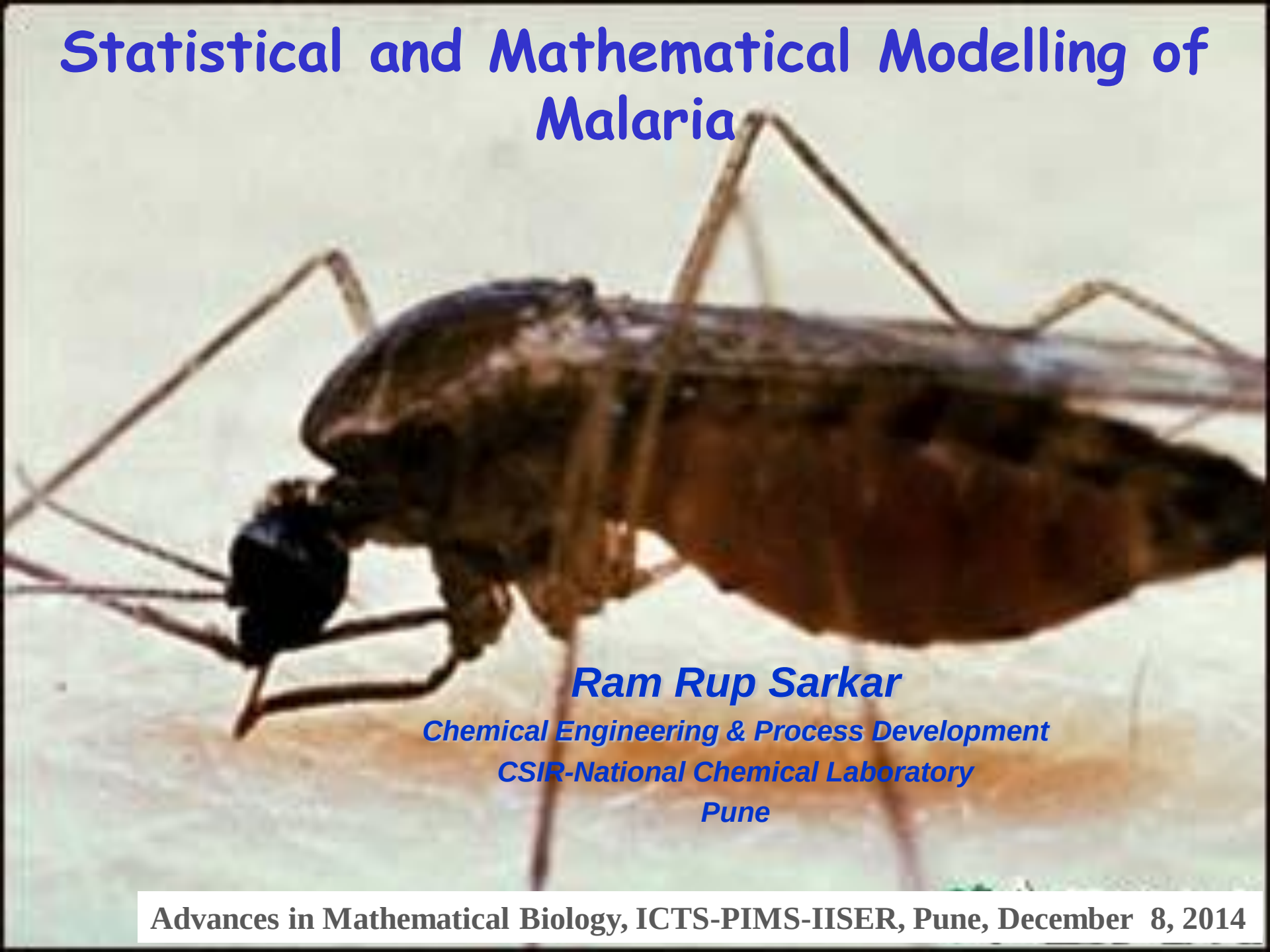
Statistical and Mathematical Modelling of Malaria



Ram Rup Sarkar

Chemical Engineering & Process Development

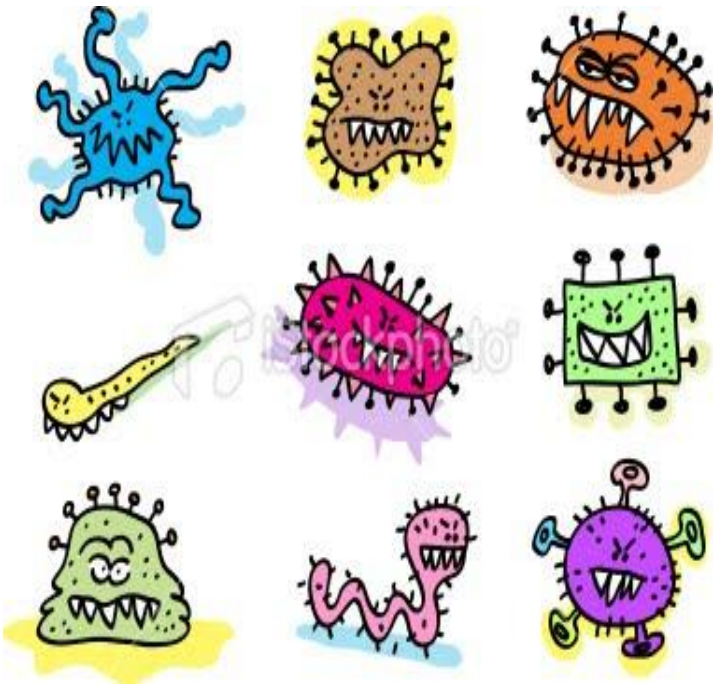
CSIR-National Chemical Laboratory

Pune

Advances in Mathematical Biology, ICTS-PIMS-IISER, Pune, December 8, 2014

Infectious Diseases Are Big Problems

Infectious diseases are big problems in India and worldwide, for people of all ages, as well as for livestock.



2005: More than 130,000 cases of cholera occur worldwide

2006: More than 350,000 cases of gonorrhea are reported in the United States

2007: 33.2 million people worldwide have HIV infections

Each year in the United States, 5% to 20% of the population gets the flu and 36,000 die

2012: Total new and relapse cases of **TB – 12,89,836**; Total cases notified-**14,67,585**

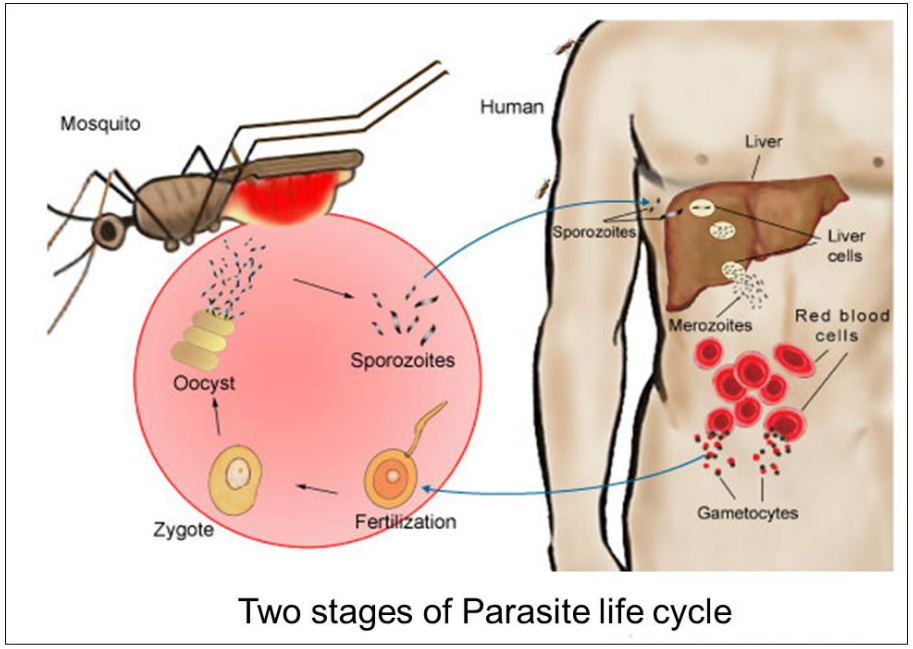
2000-2012: On an average **2% of the entire population** of India tested positive for **Malaria**

Source: World Health Organization (http://www.who.int/en/)

Malaria - the world's most important tropical parasitic disease

- causing serious threats to human communities in recent times (seems to be on a come-back mode)

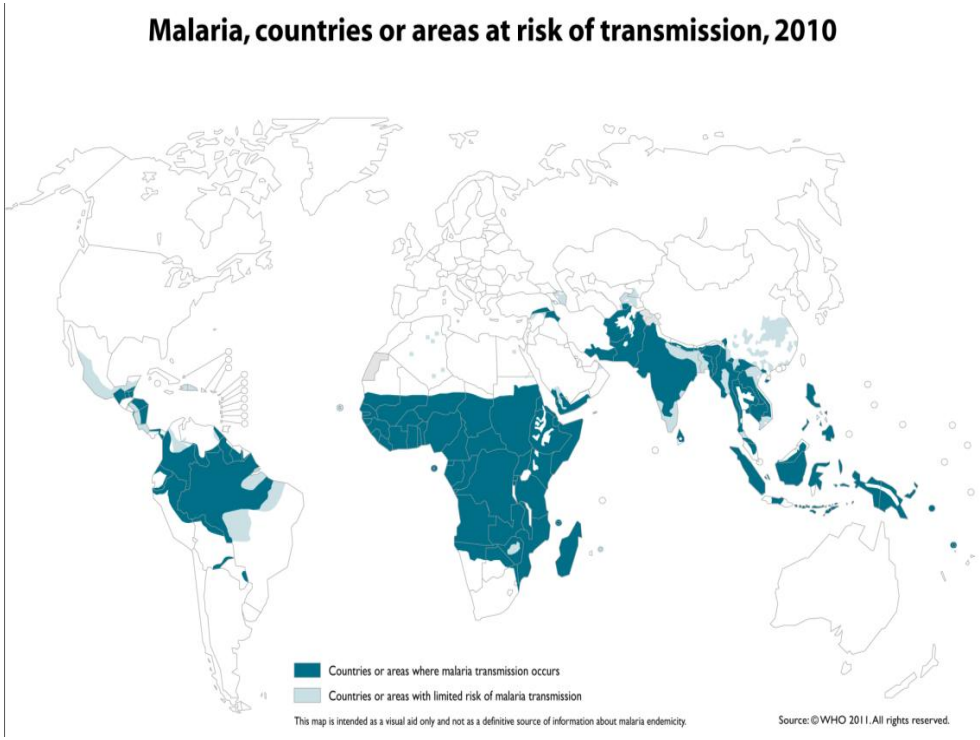
Caused by the parasite Plasmodium: *Plasmodium falciparum, P. vivax, P. malariae* and *P. ovale* and the vector (Anopheles mosquitoes)



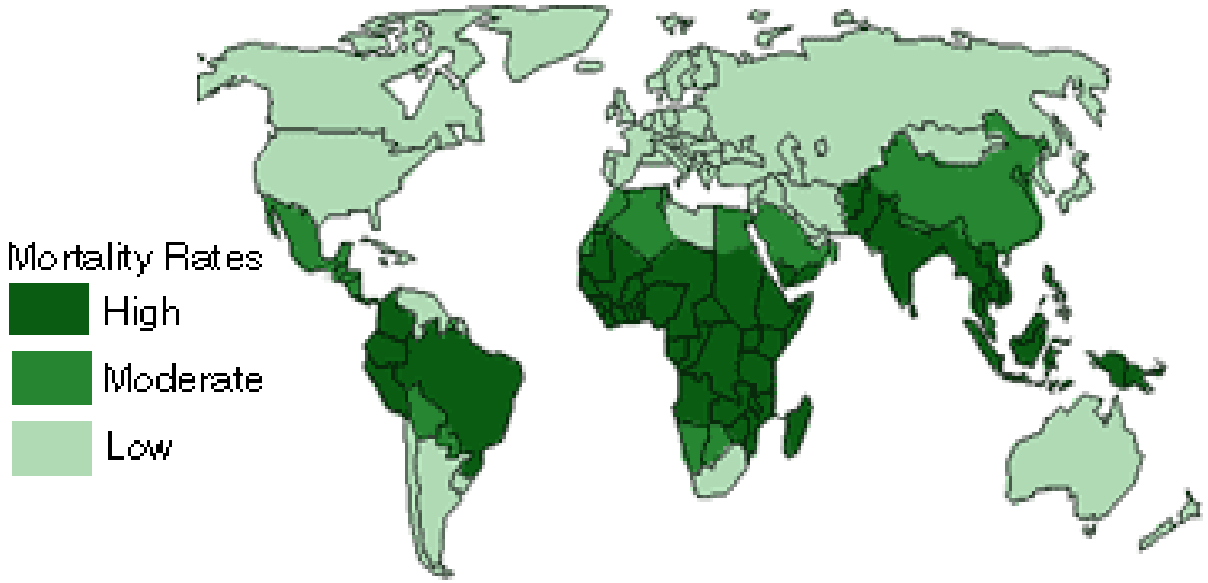
Two stages of Parasite life cycle

Global Statistics

- 40% of the world's population is at risk
- 300-500 million new cases/year
- 1.5-2.7 million deaths/year
- Malaria is endemic to over 100 countries and territories
- More than 90% of all cases are in sub-Saharan Africa

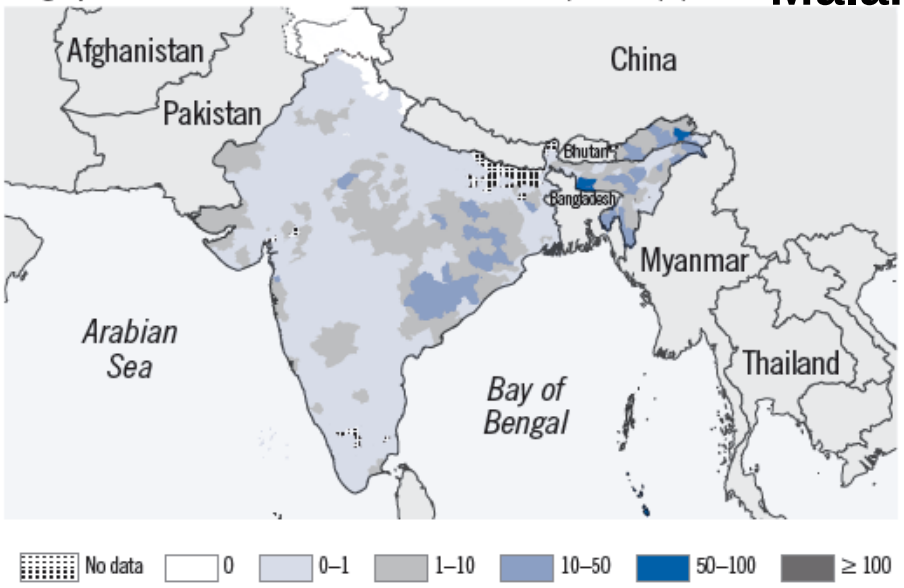


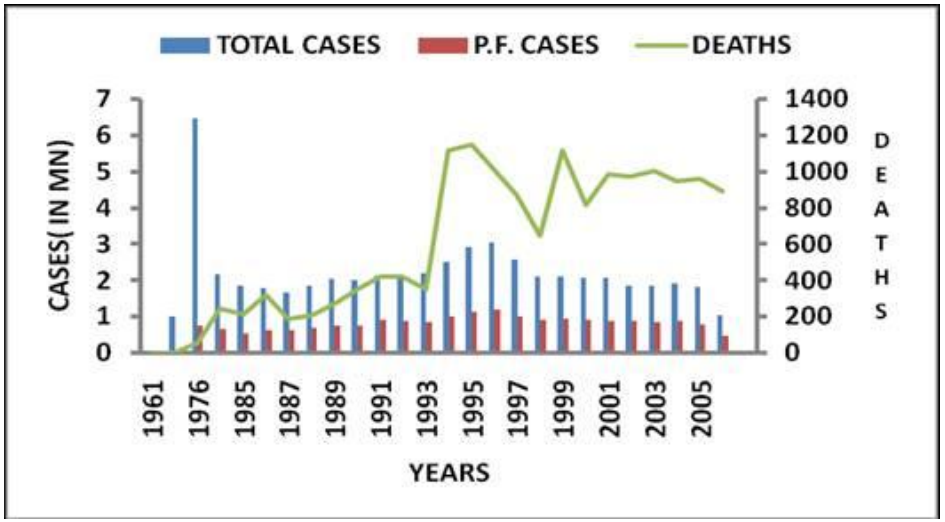
WHO Report, 2010)



Geographical distribution of confirmed malaria cases (per 1000 population)

Malaria Map India (WHO Report, 2010)





THE STATE OF MALARIA IN INDIA FROM 1961 TO 2006

- **Malaria** is also one of the major epidemics observed in India.
- Prevalent in all parts of the country
- **PF cases:** A constant high from 1993 except the period between 1995 and 1999

Defining Epidemiology

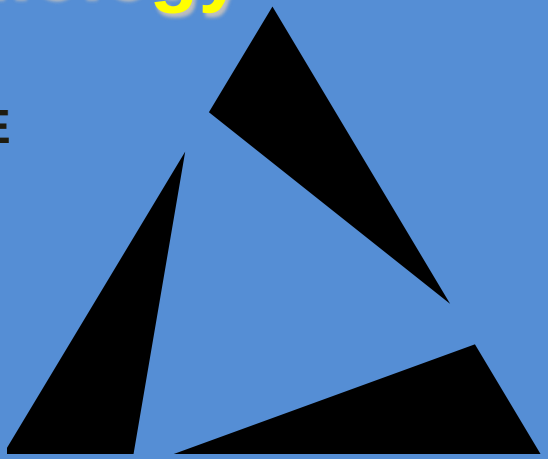
"The study of the distribution and determinants of health related states and events in populations, and the application of this study to control health problems."

John M Last
Dictionary of Epidemiology

Basic Triad of Descriptive Epidemiology

THE THREE ESSENTIAL CHARACTERISTICS OF DISEASE WE LOOK FOR:

- **PERSON**
- **PLACE**
- **TIME**



Stages of an Infectious Disease

(generic)

- Incubation (latent) period
- Prodromal (initial, pre-eruptive) period
- Overtly symptomatic (infectious) period
- Recovery period (no longer infectious)

Depending on the disease, a person may or may not be able to transmit the disease during incubation and prodromal periods

Relative infectiousness in the prodromal and the symptomatic periods determine the optimal control strategy

Definitions

Latent period (τ): The period from the point of infection to the beginning of the state of infectiousness during which the infected individuals stay in the **Exposed** class.

Incubation period: The period from the point of infection to the appearance of symptoms of disease

Asymptomatic: In some infections, symptoms never appear in the individual and are called asymptomatic infection.

Clinical immunity: The immunity, which reduces the probability of clinical disease, is called Clinical immunity.

Anti-parasite immunity: The immunity, which is responsible for clearance of parasite is called Anti-parasite immunity.

Effectiveness of treatment (ε): The ratio of the duration of infection for the untreated and treated sensitive parasites.

Vector borne Diseases

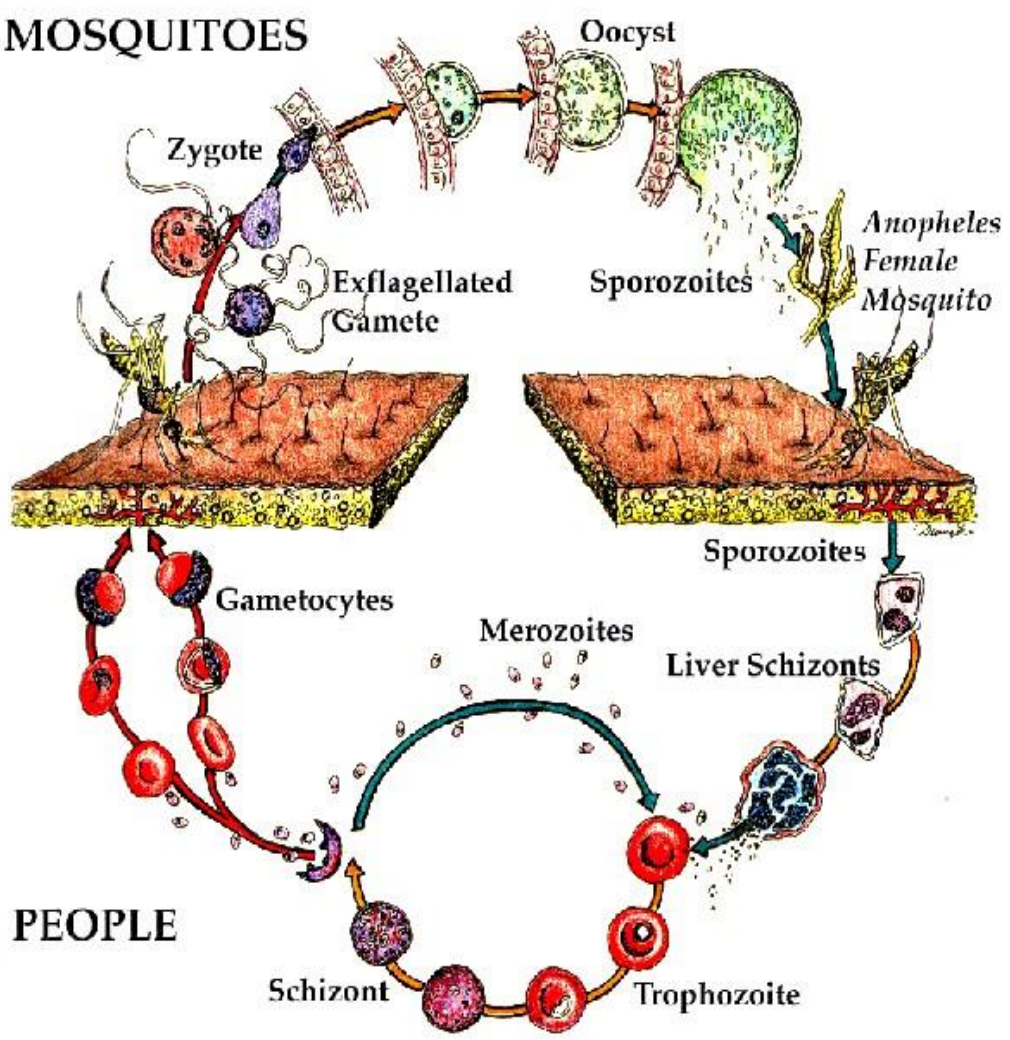
Malaria, Leishmania etc.

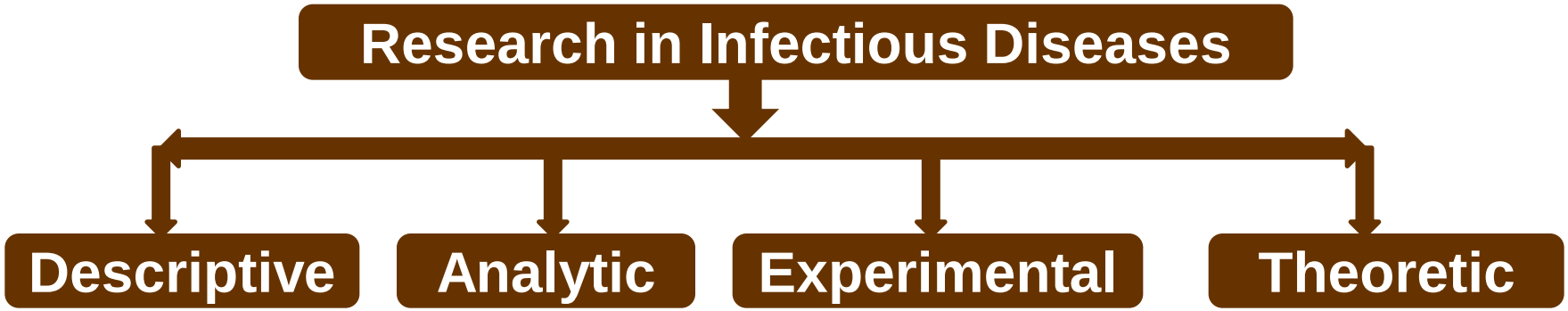
Entomological Inoculation Rate (EIR):

Rate of infectious bites per person is termed as Entomological inoculation rate.

Vectorial Capacity (C):

All the information (vector density relative to host, biting rate, life expectancy etc.) about the vector populations is incorporated through vectorial capacity, which is defined as a **measurement of efficiency of vector borne disease transmission.**





Models can provide an explicit framework to develop and describe an understanding of infectious disease transmission dynamics

- Mathematical models for **smallpox** – 1760 **Bernouilli**
- **Hamer** - discrete-time model for the spread of **measles** in 1906.
- **Ross** (1911) - transmissions of **malaria** between human beings and mosquitoes. (**Second Nobel Prize in Medicine**)
- **Kermack and McKendrick** -**SIR** (susceptible–infective–recovered) & **SIS** compartmental model, in 1926 - outbreak of **Black Death** in London, and the outbreak of **plague** in Mumbai .

Sir Ronald Ross

▪ **Theoretical Approaches** - **population dynamics**, **behavior of disease transmissions**, **features of the infectious agents**, and **connections** with other **social** and **environmental factors**.

Two broad methods in modeling

– Mathematical modeling and Data based / Statistical modeling

Why Mathematical Modeling?

STRENGTHS AND BENEFITS

Mathematical modeling is typically the only way to
examine the possible impact of different release and control scenarios

Questions that can be addressed:
What fraction of the population should be quarantined and/or vaccinated?
How fast have control measures to be implemented?, etc.

PROBLEMS

Simple models cannot capture the complexity of epidemics and their dynamics
Complex models are intransparent and difficult to validate

Several important aspects of epidemics are difficult to quantify
(e.g. response of population to certain events)

•The **concept of thresholds** established the fundamentals of the theory of epidemic dynamics.

• More intensive studies on epidemic dynamics took place after the middle of the 20th century.

• A landmark publication is the book by **Bailey** (first edition in 1957; second edition in 1975)

• More developments and progresses - during the past 20 years.

• Massive mathematical models have been formulated and developed to study various infectious diseases, ranging **from more theoretic to**

general ones [Waltman (1974); Burnett and White (1974); Hoppensteadt (1975); Frauenthal (1980); Anderson and May (1982); Evans (1982); Webb (1985); Kranz (1990); Busenberg and Cooke (1993); Capasso (1993); Isham and Medley (1996); Daley and Gani (1999); Diekman and Heesterbeek (2000)]

more specific ones e.g. measles, malaria, tuberculosis, sexually transmitted diseases (STD), or AID/HIV [Hethcote and Yorke (1984); Hethcote (2000); Hyman and Stanley (1988); Brauer and Castillo-Chavez (2001); Brauer *et al. (2008)].*

Challenges…….

- **Massive diverse progress in malaria research**

  - yet the tools and methods used to measure the intensity of its transmission are still not standardized

- Even in the maximum cases one **does not take into account different factors like ecological, demographic, and socioeconomic differences across populations** [*Hay et al., 2000, 2005; Robert et al., 2003; Kelly-Hope & McKenzie, 2009*]

  - Transmission intensity of malaria is highly variable and significantly correlated with monthly rainfall and daily temperature variation.

  - **Indicate the need for consideration of an appropriate climatic variation term in explaining malaria dynamics in different geographical locations.**

Challenges…….

- **From the long term (more than one year) prevalence pattern of malaria, it is observed that along with the monthly variation of malaria cases there appears another long term year wise trend on the pattern.**

  Malaria cases continuously rise for few years, reaches to peak values, and subsequently come down to its normal variation.

- **Effect of other external factors different than environmental fluctuations**

  Resurgence in construction activities and water logging in construction sites, lead to increase in malaria cases, whereas number of cases may decrease due to mass vaccination policy or pesticide spray so as to control the situation in that place.

Even the effects of different factors are known, but **how they are influencing the disease prevalence is an important problem - for better understanding and future predictions**

Challenges…….

Most control strategies of Malaria currently are **based on African data** (e.g., age-structure of infection, gender difference in infection, etc), **whereas, the disease dynamics may be different in other countries like, India**

Major challenge:

To develop a universal approach to ascertain the **detailed understanding of the components of the disease under the effect of environmental and social variations**, so that realistic outcome can lead to clear and reliable predictions towards the control

Statistical Models

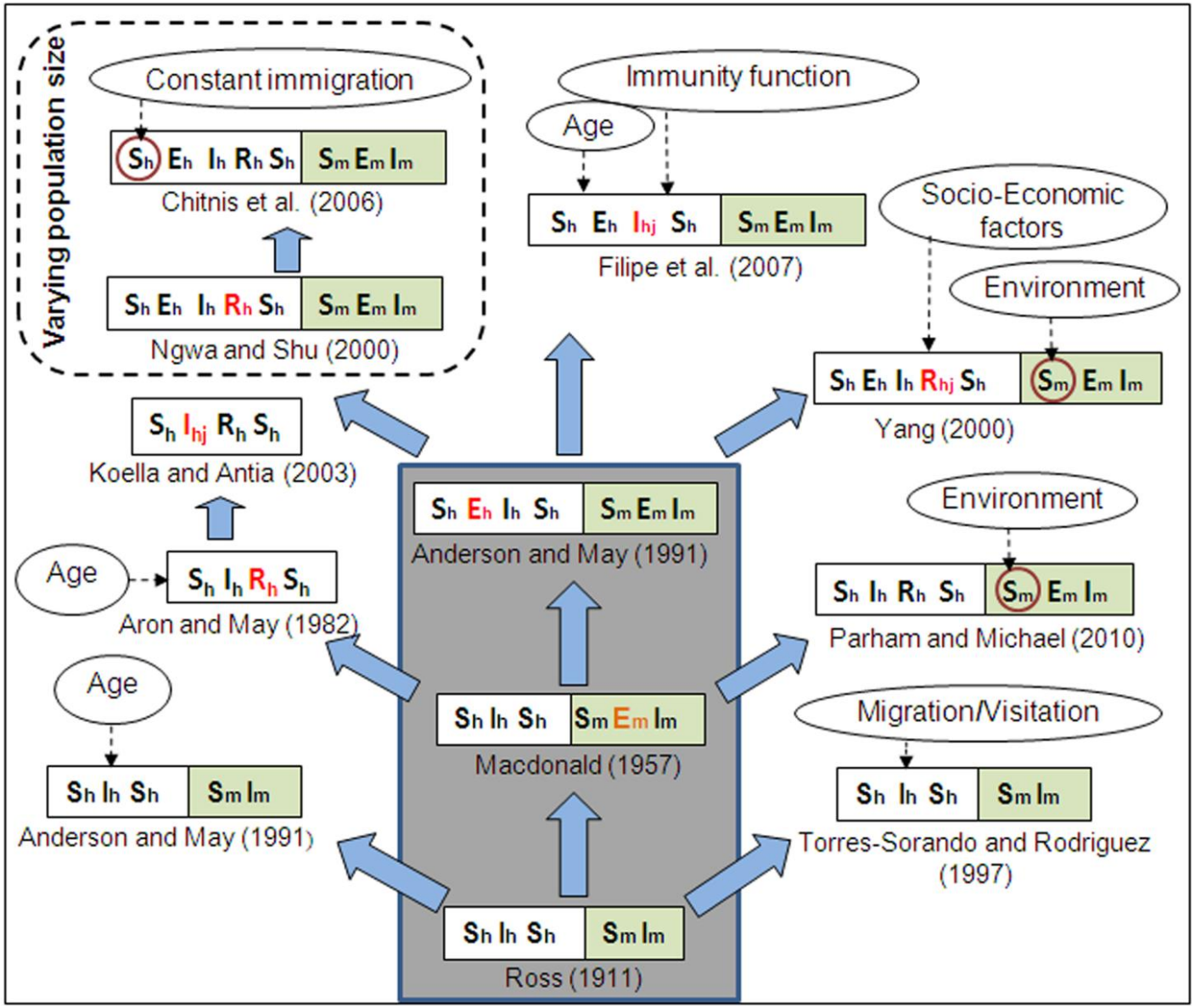
- Account for bias and random error to find correlations that may imply causality.
- Often the first step to assessing relationships.

Dynamic Models

- Systems Approach: Explicitly model multiple mechanisms to understand their interactions.
- Links observed relationships at different scales.

By developing dynamic models in a probabilistic framework we can account for dependence, random error, and bias while linking patterns at multiple scales.

Evolution and grouping of different types of malaria models



Mandal, Sarkar, Sinha, *Malaria Journal*, 2011

Biostatistical Analyses

Data-based statistical modeling on the available disease prevalence data for different categorizations:

An integrative approach that uses the past data to predict the future trend

Applicability of such models becomes important while assessing their numerical outputs, estimating parameters, and predicting future values from past observations.

A combined mathematical-statistical approach (**the Liverpool Malaria model**), that uses the dynamic transmission models of the SIRS-type framework and statistical methods for correlating environmental dependencies, was developed recently to successfully forecast the evolution of malaria epidemiology in western Africa [*Gaudart et al., 2009; Ermert et al., 2011*].

Few other problems…..

Data collection is mostly restricted to short time scales (months to few years)

Recently researchers have used elaborate time series analysis models to show seasonality pattern in malaria incidence, and the Monte-Carlo Markov Chain methods with Bayesian techniques of a-priori probability assignment, to estimate the risk factors.

[Briet et al, 2008; Abellana et al, 2008; Cancre et al, 2000]

Drawbacks: Suitability of these models being fitted into the incidence pattern of the disease is questionable.

For instance, discontinuity is observed in the time series at high temporal resolution while studying extensive dataset for elucidating climatic role in the transmission of malaria in Africa and Europe. *[Kuhn et al, 2002]*

Different types of Statistical Models For Long-term Study

Time stratified model

Modelling long-term patterns in the outcome data by spliting the study period into intervals and estimate for each interval

- simply including an indicator variable for each time interval in the Poisson model.

Pros: easy to understand, captures main long-term patterns quite well.

Cons: potentially large number of model parameters; implicitly assumes biologically implausible jumps in risk between

Periodic functions

Fitting Fourier terms in the Poisson model

- pairs of sine / cosine functions of time with an underlying period reflecting the full seasonal cycle (i.e. calendar year)

Pros: models long-term patterns smoothly, relatively few parameters.

Cons: mathematically complex; Can capture very regular seasonal patterns; modelled seasonal pattern is always forced to be the same from one year to

ar to

vell).

Residual variation around the long-term pattern

Flexible spline functions

Fit a spline function of time

- Essentially a number of different polynomial (most commonly cubic) curves that are joined smoothly end-to-end to cover the full period.

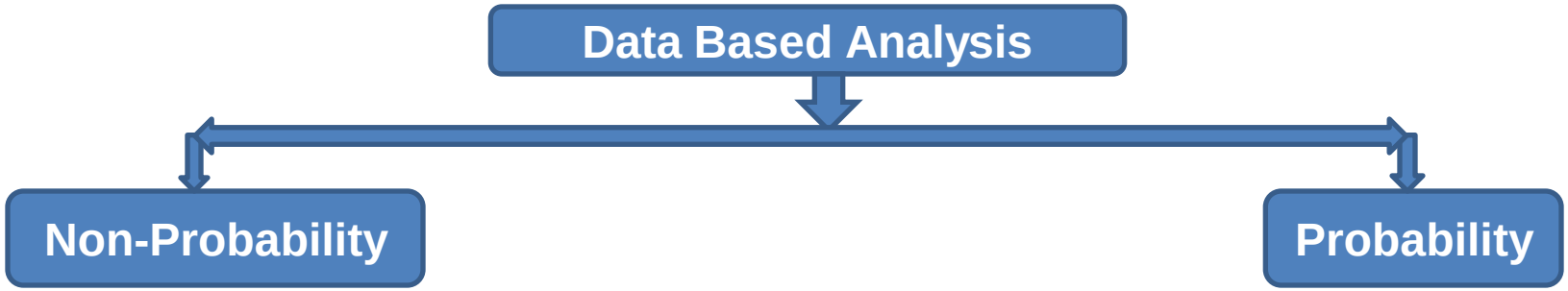
Pros: models long-term patterns smoothly; can capture seasonal patterns that vary from one year to the next; and capture long-term non-seasonal trends in the data.

Cons: more mathematically complex than the other methods

- ❑ **Exposure-outcome associations and confounding**
- ❑ **Confounding by other time-varying factors**
- ❑ **Allowing for delayed exposure effects**
- ❑ **Model checking and sensitivity analysis**
- ❑ **Precision and power considerations**

- ➢ **Non-linearity**
- ➢ **Investigation of effect modifiers**
- ➢ **Analysis of data from multiple locations**

  Separate analysis by location (e.g. specific cities) can increase power and provide information on heterogeneity and adaptation to environmental exposures



Regression Methods

- **Linear**
  - Without Transformation of Variables
  - With Transformation of Variables
- **Non-Linear**
  - ***Polynomial*** *(Ruru & Barrios, 2003)*
  - ***Logistic regression modeling***
    (spatial mapping of malaria**, MARA Project**)
  - ***Binary logistic regression modeling***
    (*Ye et al.,* ***BMC Public Health*** *2007*)

Time Series Analysis

- **Monte-Carlo Markov chain** methods
- **ARIMA** *(Briet et al.,* ***Malaria J****, 2008)*
- **Auto Regressive Conditionally Heteroskedastic models** (ARCH)
- **Random Walk model**

Poisson regression model

(Uses both Probability and Regression approach, *Ruru & Barrios, 2003*)

Stepwise Regression Method with Transformation

(Modelling & Forecasting *malaria incidence* in Kolkata, India, ***Ecol. Modelling***, *2004*)

Multi-step Polynomial Regression without Transformation

(Modeling and Forecasting *malaria incidence* in Chennai, ***PLoS ONE***, *2009*)

Response Surface and Model Reduction

(Modeling and Forecasting *malaria* in two cities: Mangalore & Chennai, ***Malaria Journal***,*2011*)

Malaria control – need for models

To understand the causative factors, degree of influence and disease transmission dynamics over a horizon

Variability due to different environmental factors (temperature, humidity, rainfall etc.) on the disease dynamics cannot be ignored.

Objective

To see, how the environmental factors and previous disease incidence influencing the incidence of Malaria?

To predict future incidence of the disease – Using simple statistical modelling techniques

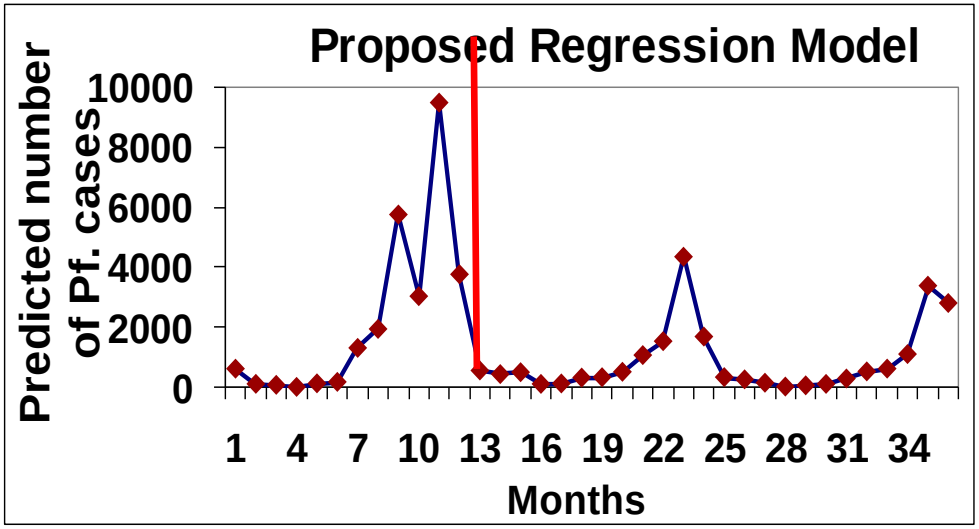
- **Earlier works** concentrated on inducing linearity between variables - fitting a general linear model to different influential factors
- Those who have considered non-linear models, have predefined the functional form of the model before getting into the dynamics between the variables

Proposed suitable statistical model to study the pattern followed by Malaria cases (or SPR values) and/or deaths over the years and eventually predict the course of the disease – Kolkata & Chennai

➢ **A stepwise regression model with Transformation**
- environmental factors along with a socio-economic factor based on the available data from Kolkata Municipal Corporation (KMC)

✓ **Compared this model with the basic malaria model by introducing environmental stochasticity.**

✓ **Predictions could be useful to design proper control strategy**



J. Chattopadhyay, **R.R. Sarkar**, S, Chaki and S, Bhattacharya (2004) ***Ecological Modelling***

Problem?

Completely based on linear regression after transformation of variables

Unable to capture the non-linear relationships between the Environmental and Social Factors influencing the disease dynamics

With no a-priori assumption of independent variables and not presuming about the nature of the dependent variable –whether discrete or continuous

We developed a non-linear multistep regression methodology to model and forecast malaria incidence in Chennai City, India

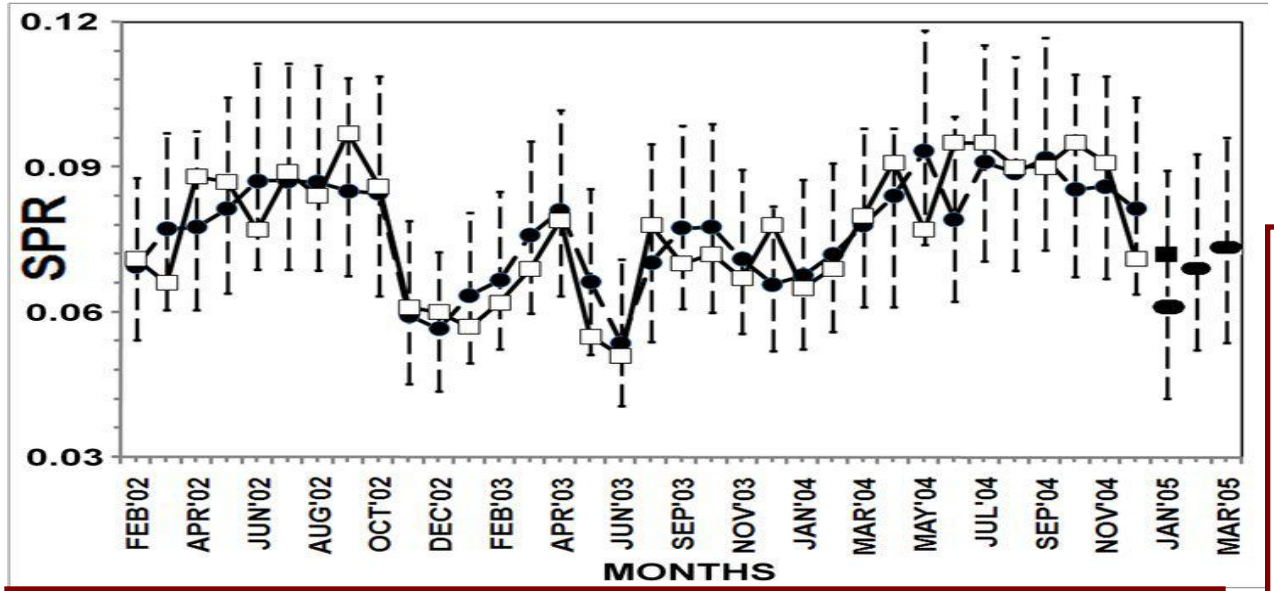
Three types of data:

i) **Longer (**January 2002 to January 2005) **time series data of Slide Positivity Rates (SPR)** for two types of malaria *P. vivax* and *P. falciparum*;

(ii) **Shorter time series data (deaths due to *Plasmodium vivax*) of one year** (January –December 2006);

(iii) Spatial data (zonal distribution of *P. vivax* deaths) for the city

Along with Climatic factors, Population and previous incidence of the disease.



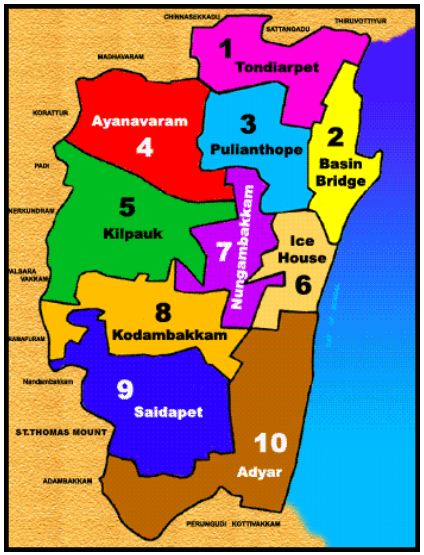
Straight line with white squares - *observed values of SPR;*

Dashed line with black circles – *predicted values from Model.*

Error bars are of 95% confidence intervals of the predicted response.

Black ellipses - *forecasted values of SPR (autoregressive forecasts)*

Black square – *observed SPR value for January, 2005.*



Zones of Chennai Corp.

Log (SPR) = - 19.708 + 15.007*log (Max.Temp) - 1.05*log (Max.Temp)^4 + 2.963E-28* (Population)^4 + 4.517*(SPR-at-lag-one) - 9740.329*(SPR-at-lag-one)^4 + 86473.283* (SPR-at-lag-one)^5 + 3.687E-02*(Minimum Humidity) - 3.14E-04* (Minimum Humidity)^2.

(Model proposed on 34 Time points – Forecast for 3 time points)

C.Chatterjee & R.R.Sarkar,
***PLoS ONE*, 4(3): e4726. 2009**

Our method introduces successively higher powers of the chosen independent variables

Problem Again?

Leads to a complex model that does not ensure a trade-off between the number and numerical order of terms and, the goodness of fit of the model

Method was too much area specific, **unable to capture -**
considerably different levels of disease severity, population density, rainfalls and vegetation patterns in different regions

A Tale of Two Cities

Epidemiological study of Malaria prevalence

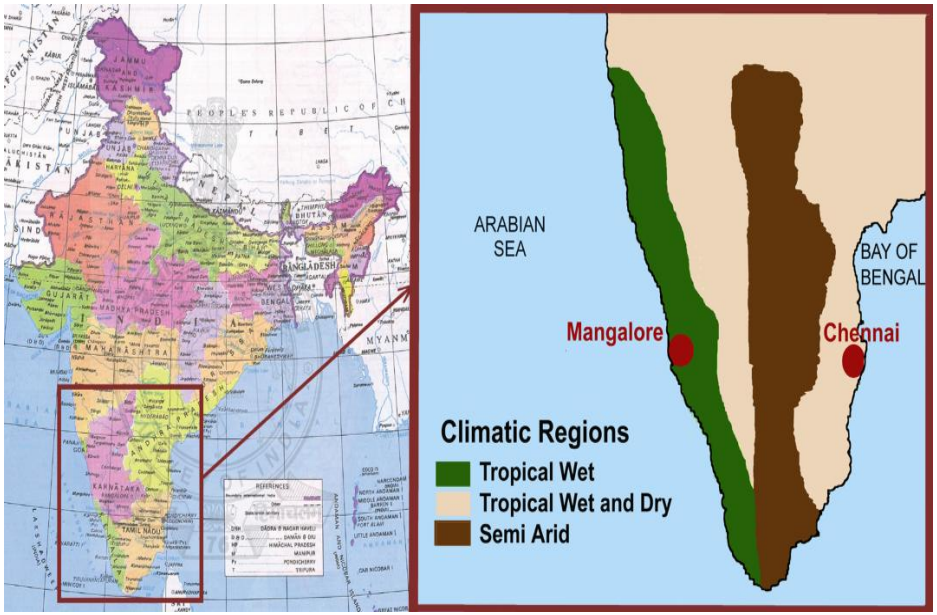
Study the relationships between malaria incidence and climatic and non-climatic factors –

Prevalence data for three to five years from two cities in southern India from two different climatic zones (varying strengths of disease prevalence and environmental conditions)

- Used the statistical technique of **Response Surface Method (RSM)** for the **first time** to study any epidemiological data
- **Proposed a new model reduction technique in combination with RSM to refine the model -** provides a simpler structure and gives better fit

Sayantani Basu Roy, Ram Rup Sarkar and Somdatta Sinha, *Malaria Journal* 2011, 10:301

Preliminary Data and some features



Similarities:

• Both are endemic for malaria since few decades, though Chennai has lower endemicity.

• Victims of rapid industrialization, construction activities

- problem of mosquito breeding in man-made clear water sources like wells, overhead tanks, sumps, cisterns as well as other defective and illegal drainage systems.

Several differences in demographic and climatic factors.

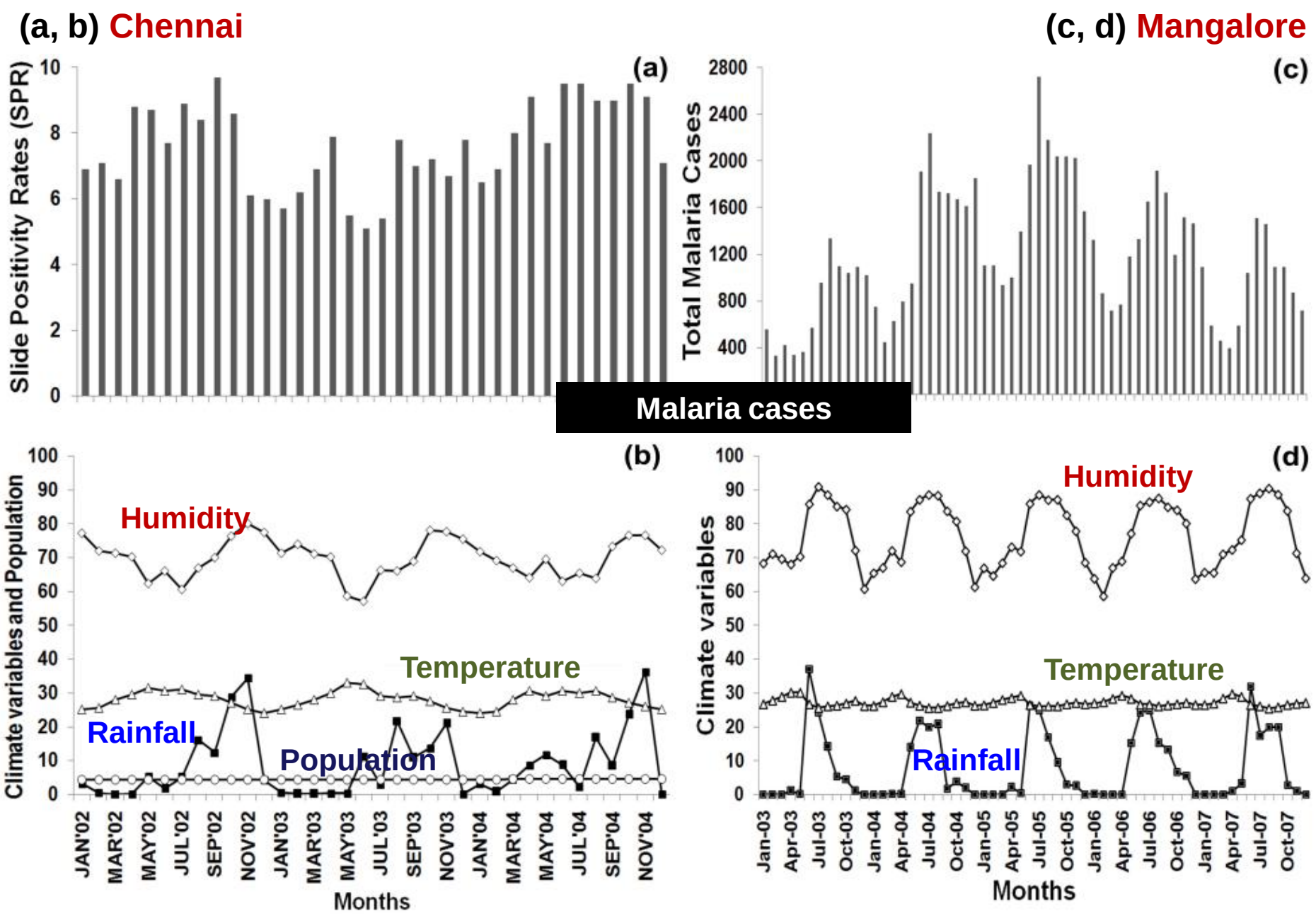
• **Chennai** – **Tropical Wet & Dry**; **Mangalore** – **Tropical Wet**

• **Mangalore** has less population (398,745) as compared to **Chennai** (5.6 million, third largest city) - *2001 census of India*

• Urban area in **Mangalore** had 32 recognised slums, and nearly 22,000 migrant labours lived in slums within the city limits,

whereas, **Chennai** has the fourth largest population of slum dwellers among major cities in India, with about 820,000 people (18.6% of its population) living in slum conditions

Malaria incidence and climate data



Data

CHENNAI:

- ✓ **No. of blood slides examined** (denoted, **A**),
- ✓ **No. of malaria cases** (denoted**, M**) from Jan. 2002 – Dec. '04 [monthly figures]

→ **Slide Positivity Rate values or SPR (= M/A)** : Dependent variable

Average Temperature, Average Humidity, Rainfall, Population → Independent variables

MANGALORE:

- ✓ **No. of *P. falciparum* and *P. vivax* cases** from Jan. 2003– Dec. '07 [monthly figures]

→ **Total Malaria cases or TMC** : Dependent variable

Average Temperature, **Average Humidity**, **Rainfall** → Independent variables

If the percentage of parasitaemia values are 2-3% or above, then they are considered to be positive slides for thin blood smear films.

The total malaria cases include both symptomatic and asymptomatic cases.

Temporal scale of SPR is obtained from the no. of blood smears collected every month by the Corporation of the city of Chennai.

*Total number lies in the range of **16306-63717** depending on the seasons, and the number of cases tested positive are in the range of **1184-4275**.*

Climate variables: data from daily weather sources - averaged to ensure compatibility with the other variables (SPR, TMC) at the monthly scale.

- ❑ **Exposure-outcome associations and confounding**
- ❑ **Confounding by other time-varying factors**
- ❑ **Allowing for delayed exposure effects**
- ❑ **Model checking and sensitivity analysis**
- ❑ **Precision and power considerations**

- ➢ **Non-linearity**
- ➢ **Investigation of effect modifiers**
- ➢ **Analysis of data from multiple locations**

  Separate analysis by location (e.g. specific cities) can increase power and provide information on heterogeneity and adaptation to environmental exposures

Statistical Methods used in our Study

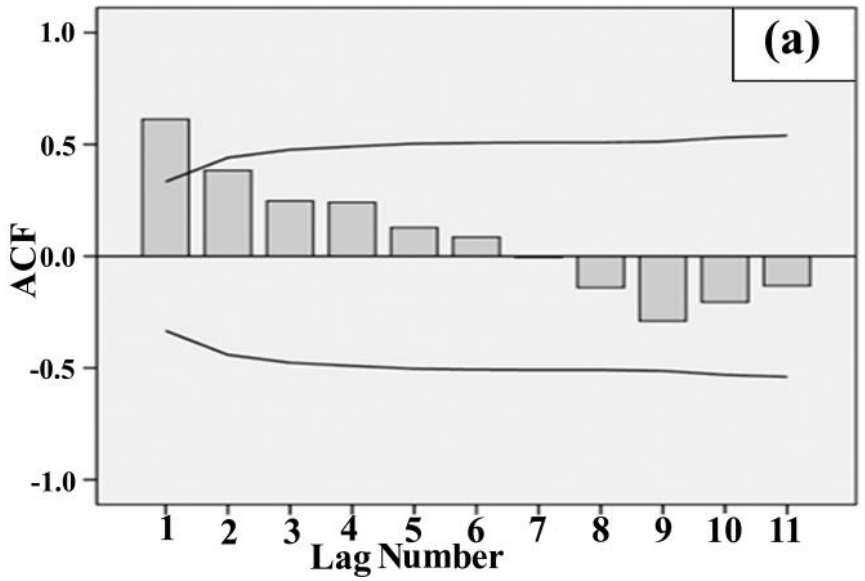
- **Data organisation and Preliminary analysis**
  - *Autocorrelation function*
  - *Correlation and Residual plots*
- **Model development using Response Surface Method**
- **Model Reduction Process**
- **Model Validation and Forecast**
  - *Confidence Intervals*
  - *Akaike's Information Criterion (AIC)*
  - *Prediction Intervals in Forecasting Method*
  - *Collinearity Analysis*

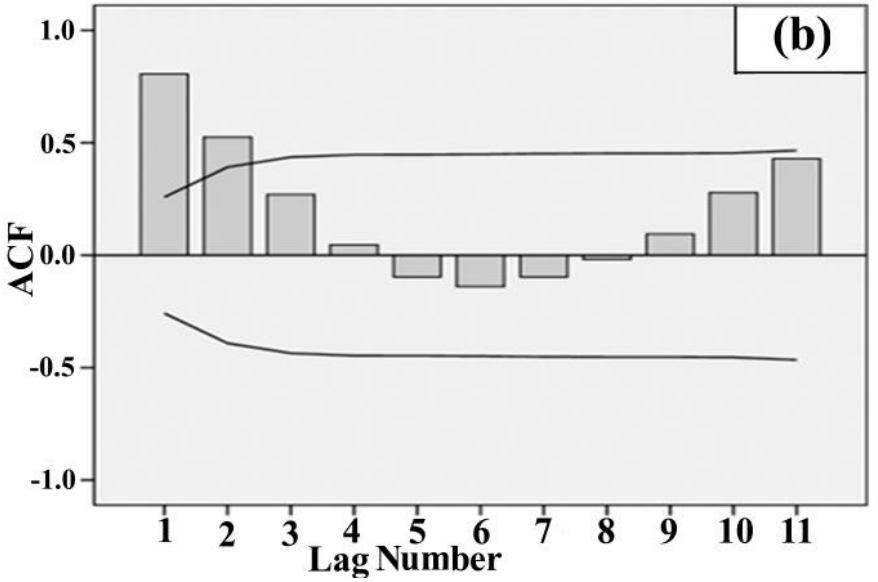
Preliminary Statistical Analysis

I. Autocorrelation function (ACF):

Measure of whether a variable is related with its own values as a function of a time difference

$$\rho_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n-k}(X_t - \bar{X})^2 \sum_{t=k+1}^{n}(X_t - \bar{X})^2}$$





a) *SPR values of Chennai from January 2002 to December 2004;*

b) *TMC of Mangalore from January 2003 to December 2007,*

all at Lags 1 through 11. The lines are for 95% confidence limits.

Significant autocorrelation values at lag one (95%).

>> SPR / TMC values as well as the Average Temperature at time lag one considered as additional independent variables in our model.

Ljung-Box Statistic:

This statistic gives the significance of autocorrelation at different lags, cumulatively.

Ljung-Box statistic for SPR values of Chennai and TMC values of Mangalore with p-values at 95% level of significance

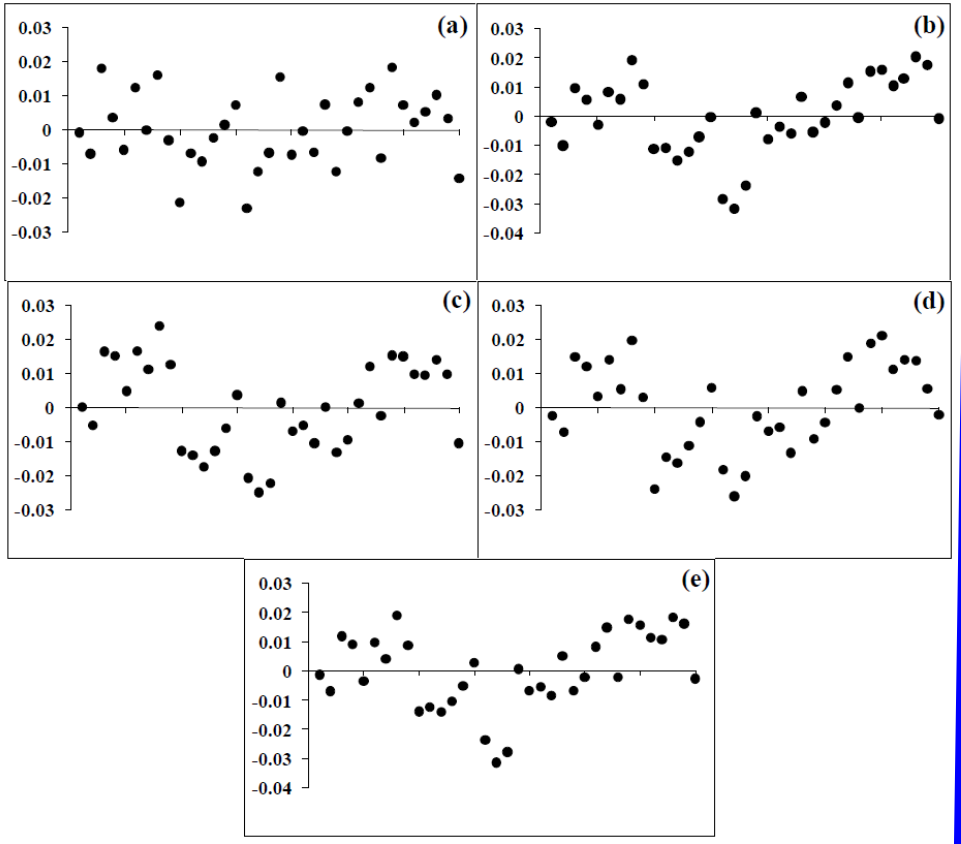
Lag↓	ACF_1	LB_1	p_1	ACF_2	LB_2	p_2
-1	0.613681	14.34421	0.000152	0.819723	38.96123	0.0001
-2	0.387528	20.23755	0.00004	0.535942	55.90804	0.0001
-3	0.285586	23.53816	0.000031	0.260731	59.99051	0.0266
-4	0.284155	26.91116	0.0205	0.013578	60.00178	0.0052
-5	0.147157	27.84595	**0.1671**	-0.14507	61.31252	**0.2084**
-6	0.115245	28.43903	**0.3014**	-0.18232	63.42165	**0.8515**
-7	-0.01087	28.44449	**0.9142**	-0.12218	64.38713	**0.5598**
-8	-0.19157	30.20476	**0.1287**	-0.0451	64.52124	0.0231
-9	-0.45511	40.52099	0.0003	0.083952	64.99529	0.0142
-10	-0.37242	47.70555	0.0043	0.31126	71.6446	**0.9259**
-11	-0.25344	51.17141	**0.0516**	0.503997	89.44137	0.0002

High correlations were observed between SPR and SPR-at-lag-one, and TMC and TMC-at-lag-one.

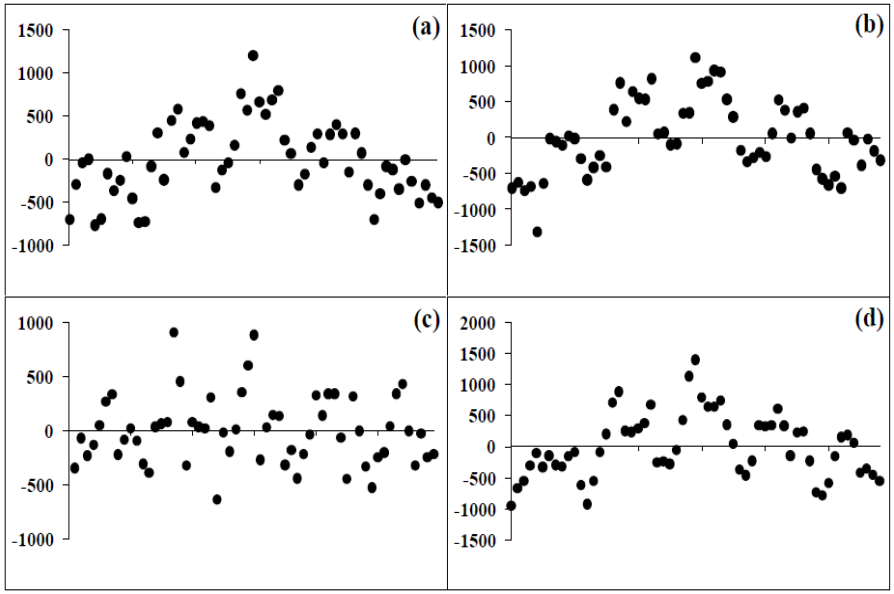
For other combinations of independent and dependent variables, where correlation was found to be low, non-linearity was confirmed by residual plots

Apart from SPR and SPR_{-1}, and also, TMC and TMC_{-1}, the correlation values of the other variables are not very significant, though they are known to be influential factors.

One possible reason for this is due to use of linear correlation measure to capture the relationship between dependent and independent variables.



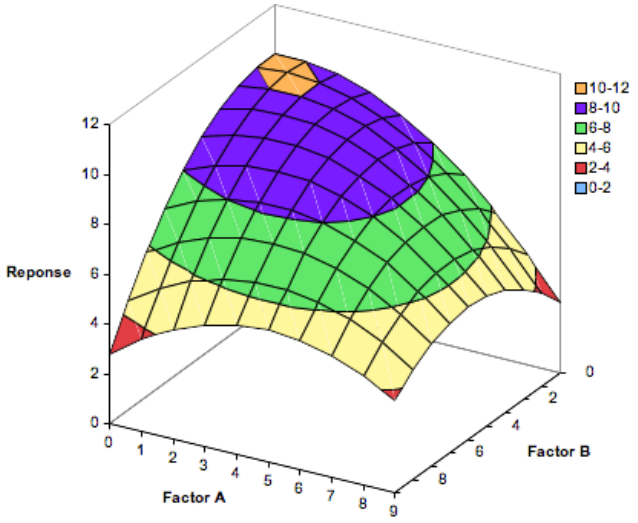
Residual plots for the dependent variable, SPR values of Chennai with the independent variables:
(a) SPR.-at-lag-one:SPR_{-1}; (b) Average Temperature:T; (c) Population:P; (d) Rainfall:R; and (e) Average Temperature-at-lag-one:T_{-1}



Residual plots for the dependent variable, TMC values of Mangalore with the independent variables:
(a) Average Temperature:T; (b) Rainfall:R; (c) TMC-at-lag-one:TMC_{-1}; and (d) Average Temperature-at-lag-one:T_{-1}

Response surface method ~ a regression outlook

- ✓ Introduced by Box & Wilson in 1951
- ✓ An experimental optimization procedure
- ✓ Easy to apply



- Technically RSM is applied to study the measured yield or output of a system as it varies in response to the changing levels of one or more physical input variables.
- In Dynamic Network analysis, RSM was found useful for sensitivity analysis of various DNA measures for different random errors.
- RSM is a sequential combination of statistical experimental design, regression modeling and optimization methods.
- RSM approach is a global approximation method which is ideally suited for solving problems with a relatively noisy response, where a gradient based method would lead to a local optimum instead of a global one.

RSM… the approach

- ✓ Dependent variable: **Y** **Slide Positivity Rates (SPR) values / TMC**
- ✓ Independent variables: $X_1, X_2, X_3, \ldots, X_n$

  SPR/TMC-at-lag-one, Average Temperature, Rainfall, Population, Average Temperature-at-lag-one
- ✓ First Order Model: $$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$$
- ✓ Second Order Model: $$Y = \beta_o + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n} \alpha_i X_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \beta_{ij} X_i X_j$$

- When the experimenter is relative closed to the optimum, the second-order model is used to approximate the response.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

- We find the stationary point. Subsequently we find the maximum response, minimum response or saddle point.
- We determine whether the stationary point is a point of maximum or minimum response or a saddle point.

The Model:

$$\hat{y} = \hat{\beta}_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x},$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \cdots & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k}/2 \\ & & \ddots & \\ & & & \hat{\beta}_{kk} \end{bmatrix}$$

$$\mathbf{x}_s = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}$$

** x_s *is the stationary point*

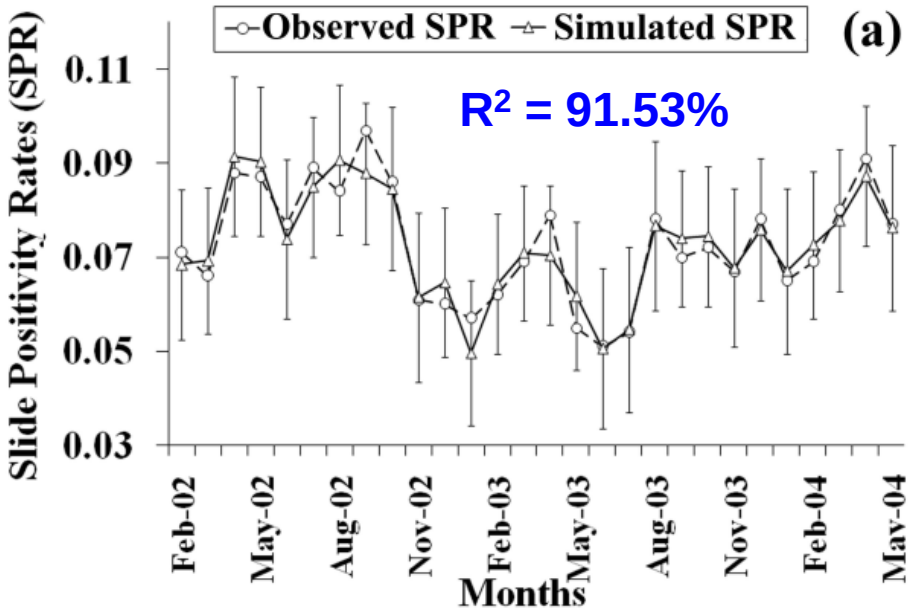
$$\hat{\mathbf{y}}_s = \hat{\beta}_0 + \frac{1}{2}\mathbf{x}_s'\mathbf{b}$$

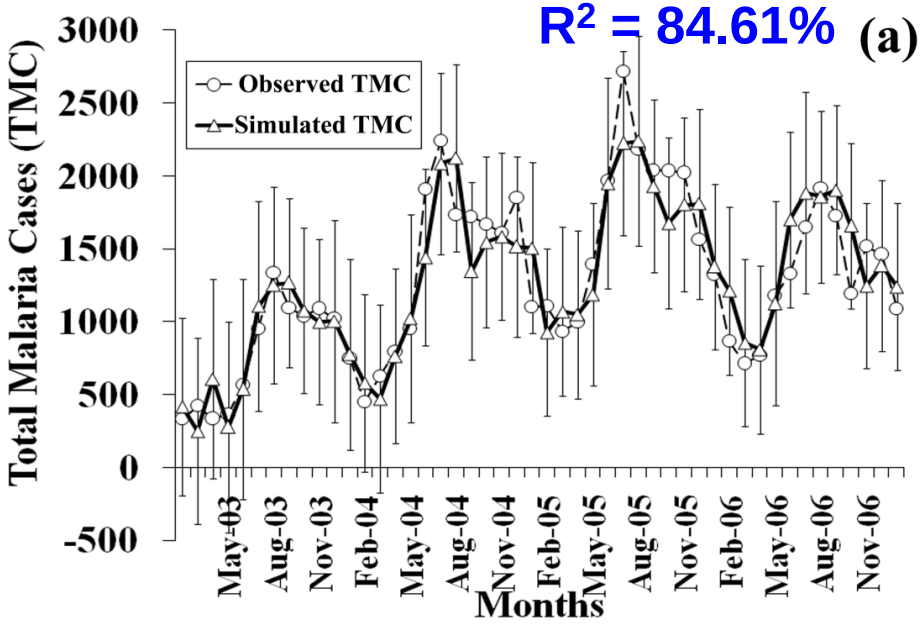
Canonical Analysis

If the second order model is found to be suitable then we resort to canonical analysis where we claim that **if all the Eigen values of the matrix B are positive (negative)**, then the quadratic surface **has minimum (maximum)** at the stationary point.

However if Eigen values are of both signs, then it is a saddle point.

The observed data and the model fit for Initial model using RSM





(Error bars show the 95% confidence intervals)

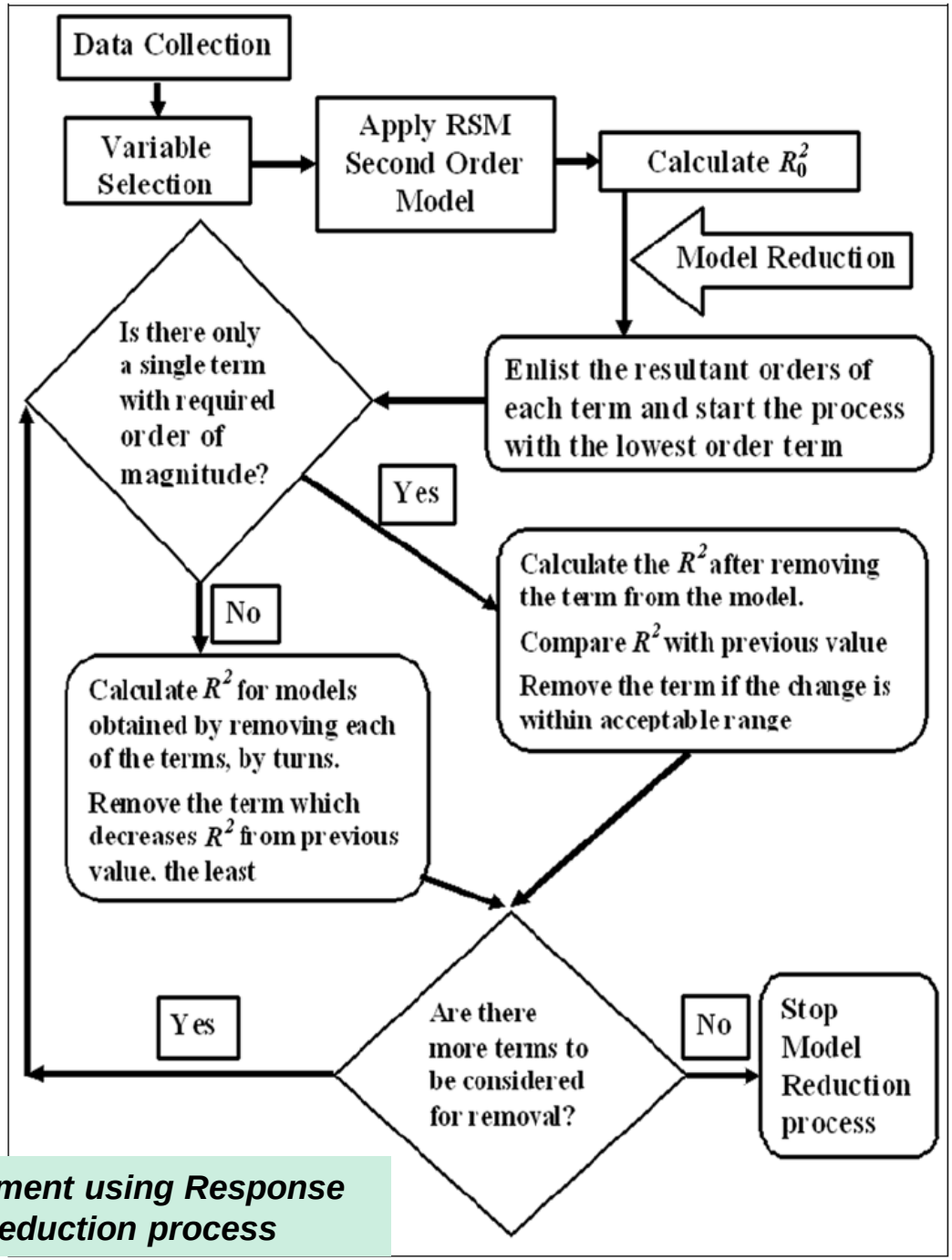
SPR = $115.25 - 3.53(SPR_{-1}) - 4.78\acute{}10^{-03}T$
$- 5.51\acute{}10^{-05}P + 6.75\acute{}10^{-03}R$
$+ 3.64\acute{}10^{-1}T_{-1} + 10.07(SPR_{-1})^2$
$- 3.61\acute{}10^{-04}T^2 + 6.52\acute{}10^{-12}P^2$
$- 3.48\acute{}10^{-07}R^2 - 1.63\acute{}10^{-03}(T_{-1})^2$
$- 8.35\acute{}10^{-02}SPR_{-1}\,T - 3.48\acute{}10^{-07}SPR_{-1}P$
$+ 1.09\acute{}10^{-03}SPR_{-1}R + 8.83\acute{}10^{-02}SPR_{-1}T_{-1}$
$+ 1.41\acute{}10^{-08}T\,P + 2.34\acute{}10^{-05}T\,R$
$+ 4.18\acute{}10^{-04}T\,(T_{-1}) - 1.57\acute{}10^{-09}P\,R$
$- 6.65\acute{}10^{-08}P\,(T_{-1}) - 2.02\acute{}10^{-05}R\,(T_{-1}).$

TMC = $6995.79 - 1087.87T$
$+ 4.03R - 1.44TMC_{-1} +$
$516.88(T_{-1}) + 40.03T^2$
$- 1.9\acute{}10^{-4}R^2 - 2.1\acute{}10^{-4}(TMC_{-1})^2$
$+ 6.08(T_{-1})^2 + 5.68\acute{}10^{-2}T\,R$
$- 5.64\acute{}10^{-2}\,TMC_{-1}T - 35.24TT_{-1}$
$- 2.6\acute{}10^{-04}\,TMC_{-1}\,R$
$- 1.59\acute{}10^{-1}R\,(T_{-1})$
$+ 1.67\acute{}10^{-1}TMC_{-1}\,T_{-1}.$

- Presence of more terms renders the model complicated
- Decrease its efficiency statistically as more parameters are estimated from limited data.

Model Reduction Process:

- **Implemented to derive a simpler model** with fewer but important terms,
- **Ensuring efficiency** (coefficient of determination) similar to that obtained in the initial model fitting through RSM.
- **A simple but useful approach, and has no underlying assumption** except that the contribution of a term in the model is directly proportional to its order of magnitude.
- **At each step of the model reduction, it is ensured that the predictions remain within the confidence limits that minimize error variance.**



Flowchart for model development using Response Surface Method and Model Reduction process

Model Reduction Steps for the Second Order Response Surface model for Chennai & Mangalore

	Terms Removed	R-square	Decision
Step 1	X4^2	0.9118	Remove
Step 2	X1*X4	0.909	Remove
Step 3			
3a	X2*X4	0.9054	Retain
3b	X4*X5	0.9089	Remove
Step 4			
4a	X4	0.8989	Remove
4b	X3*X4	0.8956	Retain
Step 5			
5a	X1^2	0.8937	Retain
5b	X1*X2	0.8989	Remove
5c	X1*X3	0.8954	Retain
5d	X1*X5	0.8984	Retain
Step 6			
6a	X1*X3	0.895	Retain
6b	X1*X5	0.8975	Remove

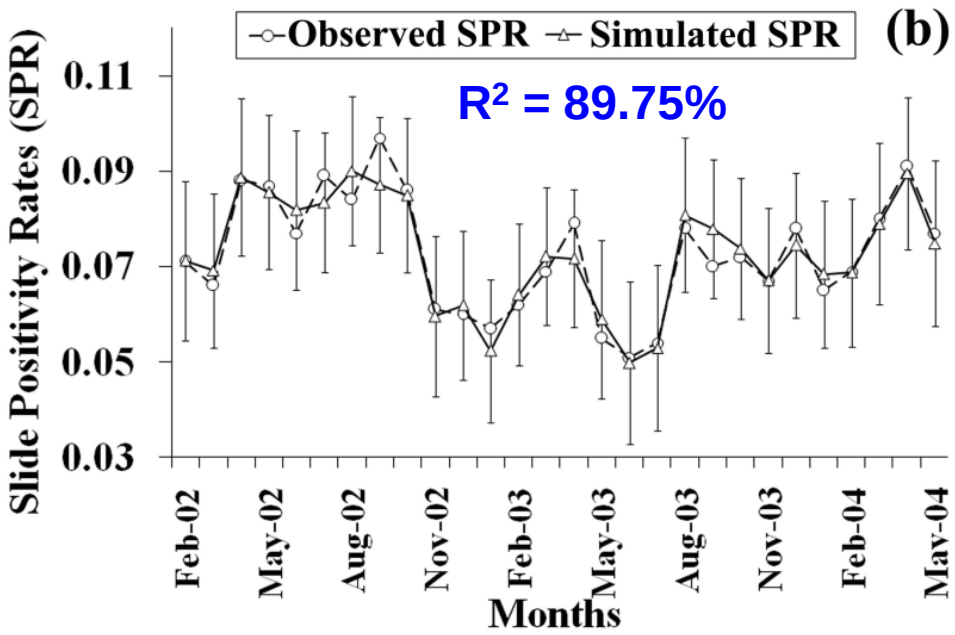
X1: SPR-at-lag-one or SPR_{-1}; X2: Average temperature or T; X3: Population or P; X4: Rainfall or R; X5: Average Temperature-at-lag-one or T_{-1}

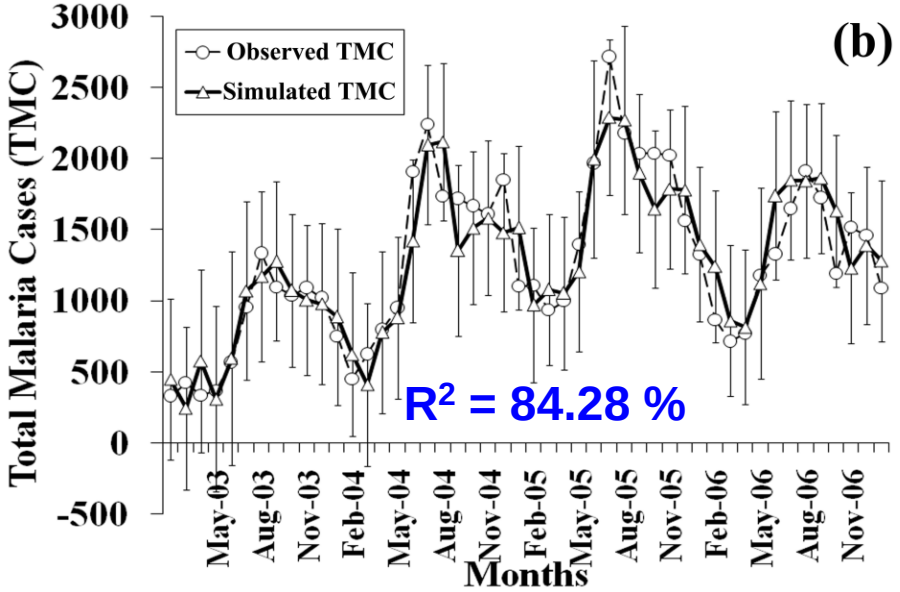
	Terms Removed/Added	R-square	Decision
Step 1	X2^2	0.8456	Remove
Step 2	X2*X3	0.8439	Remove
Step 3	X1*X2	0.8429	Remove
Step 4			
4a	X2*X4	0.8411	Remove
4b	X2	0.8401	Retain
Step 5	X1*X3	0.8411	Remove
Step 6	Add X1*X2	0.8428	Add (epidemiologically important, better fit)

X1: Average Temperature or T; X2: Rainfall or R; X3: TMC-at-lag-one or TMC_{-1};
X4: Average Temperature-at-lag-one or T_{-1}

Remove: If the change in R^2 is in the second or better, in the third place of decimal;

When for both choices available, the change is in the second or third place of decimal, choose the one that is reducing the R^2 less.





SPR = $75.96 - 20.40(SPR_{-1}) - 1.31\acute{}10^{-01}T - 3.71\acute{}10^{-05}P + 5.30\acute{}10^{-1}(T_{-1}) + 8.474(SPR_{-1})^2 - 1.07\acute{}10^{-04}T^2 + 4.46\acute{}10^{-12}P^2 - 1.21\acute{}10^{-03}(T_{-1})^2 + 4.4\acute{}10^{-06}SPR_{-1}\,P + 3.28\acute{}10^{-08}T\,P + 2.56\acute{}10^{-05}T\,R + 2.402\acute{}10^{-04}T\,(T_{-1}) - 1.62\acute{}10^{-10}P\,R - 1.04\acute{}10^{-07}P\,(T_{-1})$.

Removes six terms from the initial model (**about 30% reduction in coefficients**) compromising only less than **2% decrease in the value of R^2**.

TMC = $9800.71 - 1077.33T + 3.43R - 2.30TMC_{-1} + 348.61(T_{-1}) + 18.64T^2 - 2.2\acute{}10^{-4}(TMC_{-1})^2 + 9.958(T_{-1})^2 - 4.378T\,(T_{-1}) - 1.1004R\,(T_{-1}) + 1.40\acute{}10^{-1}\,TMC_{-1}T_{-1}$.

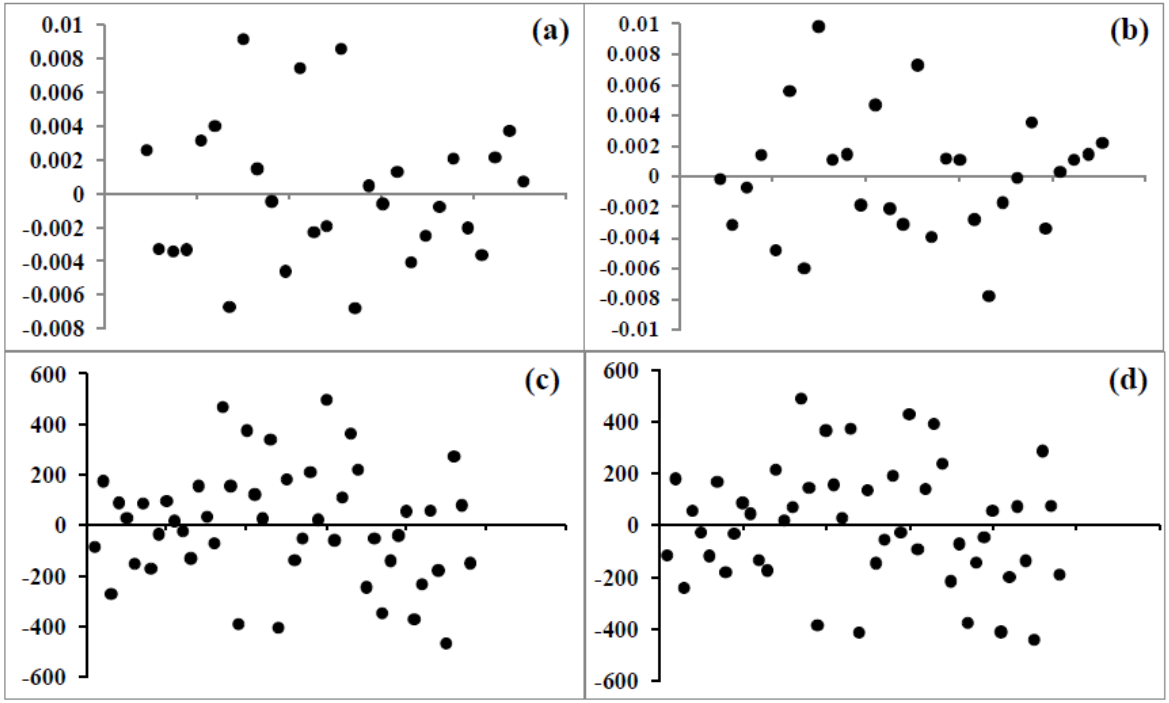
Reduces the number of terms from 15 to 11 **(~27% reduction)**, and hardly reduces the R^2 from 84.61% to 84.28% (**only by 0.4%**).

Confidence Intervals

Tested each estimated value of the data points used for fitting the model (both for the optimized and reduced models) - lies within the 95% confidence intervals (Gauss Markov Theorem)

To assess the linear accuracy, homoscedasticity and to observe the differences after removing lower order variables,

Residual plots for the initial and reduced models have been plotted for both SPR and TMC values



Residual plots.
For the dependent variable, SPR values of Chennai from January 2002 to December 2004:
(a) Initial model; (b) Model after Reduction;
For response variable, TMC values of Mangalore from January 2003 to December 2007:
(c) Initial model; and (d) Model after Reduction

There is slight compression of the residual cloud from the initial to the reduced model for SPR (Figure (a), (b)) as well as TMC (Figure (c), (d)).

The spread in general is uniform on either side of the x-axis which indicate homoscedasticity.

Collinearity Analysis

For least square regression approach, it is imperative for the residuals to be able to adhere to the assumption of constant variance (homoscedasticity).

Two different measures are considered, namely, **the Variance Inflation Factor** and **Breusch-Pagan test** , for analyzing this property shared between different cofactors in the models

- ascertains that there is no significant evidence that the assumptions underlying the fitting and estimation methods are violated.

- The homoscedastic null hypothesis is not rejected at 99% level of significance.
- This reinstates that the residuals show no erratic behavior as a function of the value of the responses.
- It can be assumed that there is no significant level of heteroscedasticity.
- Provides a comprehensive reason for the use of our model fitting approach for the available data.

Table A7 (a) *Variance Inflation Factor for the initial model for SPR values of Chennai from January 2002 to December 2004*

Independent variable	Statistic	Decision
SPR_{-1}	1.2934	No *significant* collinearity
T	2.0716	No *significant* collinearity
T_{-1}	1.8135	No *significant* collinearity
P	4.2158	No *significant* collinearity

The values of VIFs in our case clearly indicate that there are no values beyond 5, which shows the absence of multi-collinearity.

Hence, we can conclude that the chosen independent variables in both data, do not lead to violation of the assumption of constant variance of the residuals.

T	4.0209	No *significant* collinearity
T_{-1}	2.7855	No *significant* collinearity
R	2.172	No *significant* collinearity

TMC_{-1}:TMC-at-lag-one; T: Average Temperature;T_{-1}: Average Temperature-at-lag-one; R: Rainfall;

Akaike's Information Criterion (AIC)

This statistic measures the appropriateness of forecasts of the estimated statistical models and selects the better model from the given models

Lower the AIC, the better the model

$$AIC = (-2)max(log(likelihood) + 2k$$

where, k: no. of parameters; n: no. of observations

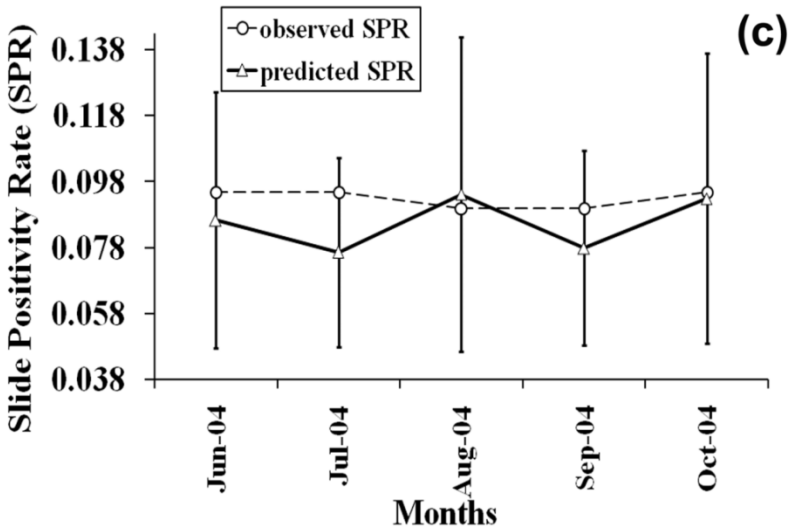
and, ϵ_i : residual for i^{th} observation .

	Chennai	Mangalore
Initial Second Order Model	3827.692	3812.048
Reduced Second Order Model	3410.028	3359.722

Predicted and Observed values

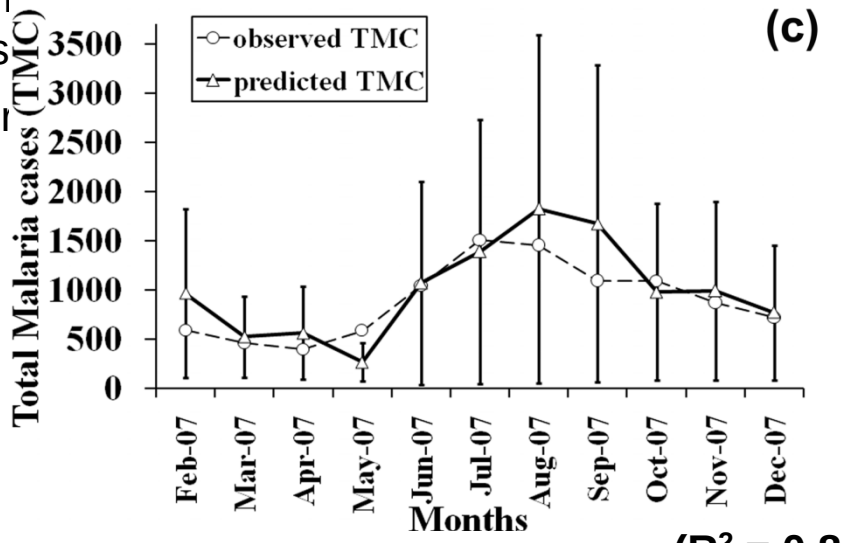
Incidence values (SPR or TMC) have been predicted for those time points that have not been used for the purpose of fitting the model (remaining 20% time points)

- Model incorporates auto-regressive terms (SPR-at-lag-one for SPR values, TMC-at-lag one for TMC values),



(R^2 = 0.8975)

Predicted SPR values for Chennai from June 2004 to October 2004



(R^2 = 0.8428)

Predicted Total Malaria cases' values for Mangalore from February 2007 to December 2007

Chi-square goodness of fit is also found to be highly statistically significant (p-value < < 0.001)

Predictions do not show any clear trend for the disease to decrease **- show more monthly periodic variations.**

Model predictions show initial decay in the total malaria cases from February to May, then an upward trend till August, and finally decreasing again till December **- following the pattern of rainfall distribution of that region**

Summary

❑ A combination of statistical modelling approach (the RSM) and a simple model reduction method is applied to describe the incidence of malaria in two geographically, ecologically and demographically different cities in India

Chennai, Tamil Nadu and **Mangalore, Karnataka**

❑ Two types of epidemiological data - Slide Positivity Rates (SPR) values of malaria for Chennai and Total Malaria cases (TMC) for Mangalore - shows the applicability of the algorithm to different measures of incidence of malaria.

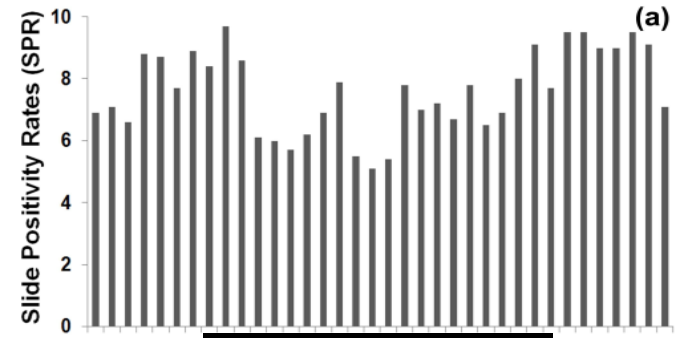
❑ Analyses of the data using the RSM approach not only capture the essential dynamics of the disease incidence, but also show the influence of different climatic and non-climatic factors.

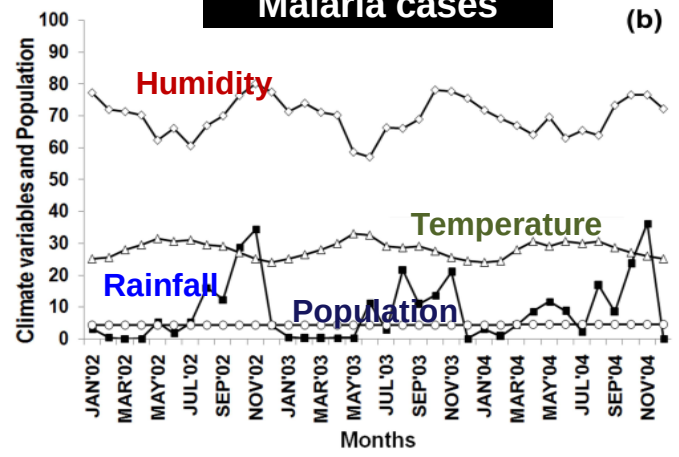
❑ Results show **the autoregressive nature of forecasting** in the long-range time series study, where **Rainfall, one-lag SPR and Temperature at lag one - values play an influential role and can be useful for better prediction**

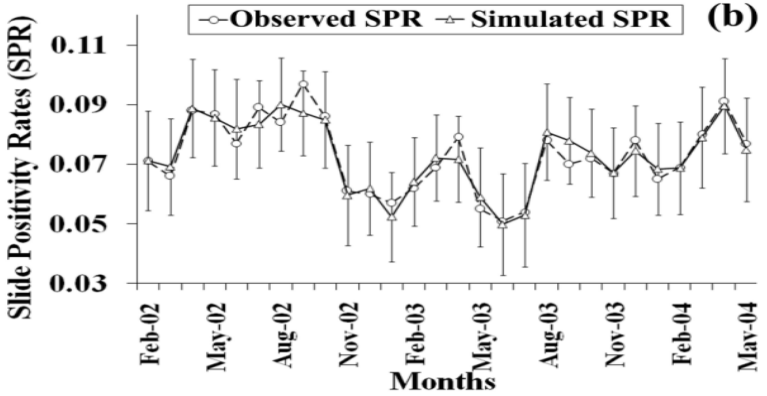
❑ **Some interesting effects of the climatic factors on the disease dynamics**

Sayantani Basu Roy, Ram Rup Sarkar and Somdatta Sinha: ***Malaria Journal*** **2011**, 10:301

Chennai







- **Complex influence of environmental variables on the epidemiological status of the region**
- **Variation in SPR (original data) values do not follow exactly the rainfall distribution in this region**, which clearly has two peaks every year due to North-East and South-West monsoons.
- **The ups and downs of the SPR values follow the temperature distribution broadly.**
- Pattern is also captured by the **RSM model, which shows more dependence of SPR values on temperature**, either through direct presence of linear and higher order terms, or through interactions with other associated factors including rainfall **- more cross interaction terms in Chennai**

Chennai being in the tropical wet-dry climatic region shows sustained prevalence of malaria through-out the year with less periodic variations: Variations in the SPR values are associated with more cross interaction terms.

$$SPR = 75.96 - 20.40(SPR_{-1}) - 1.31\acute{}10^{-01}T - 3.71\acute{}10^{-05}P + 5.30\acute{}10^{-1}(T_{-1}) + 8.474(SPR_{-1})^2 - 1.07\acute{}10^{-04}T^2 + 4.46\acute{}10^{-12}P^2 - 1.21\acute{}10^{-03}(T_{-1})^2 + 4.4\acute{}10^{-06}SPR_{-1}\,P + 3.28\acute{}10^{-08}T\,P + 2.56\acute{}10^{-05}T\,R + 2.402\acute{}10^{-04}T\,(T_{-1}) - 1.62\acute{}10^{-10}P\,R - 1.04\acute{}10^{-07}P\,(T_{-1})$$

Mangalore

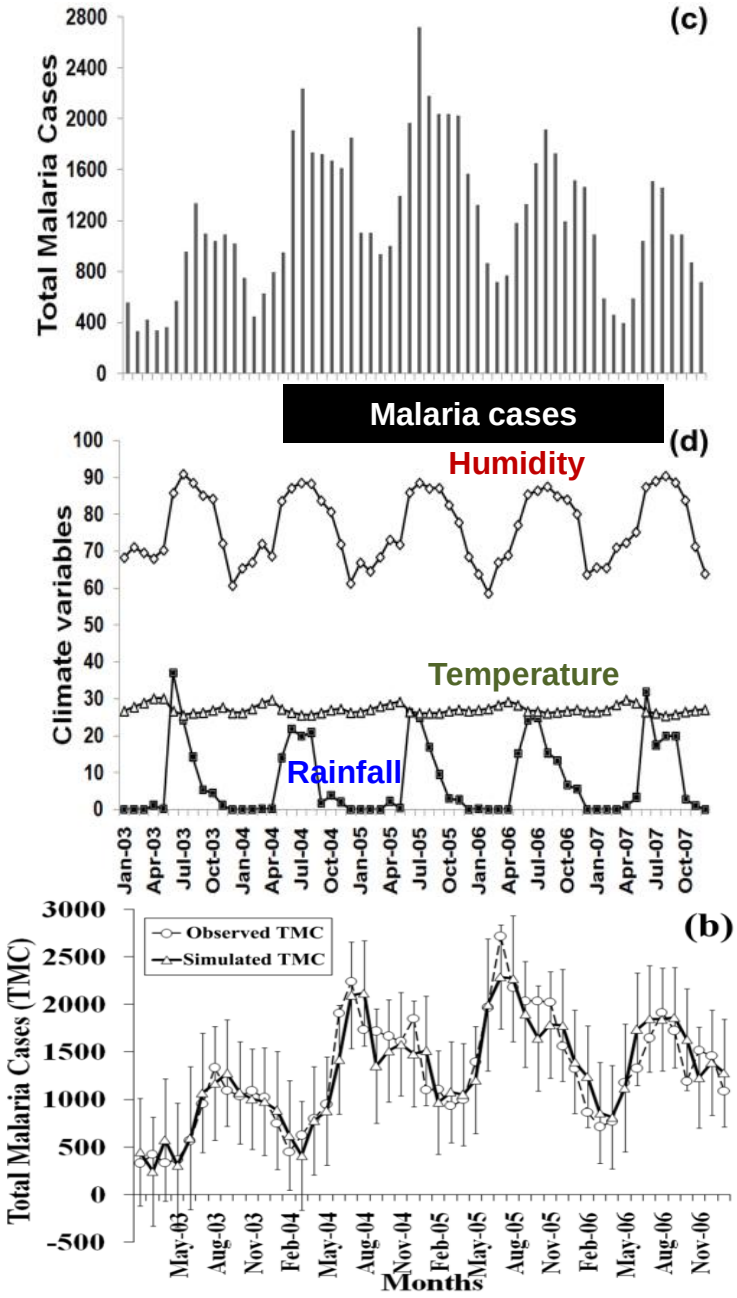
- **TMC follows a cyclic pattern each year with gradual increase leading to the peak during June - August**, followed by a gradual decrease, **quite similar to the rainfall distribution in this region due to the direct influence of the Arabian Sea branch of the South-West monsoon.**

- **This specific trend is mimicked by the RSM model using the model reduction technique.**

- **Model equation shows direct dependence on rainfall term and captures the essential dynamics of the disease along with other factors**

  - presence of more direct terms (linear or higher-order) and less cross interaction terms.

Mangalore is in tropical dry climatic region - shows more periodic variations in TMC values following the climatic changes in the region: Observed to be directly influencing the model formulation.



$$TMC = 9800.71 - 1077.33T + 3.43R - 2.30TMC_{-1} + 348.61(T_{-1}) + 18.64T^2 - 2.2\acute{}10^{-4}(TMC_{-1})^2 + 9.958(T_{-1})^2 - 4.378T\,(T_{-1}) - 1.1004R\,(T_{-1}) + 1.40\acute{}10^{-1}\,TMC_{-1}T_{-1}.$$

Summary

❑ **RSM has been used for the first time in epidemiological studies**

❑ **Second Order model given by the RSM approach, in spite of not having terms higher than degree 2, still represents the temporal variations of both datasets very well.**

- **Major challenge is to gain the detailed understanding of the diff. environmental factors, prev. incidence of the disease and their influence on the prevalence pattern, a reduced model with fewer terms was helpful for reliable predictions.**
- Model predictions capture the climatic variations, mainly rainfall and temperature, for both the regions and resembles well with the observed disease incidence.
- **This approach leads to detection of the most crucial environmental factors influencing the transmission of the disease while offering a coherent and integrated understanding of the disease process in any area.**
- **The major advantage is –**

  No prior assumptions about the disease or knowledge of parameters required - Only the previous occurrence of the disease and suitable climatic factors are required to feed into the method;

  Gives **a simple, but highly predictive model for malaria incidence** without compromising on the proportion of variation represented
- **May be useful for adopting better strategy not only to control malaria but for other infectious diseases, if suitable climatic information and disease prevalence data are available**

Few Questions…..

What about the spread of infection within a population?

Population structure & Age groups?

Different types of infections – asymptomatic or symptomatic?

Host immunity?

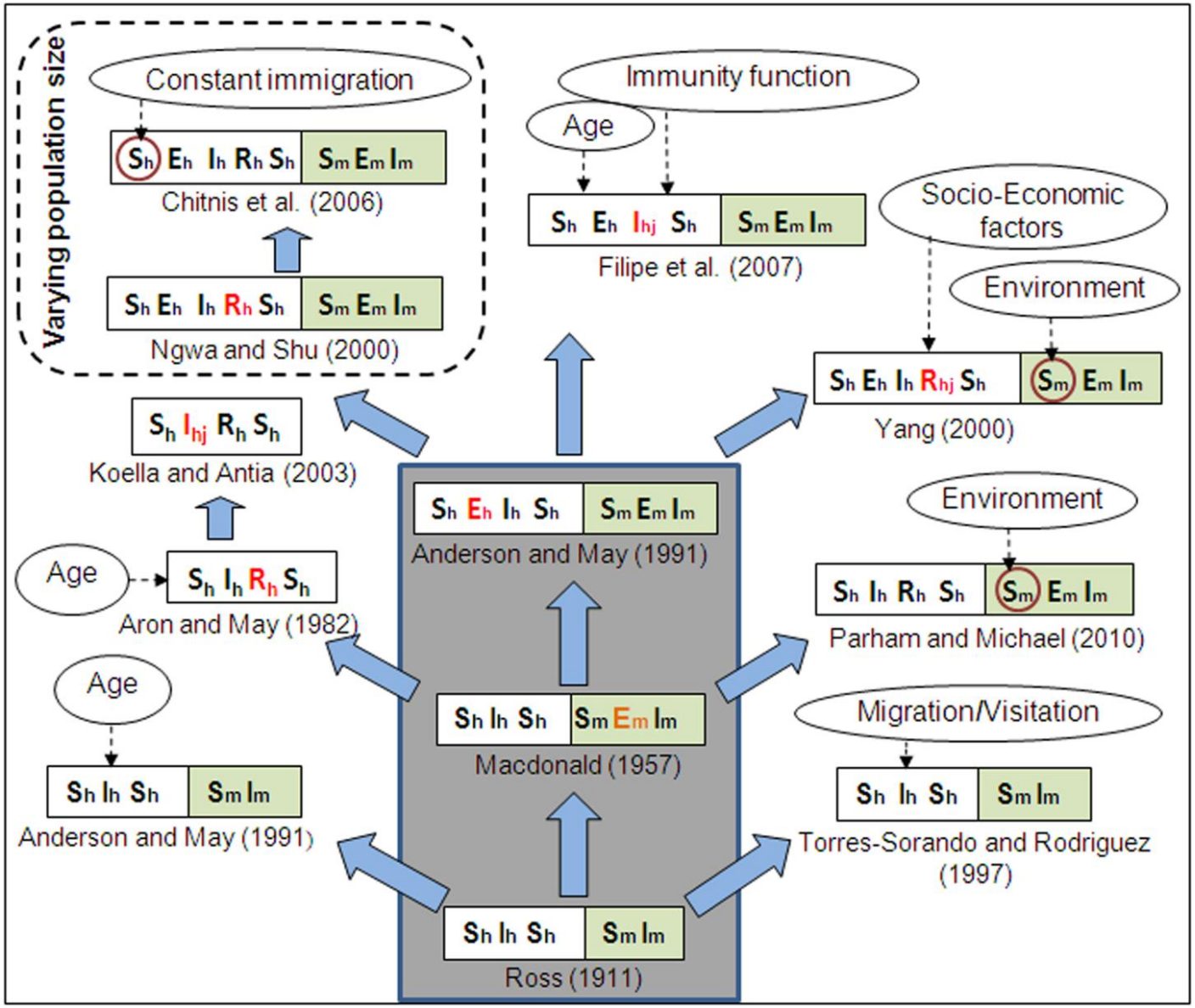
Need for a model which can capture the Interactions between different general and local processes for understanding disease prevalence pattern

- **Dependence on age and immunity of the human host**
- **Continued presence of asymptomatic infections**
- **Regulation of the host-vector interactions**

And

Variations of temperature and rainfall in a particular region

Evolution and grouping of different types of malaria models



S.Mandal, R.R.Sarkar, S.Sinha, *Malaria Journal*, 2011

Different important assumptions in our model

Immunity functions

Human immunity is considered to act in two ways: **(i) reduces the probability (η) of clinical disease,** and **(ii) regulate the rate of removal of parasites (ρ).**

Clinical immunity ($I_c(a, t)$) [some level of maternal immunity inherited by the new born (I_m), which decays exponentially with a half life d_m but simultaneously accumulates some immunity (I_f), which is determined by the force of infection and this also decays exponentially with mean life d_e]

$$I_c(a, t) = I_m(a, t) + I_f(a, t)$$

Parasite immunity ($I_p(a, t)$) [responsible for clearance of parasite and develops / matures relatively in latter age of life through a delay phase (J_A). The immunity level in delay phase (J_A) accumulates through force of infection, matures in parasite immunity ($I_A(a, t)$) at a rate ($1/d_l$), and decays exponentially with mean life d_A]

Response of the clinical immunity η

$$\eta = \frac{1}{1+\left(\frac{I_C(a,t)}{H_s}\right)^2}$$

Recovery rate from detectable parasite (ρ) is assumed to be a saturating increasing function of the immunity level $I_A(a, t)$:

$$\rho = r + \rho_0\left[1+(w_A-1)\frac{\left(I_A/H_A\right)^2}{1+\left(I_A/H_A\right)^2}\right]$$

Development of the parasite in host (human) and vector (mosquito)

- The rate (σ) at which proportion of individuals in the susceptible class (S_h) transfer to the exposed class (E_h), on receiving inoculation from infected mosquito, depends both on the number of infectious bites, and its ability to produce infection in the human host.

  If *m* is the density of mosquitoes per human, *α*, the biting rate, and *b*, **the maximum probability of inoculation in a human upon the infectious bite of a mosquito**, which is age-dependent, then

$$\sigma = m\,\alpha\,b\,(1-e^{-a/a_0})$$

  where, *a* is the age of the human host and a_0 is a constant represents the age at which half the total increase in exposure is achieved.

- Individuals remain in the exposed class (E_h) for a mean duration *1/γ* (time taken for transformation of sporozoite into merozoite inside human liver).

- Sporogony period of parasite in mosquito gut varies on the species of the parasite. The **parasite extrinsic latent period** of the parasite, in mosquito is considered as τ (**Delay Factor**).

The **probability of mosquito survival during the extrinsic latent period:**

$$\psi = e^{-\tau\,\mu_m}$$

Asymptomatic Infections

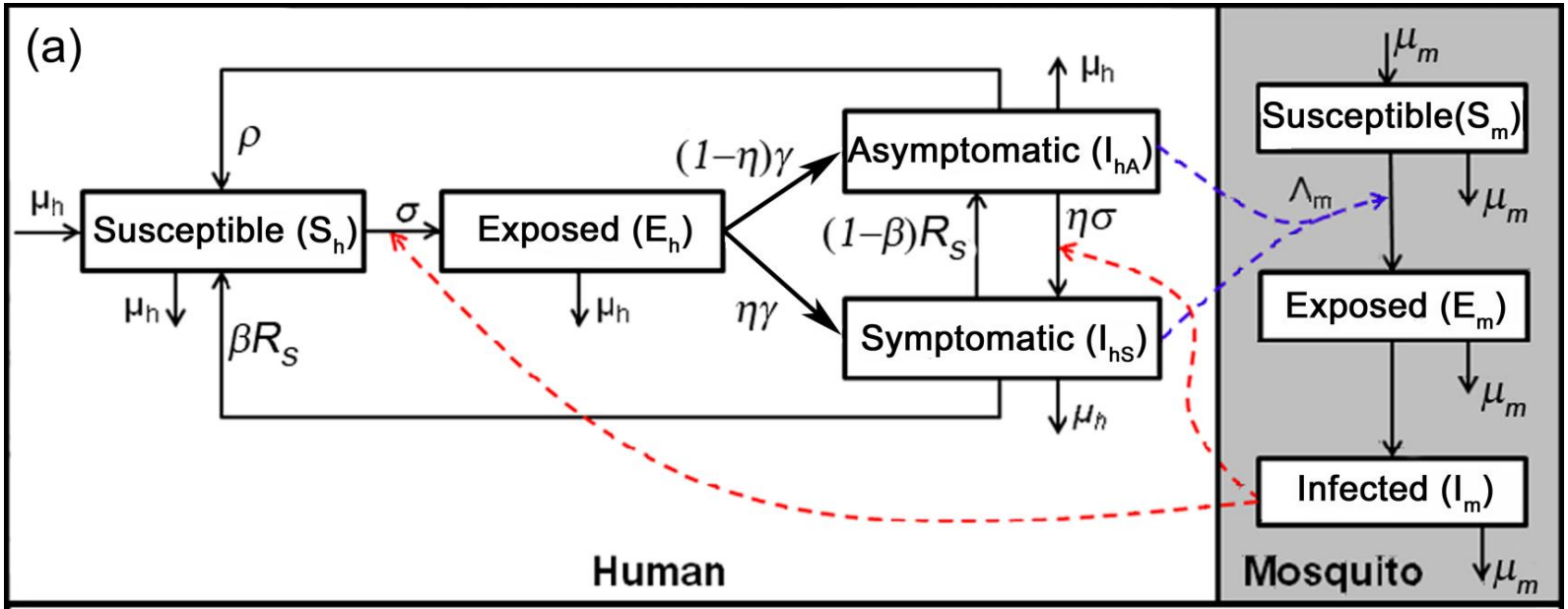
- Depending on the immunity of the human host, some proportion (η) of the exposed class produces symptomatic disease and the rest produce asymptomatic disease

- Also, population in asymptomatic class may transfer to symptomatic class, through re-infection, at a rate $\eta\sigma$.

- Asymptomatic individuals, by clearing parasites, can become susceptible again at a rate ρ.

- Susceptible mosquitoes become infected on biting symptomatic or asymptomatic patients.

Variations in temperature and rainfall of a particular geographical region



(b)

μ_h	: Per capita death rate, and the population level birth rate of human
σ	: Rate of transformation from Susceptible to Exposed class in human
η	: Susceptibility to clinical disease
γ^{-1}	: Mean duration in the Exposed class of human
β	: Probability of direct recovery from symptomatic disease
ρ	: Rate of recovery from asymptomatic disease
R_S	: Rate of recovery from symptomatic disease
μ_m	: Per capita death rate, and the population level birth rate of mosquito
Λ_m	: Rate of transformation from Susceptible to Exposed class in mosquito

Transmission of disease from mosquito to human is shown by red dotted line

The transmission from human to mosquito is shown by blue dotted line.

Arrows indicate transition among compartments

For simplicity, Birth rate and death rate (μ_h for human and μ_m for mosquito) are considered such that the total population remains constant and is normalised to one.

$$N_h = S_h + E_h + I_{hS} + I_{hA} = 1$$

$$N_m = S_m + E_m + I_m = 1$$

S.Mandal, S. Sinha, and R R Sarkar* (2013): *Bulletin of Mathematical Biology* 75: 2499-2528,

The Malaria Transmission Model

Human [at age "a"]

$$\frac{dS_h}{dt} = \mu_h + \beta\ R_S I_{hS} + \rho I_{hA} - \sigma\ I_m S_h - \mu_h\ S_h$$

$$\frac{dE_h}{dt} = \sigma\, I_m S_h - \eta\gamma\, E_h - (1-\eta)\,\gamma\, E_h - \mu_h\, E_h$$

$$\frac{dI_{hS}}{dt} = \eta\gamma\, E_h + \eta\sigma\, I_m I_{hA} - \beta\, R_S\, I_{hS} - (1-\beta)\, R_S\, I_{hS} - \mu_h\, I_{hS}$$

$$\frac{dI_{hA}}{dt} = (1-\eta)\,\gamma\, E_h + (1-\beta) R_S I_{hS} - \eta\sigma\, I_m\, I_{hA} - \rho\, I_{hA} - \mu_h\, I_{hA}$$

Mosquito

$$\frac{dS_m}{dt} = \mu_m - \Lambda_m\ S_m - \mu_m\ S_m$$

$$\frac{dE_m}{dt} = \Lambda_m\ S_m - \psi\ \Lambda_m(t-\tau)\ S_m(t-\tau) - \mu_m\ E_m$$

$$\frac{dI_m}{dt} = \psi\ \Lambda_m(t-\tau)\ S_m(t-\tau) - \mu_m\ I_m$$

Initial conditions: Defined in the seven dimensional variable space

$$C_+ = \left\{\left(S_h(\theta), E_h(\theta), I_{hS}(\theta), I_{hA}(\theta), S_m(\theta), E_m(\theta), I_m(\theta)\right) \in C\left([-\tau, 0],\ R^7_{0+}\right)\right\},$$

where $S_h(0) > 0$, $E_h\ (0) \geq 0$, $I_{hS}\ (0) \geq 0$, $I_{hA}\ (0) \geq 0$ and $S_m(0) > 0$, $E_m\ (0) \geq 0$, $I_m\ (0) \geq 0 \in$ C([- τ, 0]); C([- τ, 0], R^7_{0+}) is the space of continuous functions and is a mapping from [- τ, 0] to R^7_{0+} (R^7_{0+} = {(S_h, E_h, I_{hS}, I_{hA}, S_m, E_m, I_m) $\in \Re^7 \mid S_h, S_m > 0$; E_h, I_{hS}, I_{hA}, E_m, $I_m \geq 0$}).

Theoretical Study

Positivity and Boundedness

We observe that the right-hand-side of the equations are smooth functions of the variables (S_h, E_h, I_{hS}, I_{hA}, S_m, E_m, I_m) and the parameters, as long as these quantities are non-negative, **hence local existence and uniqueness properties** hold in R^7_+.

For constant age, if $S_h > 0$, then $dS_h/dt \geq 0$ for $I_{hS} \geq 0$, $I_{hA} \geq 0$, $I_m \geq 0$. Hence $S_h > 0$, $\forall t$. It is also true for the other variables E_h, I_{hS}, I_{hA}, S_m, E_m and I_m in the space C_+.

Proposition 1: All the solutions which initiate in R^7_+ are uniformly bounded.

Lemma 1: The system has exactly one unique equilibrium point $E_0^* = (S_h^*, E_h^*, I_{hS}^*, I_{hA}^*; S_m^*, E_m^*, I_m^*) = (1, 0, 0, 0; 1, 0, 0)$, which is disease free.

Lemma 2: The endemic equilibrium point ($E^* = (S_h^*, E_h^*, I_{hS}^*, I_{hA}^*; S_m^*, E_m^*, I_m^*)$) of the model exists for $R_0 > 1$ and it is unique.

Persistence of disease in a community is determined by the parameter ***R_0* – the basic reproduction number**.

R_0 is defined as the number of secondary infections that one infectious individual would create in a susceptible population over the duration of the infectious period.

In the mathematical theory of epidemics this is calculated as the dominant eigen value *r(K)* of the next-generation operator, *K* (Diekmann et al. 1990; Dietz 1993; Heffernan et al. 2005).

For the transmission model (2.7), R_0 can be calculated as $R_0 = r(FV^{-1})$

where *F* is a matrix reflecting the rate at which new infections arise, and *V* is a matrix reflecting the rate at which individuals enter or leave the infection classes.

$$R_0 = \frac{\alpha^2 \gamma m b (1 - e^{-\frac{a}{a_0}}) \psi [\eta C_{IS} (\rho + \mu_h) + C_{IA} \{(1 - \eta^2\} R_S + (1 - \eta) \mu_h]}{(\gamma + \mu_h)(\rho + \mu_h)(R_S + \mu_h) \mu_m}$$

OR

$$R_0 = \frac{\alpha \gamma \sigma \psi [\eta C_{IS} (\rho + \mu_h) + C_{IA} \{(1 - \eta^2\} R_S + (1 - \eta) \mu_h]}{(\gamma + \mu_h)(\rho + \mu_h)(R_S + \mu_h) \mu_m}$$

Stability of the system around the disease free equilibrium point (E_0^*)

Theorem 1: For $\tau = 0$, the system is stable around the disease free equilibrium point E_0^* if the following conditions hold true:

(i) $R_0 < 1$

and

$$\alpha\sigma < \frac{(\gamma+\mu_h)(\rho+\mu_h)(R_S+\mu_h)+[(\gamma+\mu_h)(\rho+\mu_h)+(R_S+\mu_h)(\gamma+\rho+2\mu_h)]\mu_m}{\gamma[(1-\eta)C_{IA}+\eta C_{IS}]}$$

or,

(ii) $R_0 < (\mu_h/\mu_m)$.

Stability of the system around the endemic equilibrium point (E^*)

Theorem 2: For $\tau = 0$, the system around the endemic equilibrium point E^* is stable if the following conditions hold true:

$$\alpha S_m^* \sigma[\gamma S_h^* C_{IA} + \eta C_{IS}\{\gamma(S_h^* + I_{hA}^*) + I_m^* I_{hA}^* \sigma + I_{hA}^* \rho + 3I_{hA}^* \mu_h\}] < Min[X_1, Y_1]$$

and

$$\begin{aligned}\alpha S_m^* \sigma[&\gamma S_h^* C_{IA}(R_S + 2\mu_h) + \eta C_{IS}\{I_m^* \gamma(S_h^* + I_{hA}^*)\sigma \\ &+ 2\mu_h(\gamma(S_h^* I_{hA}^*) + I_m^* I_{hA}^* \sigma) \\ &+ 3I_{hA}^* \mu_h^2 + \rho(\gamma(S_h^* + I_{hA}^*) + I_m^* I_{hA}^* \sigma + 2I_{hA}^* \mu_h)\}] < Min[X_2, Y_2]\end{aligned}$$

Theorem 3: The necessary and sufficient conditions for the system to be locally asymptotically stable around the steady state (E^*) for all $\tau > 0$ are:

(i) real part of all roots of $\Delta(\lambda, 0) = 0$ is negative

(ii) for all real ω_0 and $\tau \geq 0$, $\Delta(i\omega_0, \tau) \neq 0$.

Numerical Simulation

Time course and effect of parameter variation in disease transmission

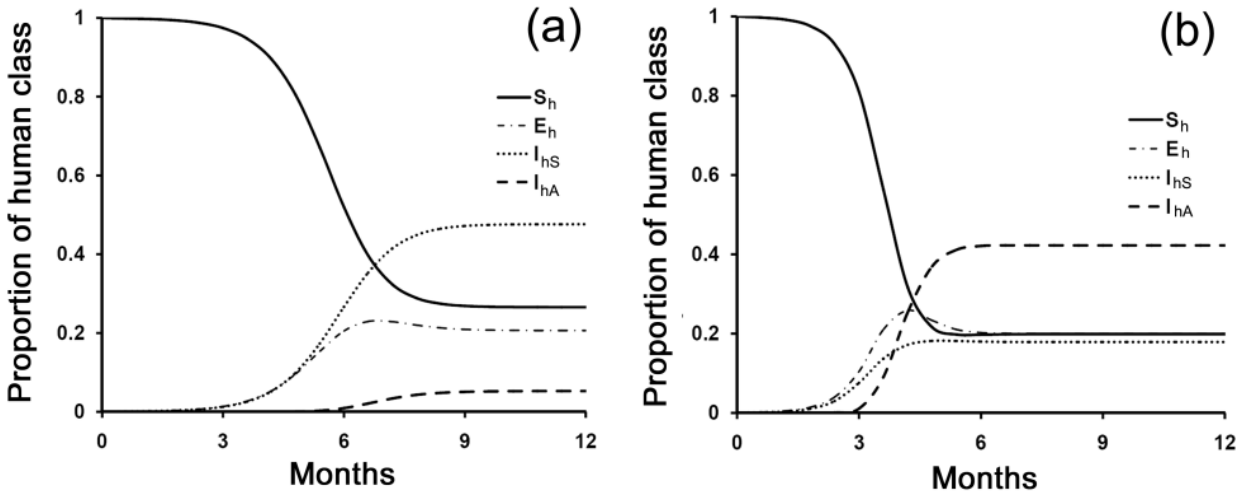
Parameters	Symbol	Value used for simulation	Reference
Human related	b	0.2	(Filipe et al. 2007)
	a_0	3 year	(Filipe et al. 2007)
	ρ_t	1/21 day^{-1}	(Drakeley et al. 2006)
	ρ_s	1/180 day^{-1}	(Filipe et al. 2007)
	ρ_0	1/180 day^{-1}	(Filipe et al. 2007)
	r	1/122 day^{-1}	Estimated
	$1/\gamma$	15 days	(Mandal et al. 2011)
	f	0.5	Assumed
	μ_h	0.0125 year^{-1}	(Mandal et al. 2011)
Mosquito related	α	0.67 day^{-1}	(Gu et al. 2003)
	C_{IS}	0.35	(Alves et al. 2002)
	C_{IA}	0.03	(Alves et al. 2002)
	τ	7 days	(Mandal et al. 2011)
	μ_m	0.1 day^{-1}	(Mandal et al. 2011)

Positive endemic equilibrium point **(S_h^*, E_h^*, I_{hS}^*, I_{hA}^*; S_m^*, E_m^*, I_m^*) = (0.265, 0.206, 0.476, 0.052; 0.470, 0.265, 0.265)** for constant age a = 2 year and m = 3.

In this case the basic reproduction rate **(R_0) is 6.93.**

For a =20 years, the endemic equilibrium point is **(0.198, 0.200, 0.178, 0.422; 0.665, 0.168, 0.167)** and the value of **R_0 is 3.37 which is also >1**.

Numerically simulation of the time course showed that these equilibrium points are always stable, satisfying our analytical result for $R_0 > 1$.



Simulation result of the model: variation of Susceptible (continuous), Exposed (dot-dashed), Symptomatic (dotted) and Asymptomatic (dashed) compartments of human population over time, for **(a) a =2 year** and **(b) a =20 year.**

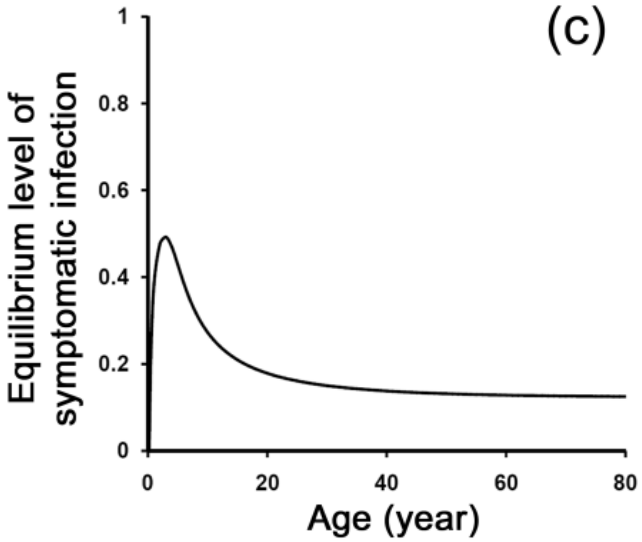
- After few months from the onset of infection all the compartments in human host reach their equilibrium levels, and remain stable.
- But the time taken to reach the equilibrium levels for 2 years aged individuals is much longer (about 8 months) compared to 20 years aged people (about 5 months).

- **Symptomatic cases are higher among lower aged individuals whereas asymptomatic cases are higher in higher aged persons.**

- **Fig. (c) shows the age dependency of symptomatic infection of human host, which reaches a maximum level for 3 years aged children.**
- **For higher aged people the prevalence is reduced and after a certain age (here *a* >20 years) it remains constant.**

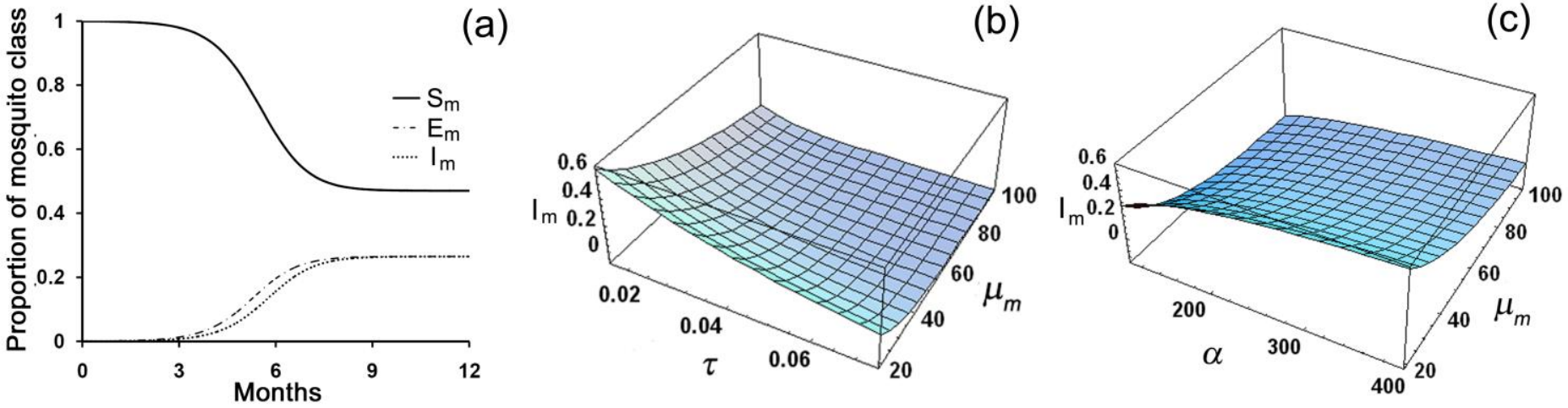
Acquired immunity of human, which grows with age and reaches to a saturation level after certain age

- **Successfully describes the time course and age dependency of infection of malaria in human (Dietz et al. 1974, WHO).**



Age-prevalence pattern of human infection

Mosquito dynamics



(a) Temporal dynamics of Susceptible (continuous), Exposed (dot-dashed) and Infected (dotted) compartments of mosquito population for m=3. (b) Change of equilibrium level of Infected mosquito with delay and mortality rate (for α = 244.55/year). (c) Change of equilibrium level of Infected mosquito population with biting rate and mortality rate (for τ = 0.019 year)

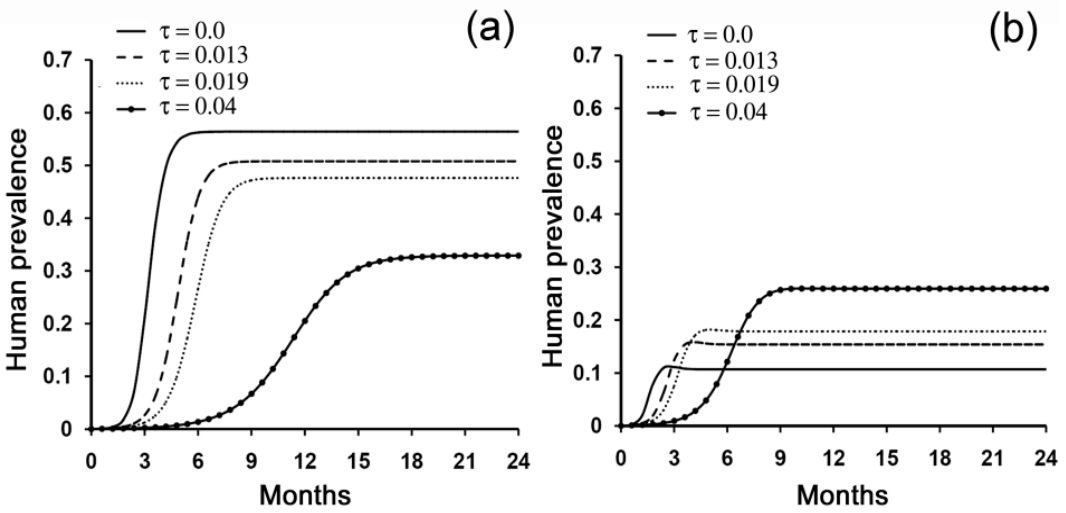
- **Prevalence pattern of mosquito population at each compartment, (S, E, I) supports our analytical observations, the dynamics is asymptotically stable for $R_0 > 1$.**
- With changes of **mosquito mortality rate (μ_m)** from 20/year to 100/year, the proportion of **infected mosquito reduces rapidly**.
- Similar phenomenon is observed for the **delay factor (τ)**, which is the time required for the parasite to develop in mosquito gut.
- The τ varies with temperature, as well as, the species of the parasite. **Increase in delay period decreases the equilibrium population density of infected mosquitos.**
- As **the biting rate (α)** of mosquito increases from 100/year to 400/year **the proportion of infected mosquito also increases.**

Effect on host dynamics due to variation of latent period & mortality of mosquito

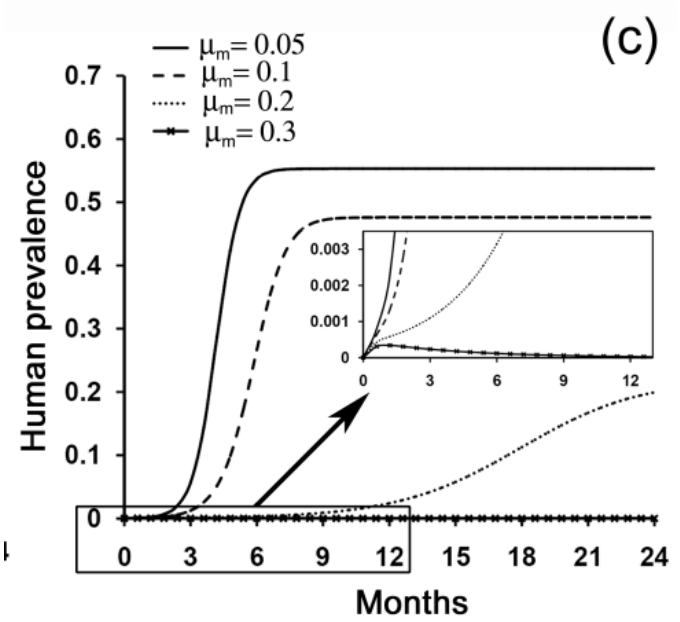
Sporogony period of parasite in mosquito gut depends on type of parasite species and with temperature.

Generally this period varies from 5 to 15 days (Anderson and May 1991; Koella 1991), and considered as a delay factor (τ) in the model.

- As the delay period increases, time to reach the saturation levels also increases.
- An increase in the delay-period induces a reduction in the level of saturation in symptomatic cases for lower aged individuals but shows opposite behavior for higher-aged people.



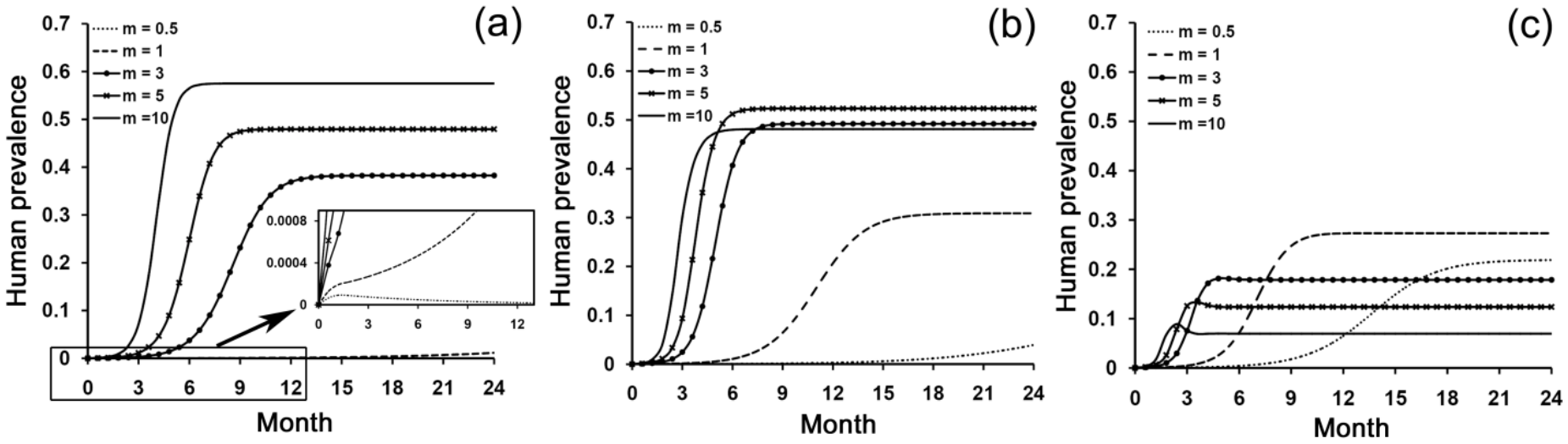
Effect of delay on symptomatic disease cases in human population for age **(a) 1 year and (b) 20 years.** μ_m= 0.1/day.



- The prevalence level varies widely for different mosquito mortality rates **- the increase in mosquito mortality reduces the symptomatic cases in human populations.**
- **For very high mortality rate of mosquito (e.g. for μ_m= 0.3/day), the disease may even get eradicated from the population.**

(c) Effect of different mosquito mortality rates on human disease prevalence in 2 years aged individuals for τ = 0.019 year. The inset shows magnification of the plot for the early time points as prevalence is very low for high mosquito mortality rate.

Effect on host dynamics due to change in mosquito density



Symptomatic disease prevalence in human population for different mosquito densities: for (a) 1 years (b) 3 years and (c) 20 years aged population; (μ_m= 0.1/day, τ =0.019 year). For lower age (say for 1 year) and for low mosquito density the prevalence is very low as shown in the inset of figure (a).

- **Equilibrium level of symptomatic disease prevalence for 1 year aged population increases as the mosquito density increases.** Same behavior for age 3 years up to certain level of mosquito density **- a reduction of equilibrium level when mosquito density '*m*' changes from 5 to 10.**
- **For higher aged people (say, for 20 years), the equilibrium level of symptomatic infection reduces more for higher mosquito density** - **due to the acquired immunity**, which develops in adults with more infected mosquito bites.
- After the onset of the disease, **the prevalence grows rapidly and the rate of infection increases** as mosquito density increases, irrespective of the age of the human.
- **Saturation level of infected human varies with mosquito density differently for different aged people.**
- **For higher aged people saturation level reduces to a lower value for higher mosquito density.**

Summary

- **Analytical and numerical study -** it is clear that this transmission model can be used to observe the effect of different mosquito related factors in malaria prevalence.
- **Successfully describes the age prevalence pattern of malaria in humans.**
- **Several predictions** can be made from the model with respect to changes in different human and mosquito specific parameters in regulating the disease
  - **Basic reproduction number** decreases with increase in delay-period for young persons. e.g. for age 2 years, $\mathbf{R_0 = 11.13}$ in case of no delay and **3.42** when delay is 0.04 year (15 days). **R_0 becomes < 1, if delay is more than 0.072 year (26 days).**
  - **Rate of mosquito mortality has a huge impact on disease transmission process.**

    Very high mortality rate of mosquito (e.g. for μ_m= 0.3/day), the disease may even get eradicated from the population - By different prevention method one can increase the mortality rate of mosquito through which the disease transmission can be controlled (Briet 2002; Killeen et al. 2011).
  - **Changes in mosquito birth/death rates, density, biting rates, etc.** can be achieved by killing mosquitoes through spraying insecticide, through genetically modified (GM) mosquitoes, which are resistant to parasites, or by using insecticide treated bed nets (ITN) or mosquito repellent.

Based on the outcomes of the model, suitable control strategies can be adopted for reducing malaria prevalence.

Question: Does it support real life scenario?

Testing the model with malaria prevalence data

- **Normally, the natural mortality rate and the latent period in mosquito are considered constant for a specific parasite and mosquito species, and for the particular malaria endemic area.**
- **Variation in mosquito density with respect to human population is quite different during different seasons in the year - may underlie the seasonal pattern of malaria prevalence observed in reality.**
- **To capture the actual seasonal pattern of malaria, there is a need to study variation of mosquito density (the mosquito density (*m*), which is considered as constant over time is not at all constant)**
- **Malaria prevalence data** from different regions in **India** and for different time periods is shown to generally **follow a bimodal shape in a year** (Dev et al. 2004, 2010; Malaria Site)

– WHY?

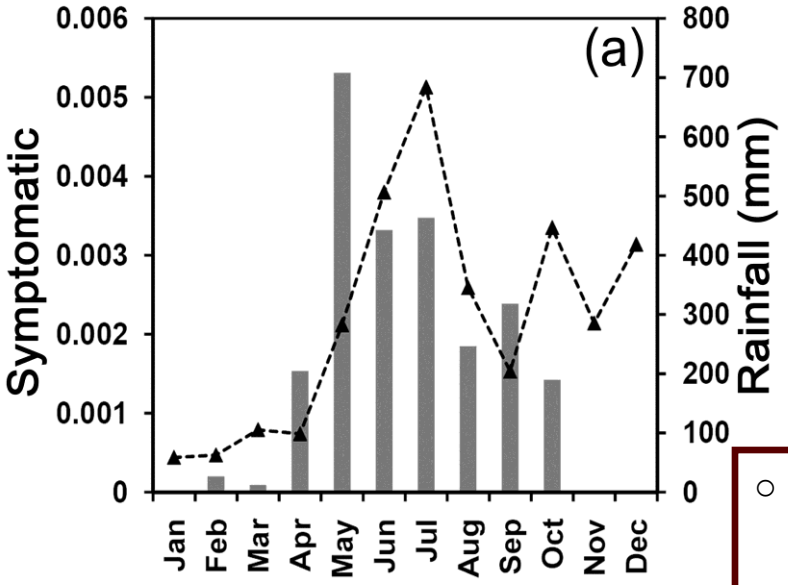
This pattern takes such forms due to the effects of seasonal variation of mosquito density, which may be due to the rainfall pattern (specifically in tropical regions).

The major reasons behind the time variation of mosquito density

- Direct or indirect influence of environmental changes as, rainfall pattern directly affect the larval development of mosquito in stagnant water,
- Relative humidity influence the life spans of mosquito (get shorten if humidity is below 50 percent or above 80 percent) etc.

(Russell et al. 1963; Pampana 1969; Sharma and Bos 2003; Bhattacharjee et al. 2006).

Malaria cases in West Garo Hill district, Meghalaya, India



Observed diseased cases and rainfall pattern (in mm) for the year 2006.

- ***Anopheles minimus*** is the predominant species of mosquito for spreading malaria.
- ***P. falciparum*** and ***P. vivax*** cause malaria, but ***P. falciparum*** **is the dominant (>82%) species and solely responsible for each malaria-attributable death case.**
- **Entomological parameters**: Mosquito biting rate (*α,* varies from 2 to 4/person/night); infection status of collected mosquitoes (I_m is 2.27%) by detection of sporozoites in their salivary glands, etc.. **- Dev et al. 2010**

- **Malaria transmission was perennial and persistent**, **with seasonal peak during May-July -months of high rainfall.**
- **Second peak appears in the month of September.**
- **Disease prevalence shows two peaks in July and October .**

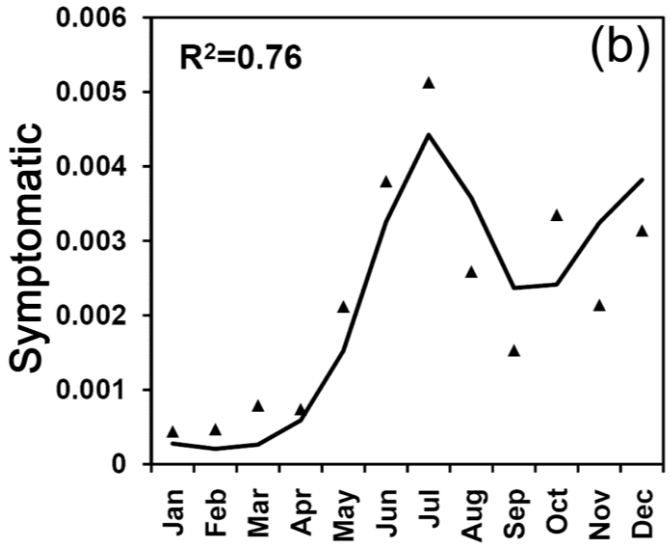
Variations in mosquito density, *m*, a time dependent periodic function is proposed :

$$m(t) = 0.11 + 3.55\, Sin^2\{\pi(t-\theta_1)\}\, Cos^2\{\pi(t-\theta_2)\}$$

t : time of the year, $\theta_1 = 0.1$, and $\theta_2 = 0.175$ effectively determine the peak position of mosquito density in that year - estimated from the prevalence pattern.

Excellent goodness of fit.

Establishes the capability of our model to explain the realistic disease prevalence pattern with minimal changes.



Model result and observed data

S.Mandal, S.Sinha, RR. Sarkar* (2013): *Bull. of Math.Biology* 75: 2499-2528

DELETE

Proportion of observed (dotted line) and simulated (solid line) malaria cases for One year in different regions – Short Time Series

Error bas represent the standard deviation of ten model simulations. Bar plots in first three figures represent the rainfall (in mm) of the respective year but for Mpumalanga (figure D) bar plot represents the average rainfall pattern.

DELETE

Model simulation and prediction of malaria in Mangalore and Mpumalanga – Long time Series

In each figure error bars represent the standard deviation of ten model simulations

Conclusion

- **Statistical & Mathematical Modelling of malaria transmission across a range of environmental conditions, including climate change**, is an important and emerging research area.
- Models that incorporate the **essentials of host-vector interaction**, **proper clinical population subdivisions for disease transmission**, **analyse the effect of seasonal forcing, rainfall, correlation between different variables and parameter changes**, and **also describe real data sets**, promise to be useful for real world implications, and ***may help us to understand hitherto unknown scenario***.
- **A new model of malaria transmission is proposed considering realistic immunity factors of hosts (*clinical and parasite*) and mosquito dynamics under environmental variations.**
- **Model is formulated dividing human population into different easily detectable epidemiological compartments and considering age-dependent immunity of human host.**
- From the analytical and numerical results, we observe that several mosquito related parameters like **mosquito birth/death rate, mosquito density, delay period of parasite development in mosquito gut etc.** are involved in the expression of endemic equilibrium point of model **- these parameters can also be used for controlling the endemic level of the disease.**

Introduction of the effect of seasonal changes on mosquito population (mosquito density) - capability of predicting malaria prevalence in a real situation (data)

By estimating the appropriate function for time variation of mosquito density, the deterministic model proposed **can successfully predict the prevalence pattern of any region**.

Confers the model with wide applicability to different epidemiological situations -- an important approach in epidemiological modeling studies

Serve two purposes:

The general model **can give important insights about real processes** that contribute to disease transmission (parameter dependencies and age prevalence), and **predict possible ways to control the outbreak of malaria**;

and

Depending on the requirements **one can use this model for specific purposes to model real disease prevalence data successfully**

Acknowledgement

- **Department of Science and Technology, Govt. of India,**
- **Corporation of Chennai & Kolkata**
- Dr. Somdatta Sinha (IISER-M), Sayantani Basu Roy, Dr. Sandip Mandal (Public Health Foundation of India)

Mosquitoes don't follow models but Human intervention is needed for control……

Thank you!