Temporal and Spatial Estimates of Tuberculosis Infection Rates in India Using Ensemble Kalman Filter

[1]Vihari Piratla, [2]Ankit Bansal, [3]Sarita Azad, [4]Pietro Lio

[1,2,3]Indian Institute of Technology Mandi, India

[4]Computer Laboratory, University of Cambridge, UK

Abstract

Ensemble Kalman filter is a highly efficient data assimilation technique that allows the estimation of parameters that are too complex to be handled by standard-likelihood methods because of inherent seasonality or non-linearity in the data considered. We assess the utility of this method to estimate the parameters of a generalized compartmental epidemiological model, using quarterly data from the India Revised National Tuberculosis (TB) Control Programme (RNTCP). The computed results indicate that the proposed framework capture well the temporal seasonal variations present in the data and provide useful information on tracing the high-infection rate Indian states as well as making good predictions. Our model shows that the infection rate is highest in Manipur; a result which tallies with epidemiological observations in India. Furthermore, the high infection rate of TB in the central west zone calls for further intensification of tuberculosis control efforts.

Keywords: Tuberculosis, Indian states, Ensemble Kalman filter, Infection rate.

Introduction

India is one of the most Tuberculosis (TB) endemic countries in the world, which accounts for more than 25% of the world's incident cases [1]. Re-emergence of TB in developing countries--especially Multi-Drug Resistant (MDR) cases--has sparked a new interest in TB research [2,3]. A large number of factors contributing to the spread of TB; such as crowding, poor hygiene, illiteracy and lack of awareness, family size, socio-economic conditions, etc., makes the TB situation critical in Indian context. All these factors directly contribute to high infection rate among population. The Revised National TB Control Programme (RNTCP), based on the internationally recommended *Directly Observed Treatment Short*-course (DOTS) strategy, was launched by The Government of India in 1997 with support from the World Bank [4,5]. The program was intended to improve the case detection and cure rates of smear positive TB patients. TB incidence and death rates are directly related to case detection rates of various types of TB (smear positive, smear negative, extra pulmonary etc.). Worldwide WHO has a target of 75% case detection and 85% cure rate, which would lead to a reduction of 11% in the incidence rate per year.

Mathematical models, deterministic or statistical, are important tools to understand TB dynamics and analyze voluminous data collected by various agencies like WHO, RNTCP [6-11]. A very popular mathematical model in epidemiology is the three-compartment nonlinear Susceptible-Infected-Removed (SIR) model developed by Kermack and McKendrick [12]. In this model the entire population is placed in three compartments; namely, Susceptible, Infected and Recovered (denoted by S, I and R). A susceptible person is infected through contacts with an infectitious person, and moves from susceptible compartment to infected compartment. Similarly, after a certain period an infected individual recovers from the disease and moves into the recovered compartment. A variant of the SIR model is the SIS model, where after recovery individuals become susceptible again. SIR models have been known to perform excellently well in modelling temporal evolution of epidemics. Blower et al. [13] introduced yet another compartment of latently infected individuals (who have been infected with *Mycobacterium* tuberculosis, but are non infectious) into the conventional SIR model. Blower et al. assumed that the infected individuals can develop either latent infection (in which the bacteria are contained or sputum negative infection), or active disease (sputum is smear positive when examined through microscopy and the host can transmit bacteria. It is generally accepted that only 5–10% individuals who become infected with M.

tuberculosis produce primary active TB [14] and rest are infected with latent TB. In addition, some patients will enter into latent class even after successful completion of drug therapy and may develop active tuberculosis at some annual average rate.

Sputum smear-positive cases are the principal sources of infection to others, and smear positive patients also suffer higher morbidity and mortality. In contrast, people with latent TB (smear negative and extra-pulmonary) suffer less morbidity and are assumed to be non-infectious. However, Tostmann et al. [15] studied the role of sputum smear–negative TB patients in TB transmission, and reported that patients with smear-negative are responsible for 13% of TB transmission in the Netherlands. Murphy et al [16] further upgraded the model of Blower et al. to study TB transmission dynamics in demographically distinct heterogeneous populations. Through a set of six non-linear differential equations they studied the effects of genetic susceptibility and demographic factors and found that in a population with small genetic susceptibility the transmission parameter significantly affects TB prevalence. Maliyoni et al. [17] analysed a two-strain TB model and studied the impact of diagnosis, treatment, and health education etc. on the transmission dynamics of Multi-Drug-Resistant (MDR) TB. Murray et al. [18] studied the impact of various control strategies using variation of SIR model and predicted incidence rate over five regions of the world.

The biggest challenge in using such dynamic models is that the model parameters are not known very accurately. The uncertainty in these parameters can greatly affect the accuracy of the predictions made through these models. In contrast, statistical models utilize the observed data to make predictions and estimate the parameters in the deterministic models. TB incidence rates have also been probabilistically estimated from the knowledge of susceptible and infected population, infection rate and other transmission parameters. A multivariate Markov chain model was developed by Debanne et al. [19] to project TB incidence in the United States from 1980 to 2010.

Data assimilation is a class of techniques where the error-prone observation data is assimilated into models. Estimation of parameters governing the spread of epidemic and fitting models to data is very important to understand transmission pathways, forecast epidemics, design control measures and identify environmental drivers of disease processes [20]. A number of researchers have applied the methodology of data assimilation in epidemiology [21-25]. The most important parameter in the prediction of an outbreak and effects of possible interventions is the transmission rate. To find infection transmission rate is a fundamental goal of fitting dynamic models to epidemic data, and derive predictive models of transmission for nonlinear stochastic epidemic process. Some of the best methods to solve such problems are Monte Carlo methods [26] for state–space models and generalized profiling [27]. Hooker et al. [27] employed the generalized profiling method to fit infectious disease models based on the least-squares and gradient matching techniques. Data assimilation increases the accuracy & reliability of epidemic tracking by incorporating data as it arrives, with weighing factors that reflect the observed reliability of the observations. The methodology requires a dynamic model to forecast the state of the epidemic between arrivals of new data, and observations that are used to update an ensemble of state estimates. The Ensemble Kalman Filter (EnKF) [28-31] uses an ensemble of state estimates which are then propagated through the true nonlinear system and the probability density function of the actual state is approximated by the ensemble of the estimates.

In India RNTCP has been providing estimates of TB cases for all the states. Although these estimates provide useful information on the burden of the disease, it falls short of providing severity indicator such as infection rate and fraction of smear positive cases in the total TB burden. The annual TB infection rate is the best measure for assessing the temporal and spatial trends of the disease burden in a given population and for evaluating the measures for control. This will help health agencies to focus and implement resources in achieving higher case detection rates through providing greater coverage of RNTCP programmes depending on epidemiological classes based on gender, socio-economic conditions, demography etc. While implementing any strategy the knowledge of disease dynamics, infection rate for example, among these classes and their access to health care would greatly influence the success of various control programs. The objective of this paper is to apply the data assimilation model based on EnKF technique and estimate the infection rate and fraction of smear positive cases to total TB cases reported each quarter.

Methods

Dynamic Model

In this paper, we use a variation of SIR model. There are three exclusive groups of individuals; namely, susceptible, S, latently infected, L (infected with M. tuberculosis but not infectious), and actively infected with M. tuberculosis, I (infected and infectious). The model does not take into account genetic and demographic heterogeneity. The following are the governing differential equations for the rate of change in population in various compartments.

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma(I + L)$$

$$\frac{dL}{dt} = (1 - p)\frac{\beta SI}{N} - \gamma L - tL$$

$$\frac{dI}{dt} = p\frac{\beta SI}{N} - \gamma I + tL \tag{1}$$

where β is the transmission rate, γ is the recovery rate (assumed to be 0.8 for both active and latent infections [4]), and *t* is the re-infection rate from latent to active disease (assumed to be 0.1 [16]). Patients with the latent form of infection are assumed to develop active tuberculosis at an average rate of *t*, with a 5–10% lifetime risk of a latent infection reactivating to active TB disease. We assume that latently infected individuals are neither infectious nor can be re-infected from actively infected population. The rate at which M. tuberculosis infected people infect susceptible people with active TB is given as, *p*, and the rate at which latent cases are created is proportional to (1 – *p*). Generally, the reported values for *p* are between 5–10%; however, higher values of up to 40% have been reported. Contact and transmission rates for TB are subject to a number of environmental and host specific conditions; such as population density, social patterns and characteristics of both the transmitter and recipient. In general, untreated patients can infect 10-15 persons each year. It is expected that given similar health coverage, the infection rate, β, is likely to be more in regions of higher population density. This will help the health professionals to provide better coverage in critical areas by improving case detection.

Incidence rate: The incidence rate is defined as the number of new cases of certain disease per unit time. We separately define incidence rate for latent cases (In_L) and incidence rate for infected cases (In_I) and use the reported data from RNTC on smear positive and smear negative cases to validate our model.

$$In_L = (1 - p)\frac{\beta SI}{N}$$

$$In_I = p\frac{\beta SI}{N} \tag{2}$$

State space formulation

The parameters β and p in the dynamic SIR model are modelled using state-space formulation and ensemble Kalman filter. These two parameters cannot be observed directly. Therefore, we developed a model to predict these parameters from time series of active infections (smear positive) and latent infections (smear negative and extra-pulmonary) cases reported by RNTCP. The methodology employs a two-step forecasting model. The coefficients of the SIR model are written into a very simple state space model representing a Markov process as

$$a_{t+1} = \begin{bmatrix}\beta \\ p\end{bmatrix}_{t+1} = \begin{bmatrix}\beta \\ p\end{bmatrix}_t + w_t$$

$$a_{t+1} = a_t + w_t \tag{3}$$

$$w_t = N\begin{bmatrix}0.5\\0.02\end{bmatrix} \tag{4}$$

where w_t is the uncertainty in the model parameters assumed to be given by Gaussian white noise with standard deviations 0.5 and 0.02, respectively. The measurement model for the observed number of active and latent cases can be written as

$$y_t = \begin{bmatrix}In_L\\In_I\end{bmatrix}_{t+1} = \begin{bmatrix}\frac{(1-p)\beta SI}{N}\\ \frac{p\beta SI}{N}\end{bmatrix}_t + v_t$$

$$y_t = H(a_t) + v_t \tag{5}$$

where v_t is the noise in the measured variables, assumed to be Gaussian distributed with standard deviation as 5%. The model and measurement errors are usually the unknown parameters; which can be estimated using maximum likelihood estimation.

Ensemble Kalman Filter (EnKF)

The Kalman filter equations are expressed in two steps, the forecast step, where information from the measurements is used in time series model, and the analysis step, where this information is used to obtain assimilated value using Kalman gain matrix [31]. In the forecast step, we prepare an ensemble of q forecasted states with random sampling error

$$a_{t+1}^{f_i} \triangleq a_t^{f_i} + w_t^i \qquad \text{i=1,2,3.....q} \tag{6}$$

where the superscript f_i refers to the *i*-th member of the ensemble. The ensemble mean is defined as

$$\overline{a}_{t+1}^{f} \triangleq \frac{1}{q}\sum_{i=1}^{q} a_{t+1}^{f_i} \tag{7}$$

The ensemble error matrix for the state variable is defined as

$$E_{at}^{f} \triangleq \left[a_{t+1}^{f_1} - \overline{a}_{t+1}^{f} \quad \ldots \quad a_{t+1}^{f_q} - \overline{a}_{t+1}^{f}\right] \tag{8}$$

and the ensemble error matrix for the observed variables is defined as

$$E_{yt}^{a} \triangleq \left[y_{t+1}^{f_1} - \overline{y}_{t+1}^{f} \quad \ldots \quad y_{t+1}^{f_q} - \overline{y}_{t+1}^{f}\right] \tag{9}$$

Analysis steps are defined as

$$a_t^{a_i} \triangleq a_t^{f_i} + K_t(y_t + v_t^i - y_t^{f_i}) \tag{10}$$

where K_t is the Kalman gain matrix given by

$$K_t \triangleq P_{ay_t}^f \left(P_{yy_t}^f\right)^{-1} \tag{11}$$

where error covariance matrices are given by

$$P_{ay_t}^f \triangleq \frac{1}{q-1} E_{at}^f \left(E_{yt}^a\right)^T$$

$$P_{ay_t}^f \triangleq \frac{1}{q-1} E_{yt}^f \left(E_{yt}^a\right)^T \tag{12}$$

The assimilated values of the model parameters can be used to make a forecast of incidence rates of infected population as

$$y_t^f = \begin{bmatrix} \frac{\left(1-\overline{p}^a\right)\overline{\beta}^a SI}{N} \\ \frac{\overline{p}^a \overline{\beta}^a SI}{N} \end{bmatrix}_t \tag{13}$$

The observable variables in our model are incidences of active and latent cases. The prediction of these forecasts can be compared with the measured values and the error of the model is calculated using relative root mean square error as

$$RRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\overline{y}_t^f - y_t}{y_t}\right)^2} \tag{14}$$

Description of parameters selected:

An ensemble size of 1000 was considered throughout the analysis and all the ensemble parameters; number of infected cases and number of latent cases of a particular state were initialized with some random constant value. Because the initial parameters chosen were random, ensemble covariance was initialized with high value; thus ensuring that the results are independent of the knowledge of initial values or error in the estimate. All the model parameters beta, p, r, gamma were initialized with 5, 0.15, 0.02, 0.85 respectively. While beta and p being the parameters of the model were initialized randomly. r and gamma values were chosen as the nation's average. Measurement values were considered with 5% error. For the states which have large discontinuities/missing data beta and p were considered close to average value of the nation i.e. 1.5 and 0.5 respectively; so that the computation of nation's parameters as a whole is not affected.

Data

India consists of 35 states/union territories with populations ranging from less than 0.7 to 175 million. The data for the study was taken from quarterly RNTCP performance reports published by Central TB Division, Directorate General of Health Services, Revised National Tuberculosis Control Programme (RNTCP) Ministry of Health and Family Welfare, Government of India from 2006-2011[4]. DOTS programmes report quarterly data for number of new smear-positive cases and number of new smear-negative and extra-pulmonary cases.

Results

The SIR-EnKF framework explained above is implemented in MATLAB and results are post processed and plotted in R-statistical programming language with the "*spplot*" package. The objective of this section is to show that our model can predict the trend in TB cases well over various quarters and to predict important parameters, such as infection rate and fraction of smear positive cases. In Fig. 1, reported smear positive cases of TB are plotted for Manipur state for the time period 2006 to 2011. It is observed that there is a declining trend in the number of cases. A seasonal pattern is observed in the number of cases over the years, with peak infection rate always occurring in the mid of the year (April - July). Figure 2 shows the contour plot of estimated infection rate (β) for all the states in 2011(fourth quarter). Our results show that Manipur is the worst infected state. This isn't much of a surprise as Manipur has been listed as one of the worst TB affected states in the country with many reported MDR TB and HIV-TB co-infections. The overall TB prevalence rate in Manipur state is double the national average. It was reported that lack of adequate diagnostic facilities has also been cited as one of the reason for higher rate of infection in the state. Also, it is observed from Fig. 2 that Maharastra, Chattisgarh, Bihar, Mizoram and Andaman & Nicobar have the high infection rates. Furthermore, it is observed that Delhi has relatively higher infection rate when compared to other neighbouring states. This may be due to high population density and large social contact rates. On the other hand, Gujarat, West Bengal, Uttaranchal and Tripura are found to have the least infection rates. Figure 3 shows the distribution of estimated fraction of smear positive cases, *p*, over the country in 2011 (fourth quarter). It is observed that the states with the highest infection rates have the least number of smear positive cases. For example, Manipur, which has the highest infection rate, has the least value of fractional smear positive infections. In contrast, Pondicherry has the highest value of fraction of smear positive infections. In Figs. 4 and 5, trends in the infection rates and fraction of smear positive cases are plotted from 2006 to 2011 for the best and the worst affected states. The mean value of infection rate for Manipur is 2.57, which is higher than the national average of 1.72. In contrast, Pondicherry has the least infection rate with mean value of 1.34. Figure 5 shows that the fractions of smear positive cases in Manipur are almost constant with a mean value of 0.32. While the fraction of smear positive cases in Pondicherry has increased from a value of 0.3 in 2006 to 0.7 in 2011.The mean value for India is calculated by weighted mean over all the states. On an average the value of *p* has remain constant at 0.6 for the entire country. In Table 1, infection rates and fraction of smear positive cases for 2011 (fourth quarter) are listed for all the states.

Finally, we obtain the time series of assimilated TB cases for all the states using Eq. 13 and forecast cases using Eq.6. We calculated the relative root mean square error (RRMSE) between assimilated and observed values. The root mean square error in the assimilated and observed values of smear positive cases for Manipur is calculated to be 5.51. The total number of cases for India (for 2011) is calculated by summing the assimilated values of smear positive and smear negative cases for all the states. The estimated cases are found to be 21 million cases. Estimates of TB incidence, prevalence and mortality are published annually by WHO using data from national surveillance systems (case notifications and registered deaths) and special studies such as national surveys of the prevalence of disease. The estimate by WHO gives an incidence of 23 million cases for India in 2011. In Table 1, we present the forecasted values of new smear positive cases and latent cases (2012 first quarter). Our model predicted 671611 cases for India in 2012 (first Quarter).

In Figs. 6 and 7, time series of forecasted and assimilated values of Manipur are plotted along with the observed values for smear positive (In_L) and latent cases (In_I), respectively. It is observed that prediction of cases after data assimilation agree well with the observed cases. The predicted cases without data assimilation demonstrate very large error in the beginning but the agreement with the data improves over the years with data assimilation as expected. All three values (predicted, observed and assimilated) demonstrate close agreement in later years. This shows that the model is very effective in not only predicting the otherwise unknown parameters β and *p* but also gives good estimates of TB cases.

S.no.	State Name	Insert beta symbol	*p*	In_I	In_L
		2011,fourth quarter		2012,first quarter	
1	Andaman & Nicobar	2.01	0.42	109	134

2	Andhra Pradesh	1.65	0.59	24276	19113
3	Arunachal Pradesh	1.63	0.51	417	389
4	Assam	1.62	0.55	7713	7399
5	Bihar	1.84	0.52	14911	15199
6	Chandigarh	1.46	0.59	727	373
7	Chhattisgarh	1.94	0.44	4157	5640
8	D & N Haveli	1.53	0.56	97	75
9	Daman & Diu	1.44	0.59	81	65
10	Delhi	1.82	0.48	6992	7230
11	Goa	1.63	0.52	386	306
12	Gujarat	1.34	0.69	18332	10675
13	Haryana	1.47	0.60	8188	6042
14	Himachal Pradesh	1.51	0.56	2424	2033
15	Jammu & Kashmir	1.64	0.55	2786	2368
16	Jharkhand	1.68	0.54	7767	7944
17	Karnataka	1.56	0.57	14382	12064
18	Kerala	1.62	0.52	4580	4396
19	Lakshadweep	1.50	0.50	15	12
20	Madhya Pradesh	1.69	0.54	18379	17698
21	Maharashtra	1.78	0.53	24095	22671
22	Manipur	2.31	0.36	420	639
23	Meghalaya	1.71	0.50	845	855
24	Mizoram	2.22	0.37	212	351
25	Nagaland	1.73	0.50	594	608
26	Orissa	1.58	0.53	8994	8723
27	Pondicherry	1.19	0.74	806	238

28	Punjab	1.49	0.57	7446	6163
29	Rajasthan	1.49	0.57	24150	20135
30	Sikkim	1.70	0.48	200	229
31	Tamil Nadu	1.64	0.50	14076	14312
32	Tripura	1.40	0.62	650	542
33	Uttar Pradesh	1.54	0.56	61284	55868
34	Uttaranchal	1.40	0.63	50755	12001
35	West Bengal	1.44	0.60	53957	23918

Table 1. Estimated values of parameters under study for all the states in India.

Discussion

India's DOTS programme is the fastest expanding programme, and the largest in the world in terms of patients initiated on treatment, placing more than 100,000 patients on treatment every month. Despite such efforts, the number of new cases continues to grow. The notification rate data of smear positive and smear negative cases published by RNTCP does provide estimates of overall TB burden over the country but this data alone cannot be used as an indicator of the severity of the TB transmission. Since the number of new cases reported every quarter depends not only on the transmission parameter β but also on the already prevailing cases of TB, the proper indicator for evaluating the severity of the disease is β(This makes no sense to me either, remove or rephrase it... is it under reporting that is being conveyed over here?). The infection rate β represent the average number of contacts a susceptible person requires to get infected with the TB. Since, it is primarily the smear positive infections that spread the bacteria, it is expected that if the fraction of smear positive infection are large, the number of contacts needed to spread the disease would be lesser. A susceptible person makes on an average β contacts with the infected people, all potentially leading to active TB. Total number of contacts with infected people depends on the fraction of population infected with active TB of smear positive type. Thus, for similar incidence rate of smear positive cases, a higher contact rate implies that fraction of total new infections that are smear positive is low. The same trend is verified from the contours of β and *p* as given in Figs. 2 and 3.

To make RNTCP program more effective a shift in strategies is needed for states with higher transmission rates as compared with states with higher fraction of smear positive cases. For example, states with higher transmission rate but lower fraction of smear positive cases require more focus on educational and awareness programs to mitigate the transmission of TB bacteria from infected to susceptible. Transmission of infectious agent can be reduced by changing the social behaviour of human population. Awareness about a disease can reduce the susceptibility of the population. For example, in states like Bihar and Chhattisgarh people should be educated to reduce their contacts to TB infected people. On the other hand, intensely populated areas like Delhi, infected people should be encouraged to use masks. A significantly different strategy is needed in states like Manipur and Mizoram which are con-infected with HIV.

For states with lower transmission rates but high value of *p*, the strategy should be to increase the number of susceptible people examined for TB. This will help in increasing the notification rate of smear positive cases and reduce the overall burden.

The unexpected outcome of this study is relatively high infection rates in Jammu and Kashmir and low infection rates in West Bengal in spite of its very high population density. The reason could be less awareness of TB in Jammu and Kashmir and the fact that Jammu and Kashmir tops in Smoking (which doubles the vulnerability). It is also very surprising to see the higher fractions of smear positive cases in Gujarat and Uttaranchal. The implication of these results is that the health agencies need to provide extra resources wherever the infection rates are higher, and the treatment methodology significantly should depend on the distribution of smear positive and smear negative cases.

The situation of TB in India is purely endemic, no spatial pattern in the number of infections was observed among the states, however a strong seasonal pattern is observed in every particular state. Let's make a comparative study among the states: the major causes for TB transmission are over-crowding, poor hygiene, lack of awareness, malnutrition etc. Gujarat and Maharashtra being neighbouring states are at two extremes in the case of TB infection rate. These two states share almost the same geological location, same population density, with Maharashtra as a richer state when compared in the basis of GDP income. This when investigated further, it is found that Maharashtra is the only state containing 7 cities each with population more than a million. It is also the state with highest number of slums. It is also worth noting that 12 MDR-TB cases were reported in Mumbai (Maharashtra) in 2009. On the other hand Gujarat has slums number in check while the government is taking up many programmes in the up-liftment and removal of slums. Talking about Bihar, it has case detection rate way below the nation's average. Overall, India has achieved the global target of 70% case detection of TB, but Bihar lags behind at 46% with some districts detecting only 36% of cases [33]. Bihar's per capita income and proportion of people living below poverty line are way below the nation's average. Poor facilities and lack of awareness could also be a reason. It is tough to point out reasons for the situation of TB in Chhattisgarh, but possible reasons could be un-reachability to all the infected and lack of awareness along with HIV-TB co-infection. Similar causes/reasons also follow for the other states like Andhra Pradesh and Madhya Pradesh. The implication of these results is that the health agencies need to provide extra resources wherever the infection rates are higher, and the treatment methodology significantly should depend on the distribution of smear positive and smear negative cases. Instead of just extensively implementing programs such as DOTS; if the government tries to weed out the possible causes from which the problem is occurring then the solution is more stable.

The formulation of the EnKf was proposed in context of climate and whether prediction and has been re-examined with the focus on the analysis of epidemiological data in recent years. For this study, we have implemented EnKF in conjunction with the SLIS model. There are many alternate assimilation approaches that may be tested further [ ]. The parameter estimation framework presented here captures seasonality well in the data which could not be expected from standard-likelihood methods. The estimations presented on TB incidence are reasonably good although we had small sample size data of 24 years.

The main advantage of EnKF is that the prediction is improved at each stage and it doesn't rely on the history of data once ensembles are created. The technique is also computationally inexpensive.

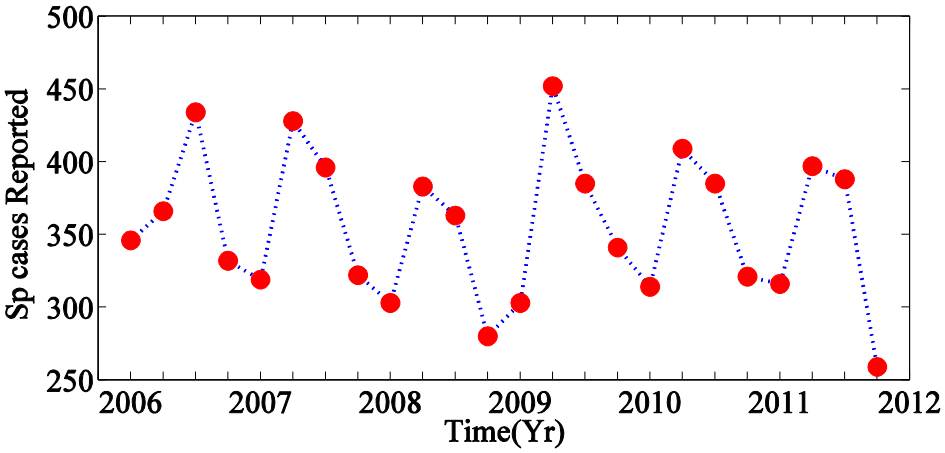

Figure 1. Reported number of new smear positive (Sp) TB cases in Manipur

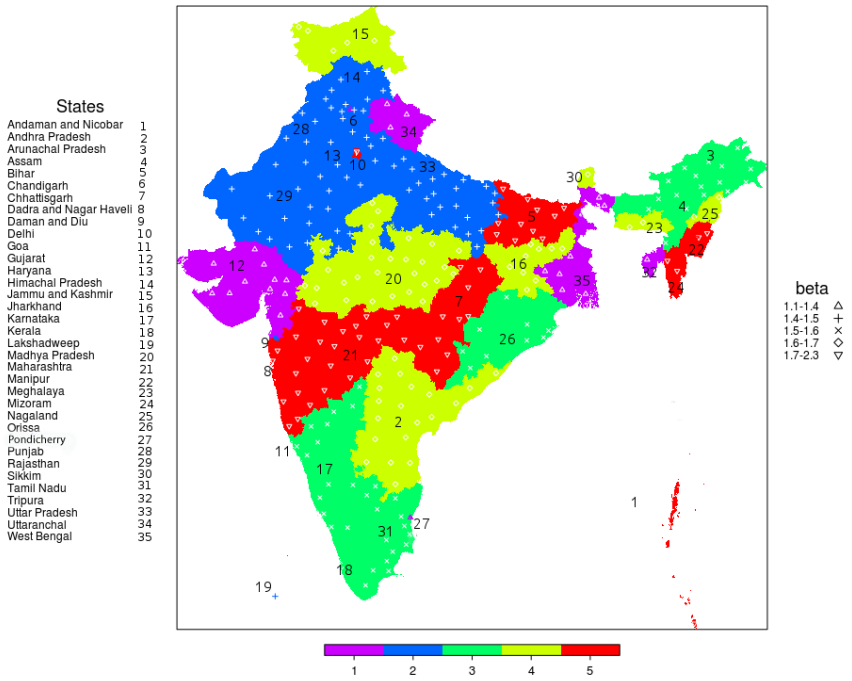

Figure 2. Estimated infection rate for Indian states (2011, fourth quarter)

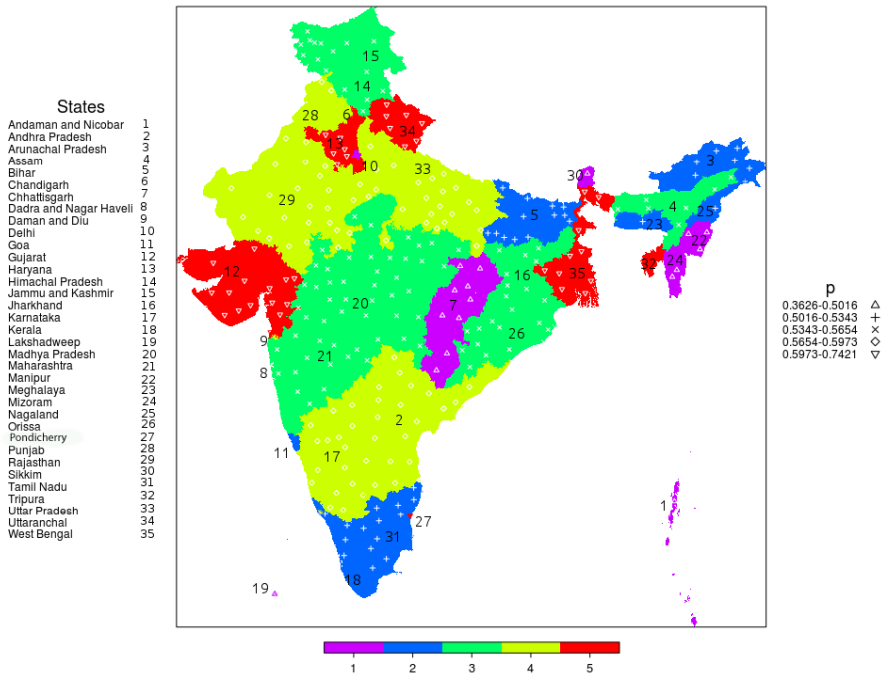

Figure 3. Estimated fraction of smear positive cases for Indian states (2011, fourth quarter)

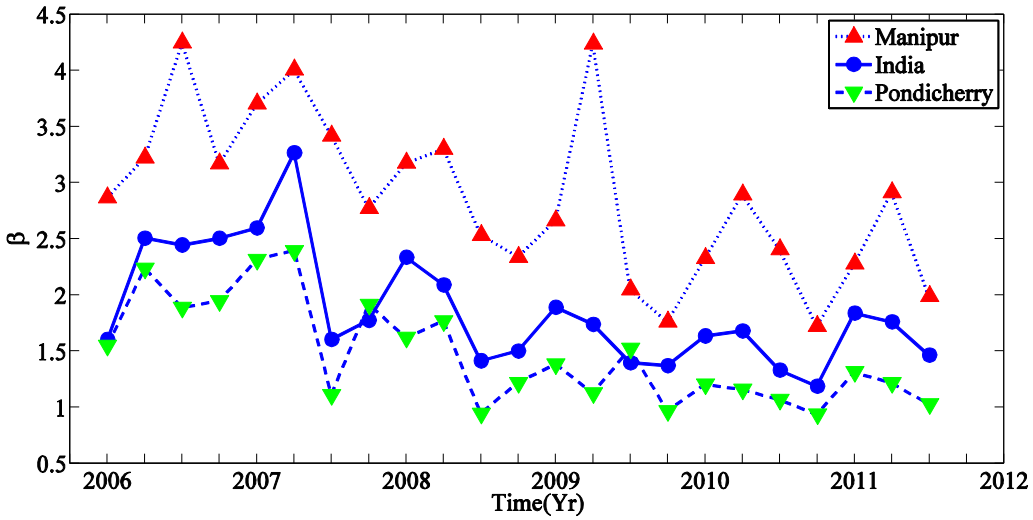

Figure 4. Estimated infection rate

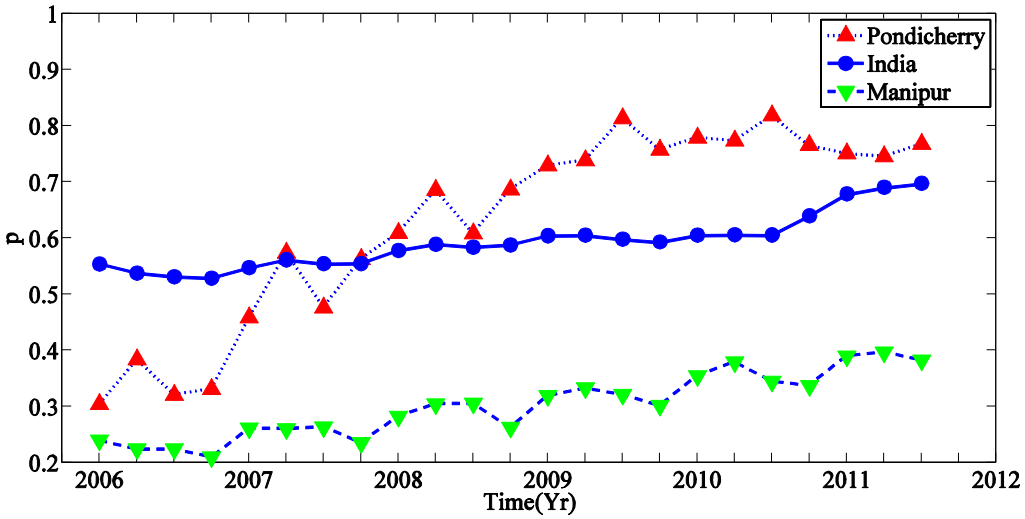

Figure 5. Estimated fraction of smear positive cases

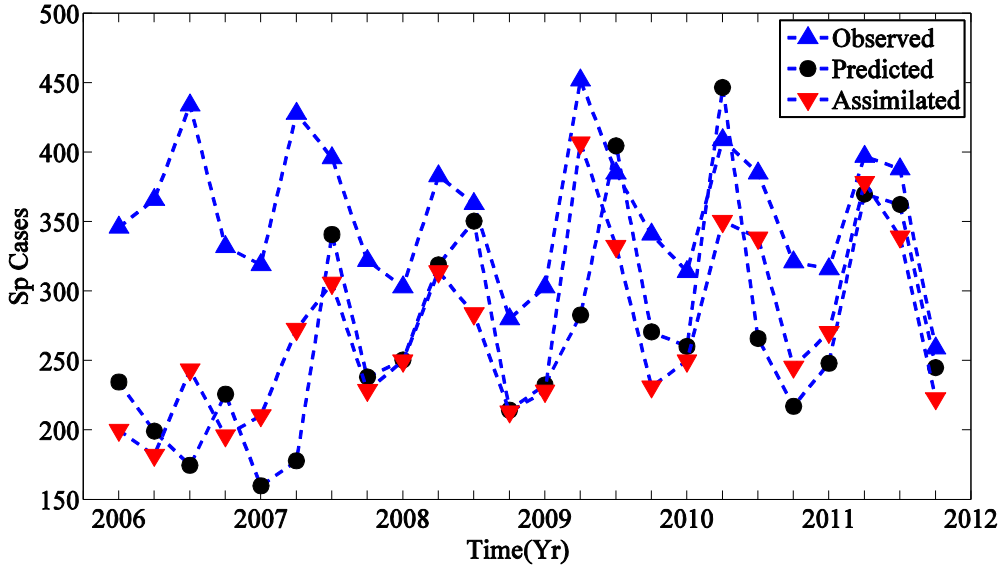

Figure 6. Smear positive cases for Manipur.

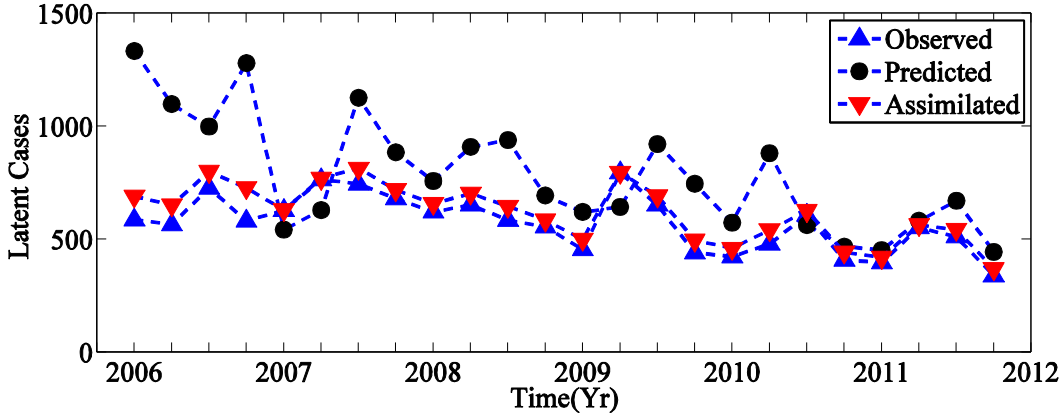

Figure 7. Smear negative and extra pulmonary cases for Manipur.

References

[1] Tuberculosis Control in South East Asia Region, (2012) SEA/TB/338

[2] Udwadia Z. F., Amale R. A., Ajbani K. K (2012) Totally Drug-Resistant Tuberculosis in India, Clin. Infect. Dis. 54 (4), 579-581.

[3]Criteria for the control of drug-resistant tuberculosis *Proc. Natl. Acad. Sci. USA 2000 97 (14) 8180-8185*

[4] RNTCP Annual Status Report (2006-2012) TB India.

[5] Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet* 1998; 352: 1886-91 doi: 10.1016/S0140-6736(98)03199-7 pmid: 9863786.

[6] Noakes C. J., Beggs C. B., Sleigh P.A., Kerr K. G. Modelling the Transmission of Airborne Infections in Enclosed Spaces (2006) Epidemiol. Infect. 134, 1082–1091.

[7] Gammaitoni L., Nucci M. C. Using a mathematical model to evaluate the efficacy of TB control measures. (1997) Emerging Infectious Diseases 3, 335–342.

[8] Beggs C. B. *et al*. The transmission of tuberculosis in confined spaces : an analytical study of alternative epidemiological models. International Journal of Tuberculosis and Lung Disease (2003)7, 1015–1026.

[9] Daley D. J., Gani J. Cannings C., Hoppensteadt F. C., Segel L. A., Epidemic modelling, an introduction (1999). In: eds. Cambridge Studies in Mathematical Biology 15,Cambridge, UK: Cambridge University Press

[10] Castillo-Chavez C, Feng Z. Jann-Yuan Wang, Li-Na Lee, Hsin-Chih Lai, Hsiao-Leng Hsu, Yuang-Shuang Liaw,

To treat or not to treat : the case of tuberculosis. Journal of Mathematical Biology 1997;35: 629–656

[11] Po-Ren Hsueh, and Pan-Chyr Yang. Prediction of the Tuberculosis Reinfection Proportion from the Local Incidence. The Journal of Infectious Diseases 2007; 196:281–8

[12] Kermack, W. O. and McKendrick, A. G., A Contribution to the Mathematical Theory of Epidemics," Royal Society of London Proceedings Series A, 115, 700-721, 1927.

[13] Blower, S. M., Mclean, A. R., Porco, T. C., Small, P. M., Hopewell, P. C., Sanchez, M. A., et al. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. Nature Medicine, 1, 815 – 821.

[14] G. Comstock, Epidemiology of tuberculosis, Am. Rev. Respirat. Dis. 125 (1982) 8.

[15] Alma Tostmann, Sandra V. Kik, Dick van Soolingen, Nico A. Kalisvaart, Maruschka M. Sebek, Suzanne Verver, Martin J. Boeree. Tuberculosis Transmission by Patients with Smear Negative Pulmonary Tuberculosis in a Large Cohort in The Netherlands. Clinical Infectious Diseases 2008; 47:1135–42

[16] Brian M. Murphy, Benjamin H. Singer, Shoana Anderson, Denise Kirschner, Comparing Epidemic Tuberculosis in Demographically Distinct Heterogeneous Populations, Mathematical Biosciences,180, 161–185, 2002

[17] M. Maliyoni, P. M. M. Mwamtobe, S. D. Hove-Musekwa, and J.M.Tchuenche, Modelling the Role of Diagnosis, Treatment, and Health Education on Multidrug-Resistant Tuberculosis Dynamics, International Scholarly Research Network, ISRN Biomathematics, Volume 2012, 459829

[18] Murray CJ, Salomon JA. Modeling the impact of global tuberculosis control strategies. *Proc Natl Acad Sci USA* 1998; 95: 13881-6

[19] Sara M. Debanne, Roger A. Bielefeld, George M. Cauthen, Thomas M. Daniel, and Douglas Y. Rowland. Multivariate Markovian Modeling of Tuberculosis: Forecast for the United States. Emerging Infectious Diseases Vol. 6, No. 2, March–April 2000

[20] Kalnay, E. (2003), Atmospheric Modeling, Data assimilation, and Predictability, Cambridge Univ Pr.

[21] Kalivianakis, M., Mous, S. L. J., and Grasman, J. (1994), Reconstruction of the Seasonally Varying Contact Rate for Measles," Mathematical Biosciences, 124, 225-234.

[22] Cazelles, B. and Chau, N. P. (1997), Using the Kalman Filter and Dynamic Models to Assess the Changing HIV/AIDS Epidemic," Mathematical Biosciences, 140, 131--154.

[23] Shaman J, Karspeck A (2012) Forecasting seasonal outbreaks of Influenza. *Proc Natl Acad Sci USA* 109:20425-20430.

[24] J`egat, C., Carrat, F., Lajaunie, C., and Wackernagel, H. (2008), Early Detection and Assessment of Epidemics by Particle Filtering," in GeoENV VI-Geostatistics for Environmental Applications: Proceedings of the Sixth European Conference on Geostatistics for Environmental Applications, Springer Verlag, pp. 23--35.

[25] Krishnamurthy, A., Cobb, L., Mandel, J. , Beexley, J. (2010), Bayesian Tracking of Emerging Epidemics Using Ensemble Optimal Statistical Interpolation (EnOSI), Section on Statistics in Epidemology--JSM, pp. 3471--3484

[26] Morton, A. & Finkenstaedt, B. (2005) Discrete time modelling of disease incidence time series by using Markov Chain Monte Carlo methods. J. R. Statist. Soc. C 54, 575–594.

[27] Giles Hooker, Stephen P. Ellner, Laura De Vargas Roditi and David J. D. Earn, Parameterizing state–space models for infectious disease dynamics by generalized profiling: measles in Ontario, J. R. Soc. Interface, 8, 961–974, (2011)

[28] Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998), Analysis scheme in the ensemble Kalman Filter," Monthly Weather Review, 126, 1719--1724.

[29] Houtekamer, P. L. and Mitchell, H. L. (1998), Data Assimilation Using an Ensemble Kalman Filter Technique," Monthly Weather Review, 126, 796--811.

[31] S. Gillijns, O. Barrero Mendoza, J. Chandrasekar, B. L. R. De Moor, D. S. Bernstein, and A. Ridley , What Is the Ensemble Kalman Filter and How Well Does it Work?, Proceedings of the 2006 American Control Conference, Minneapolis, Minnesota, USA, June, 2006