, Shaio MF: Pre-elimination of malaria on the island of Prncipe. *Malar J*, 9:26. 2010.

[41] Roy SB, Sarkar RR, Sinha S: Theoretical investigation of malaria prevalence in two Indian cities using the response surface method. *Malar J*, 10:301. 2011.

[42] Joshi PL, Chandra R, Bhattacharya M, Vaish HC: Validity of using slide positivity rate (SPR) in identification of high risk malarious segments in rural areas. *J Commun Dis*, 29:41. 1997

[43] China Malaria Elimination Plan [http:/ / www.moh.gov.cn/ publicfiles/ business/ htmlfiles/ mohjbyfkzj/ s3593/ 201005/ 47529.htm] website.

[44] China National Malaria Office of Global Fund: Management and technique program in high malaria transmission areas in China, Shanghai; 2003.