From Detection to Forecasting. Big Data and Models in Epidemiology, the Example of Dengue in Rio de Janeiro

Pierre-Alexandre Bliman (Inria, France and FGV, Brazil)
Cláudia Codeço (Fiocruz, Brazil) and Flávio Coelho (FGV, Brazil)

FGV EMAp ESCOLA DE MATEMÁTICA APLICADA
FIOCRUZ
Inria

June 10, 2016

▶ [see last slide for the main sources & references] ◀

Some facts about dengue (1/2)

- **Global burden and history**
  - **400 million dengue infections per year, of which 96 million manifest clinically: more than 30 times the 1960 number.**
  - **3.9 billion people, in 128 countries, at risk of infection.**
  - **Before 1970, 9 countries had experienced severe epidemics. Now endemic in more than 100 countries (America, South-East Asia, Western Pacific most affected regions).**
- **Clinical characteristics**
  - A severe, flu-like infectious illness that seldom causes death.
  - Caused by one of the Arboviruses (arthropod-borne viruses), like chikungunya, Zika, yellow fever.
  - Symptoms usually last for 2–7 days, after an incubation of 4–10 days after the bite from infected mosquito.
  - Severe dengue is a potentially deadly complication.

Some facts about dengue (1/2)

- **Global burden and history**
  - **400 million dengue infections per year, of which 96 million manifest clinically: more than 30 times the 1960 number.**
  - **3.9 billion people, in 128 countries, at risk of infection.**
  - **Before 1970, 9 countries had experienced severe epidemics. Now endemic in more than 100 countries (America, South-East Asia, Western Pacific most affected regions).**
- **Clinical characteristics**
  - **A severe, flu-like infectious illness that seldom causes death.**
  - **Caused by one of the Arboviruses (arthropod-borne viruses), like chikungunya, Zika, yellow fever.**
  - **Symptoms usually last for 2–7 days, after an incubation of 4–10 days after the bite from infected mosquito.**
  - **Severe dengue is a potentially deadly complication.**

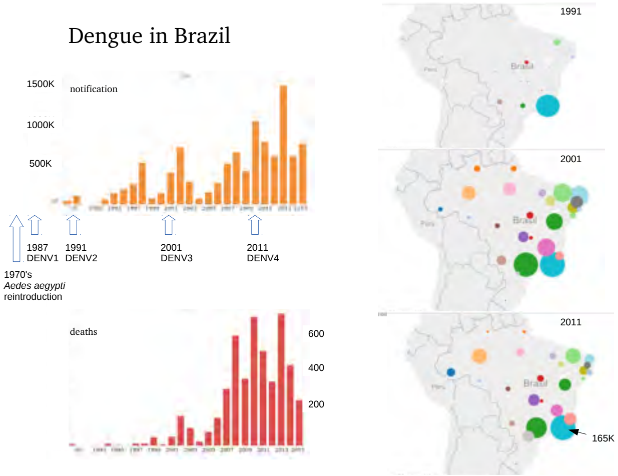
Some facts about dengue (2/2)

- **Immunological aspects**
  - **4 serotypes, DENV-1 to DENV-4.**
  - **Infection with one serotype produces lifelong immunity against that serotype reinfection.**
  - **Cross-immunity to the other serotypes after recovery is only partial and temporary.**
  - **Successive infection with two different serotypes is a risk factor for developing the severe forms of the disease.**
- **Treatment – Immunization**
  - No specific treatment.
  - For severe dengue, early detection and proper medical care lowers fatality rates from $\geq 20\%$ to $\leq 1\%$.
  - Late 2015/early 2016: registration of 1st vaccine in several countries. WHO Vaccine Position Paper to be published in July 2016.
  - Prevention and control is still the best protection.

Some facts about dengue (2/2)

- **Immunological aspects**
  - 4 serotypes, DENV-1 to DENV-4.
  - Infection with one serotype produces lifelong immunity against that serotype reinfection.
  - Cross-immunity to the other serotypes after recovery is only partial and temporary.
  - Successive infection with two different serotypes is a risk factor for developing the severe forms of the disease.
- **Treatment – Immunization**
  - No specific treatment.
  - For severe dengue, early detection and proper medical care lowers fatality rates from $\geq 20\%$ to $\leq 1\%$.
  - Late 2015/early 2016: registration of 1st vaccine in several countries. WHO Vaccine Position Paper to be published in July 2016.
  - Prevention and control is still the best protection.

The complex, yet increasing, evolution of dengue epidemics in Brazil



Mode and vectors of transmission

- **Aedes aegypti (main source) and Aedes albopictus mosquitoes are transmitters of dengue**
- **Host→Vector** **After ingestion of infected blood, dengue virus incubates in the mosquito during 8 to 12 days, after which the mosquito begins to transmit the virus biting other hosts.**
- **Vector→Host** **Newly infected host may have symptoms after 5-7 days of infection. (Immediate mechanical transmission, without incubation, can also occur when mosquito interrupts feeding on an infected host and immediately feeds on another susceptible one.)**

▶ **As the transmission cycle requires presence of the mosquito, the endemic area depends strongly (but not only) upon the vector geographical dissemination. The increase of the latter, especially due to urbanization (and possibly global warming), puts at risk new regions.**

Mode and vectors of transmission

- **Aedes aegypti (main source) and Aedes albopictus mosquitoes are transmitters of dengue**
- **Host→Vector** **After ingestion of infected blood, dengue virus incubates in the mosquito during 8 to 12 days, after which the mosquito begins to transmit the virus biting other hosts.**
- **Vector→Host** **Newly infected host may have symptoms after 5-7 days of infection. (Immediate mechanical transmission, without incubation, can also occur when mosquito interrupts feeding on an infected host and immediately feeds on another susceptible one.)**

▶ **As the transmission cycle requires presence of the mosquito, the endemic area depends strongly (but not only) upon the vector geographical dissemination. The increase of the latter, especially due to urbanization (and possibly global warming), puts at risk new regions.**

Aedes aegypti

- Most likely transported into Americas and Mediterranean on sailing ships from Africa (dengue epidemics in both places in 18th centuries).
- Disappeared from Mediterranean (unknown reason) in mid-20th century, and from USA in 1950s-1970s (eradication campaigns) and is now recolonizing.
- Freely enters homes and other buildings (rarely found more than 100m from human habitation), spends much time hidden in dark places. Feeds almost exclusively on human blood.
- Lays eggs in man-made objects that contain water: from discarded tires and buckets to saucers under flowerpots and water-storage barrels.
- Cannot tolerate cold winters.

Humans provide safe shelter, plentiful food, abundant sites for procreation: Aedes aegypti is the quintessential urban vector of dengue, aided by poverty and urbanization.

Aedes aegypti

- **Most likely transported into Americas and Mediterranean on sailing ships from Africa (dengue epidemics in both places in 18th centuries).**
- **Disappeared from Mediterranean (unknown reason) in mid-20th century, and from USA in 1950s-1970s (eradication campaigns) and is now recolonizing.**
- **Freely enters homes and other buildings (rarely found more than 100m from human habitation), spends much time hidden in dark places. Feeds almost exclusively on human blood.**
- **Lays eggs in man-made objects that contain water: from discarded tires and buckets to saucers under flowerpots and water-storage barrels.**
- **Cannot tolerate cold winters.**

Humans provide safe shelter, plentiful food, abundant sites for procreation: Aedes aegypti is the quintessential urban vector of dengue, aided by poverty and urbanization.

Aedes *albopictus*

- **Initially present from Beijing and northern Japan to tropical Asia. Widely distributed since 1985 in the southern United States (via global trade in used tires infested with eggs and larvae). Introduced into Brazil in the same date, providing secondary vector for dengue transmission.**
- **Often abundant in the peridomestic environment, particularly in areas with plentiful vegetation, but less endophilic than Ae. Aegypti.**
- **Also feeds freely on animals and birds, and so can exist far from human habitation. Since non-primates are not susceptible to dengue, generally regarded as secondary vector. But epidemics have been recorded in places where it is the only vector.**
- **Has the ability to breed in natural and artificial habitats.**
- **Possesses a diapausing egg state and therefore greater tolerance for temperate winters.**

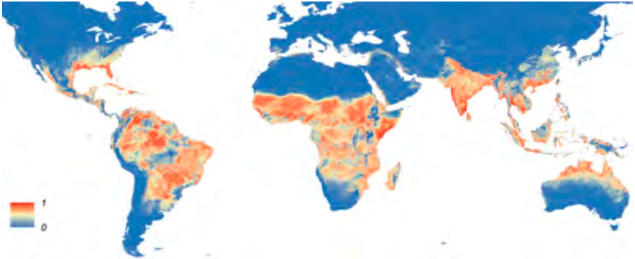
Less competent as a dengue vector, Ae. Albopictus has undergone dramatic global expansion, thanks to its ecological plasticity.

Aedes albopictus

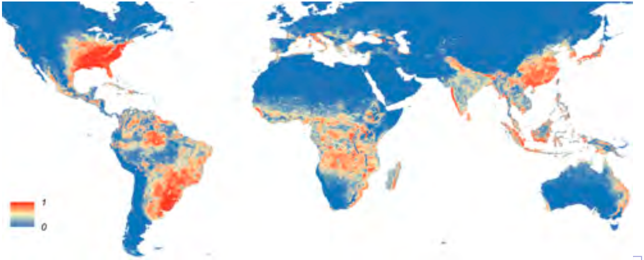
- **Initially present from Beijing and northern Japan to tropical Asia. Widely distributed since 1985 in the southern United States (via global trade in used tires infested with eggs and larvae). Introduced into Brazil in the same date, providing secondary vector for dengue transmission.**
- **Often abundant in the peridomestic environment, particularly in areas with plentiful vegetation, but less endophilic than Ae. Aegypti.**
- **Also feeds freely on animals and birds, and so can exist far from human habitation. Since non-primates are not susceptible to dengue, generally regarded as secondary vector. But epidemics have been recorded in places where it is the only vector.**
- **Has the ability to breed in natural and artificial habitats.**
- **Possesses a diapausing egg state and therefore greater tolerance for temperate winters.**

Less competent as a dengue vector, Ae. Albopictus has undergone dramatic global expansion, thanks to its ecological plasticity.

Geographic distribution of the vectors – World



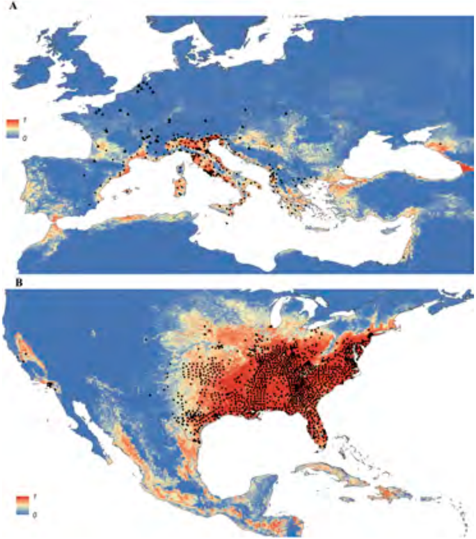
Ae. Aegypti



Probability of occurrence (from 0 blue to 1 red) at spatial resolution of 5 km $\times$ 5 km

Ae. Albopictus

Geographic distribution of *Ae. Albopictus* – USA and Europe



Probability of occurrence of Ae. albopictus in regions of rapid expansion.

Known occurrences (▲: transient, •: established) until end of 2013

Estimating effective reproduction number from epidemiological data

- **A key quantity in epidemiology is the effective reproduction number $\mathcal{R}_t$, i.e. the average number of secondary cases per primary case at time t.**

  $\mathcal{R}_t > 1$ ($\mathcal{R}_t < 1$) **suggests expansion (decline) of epidemic at time t.**

- **One can estimate $\mathcal{R}_t$ as**

$$\mathcal{R}_t = \frac{Y(t)}{\int_0^\infty Y(t-a)g(a)da}$$

  - $Y(t)$: **number of cases at time t**
  - g: **generation time distribution, i.e. probability distribution function of the time from infection of an individual to the infection of a secondary case**

Y **is obtained from epidemiological data, how to estimate the GT distribution for vectorial disease as dengue?**

Estimating effective reproduction number from epidemiological data

- **A key quantity in epidemiology is the effective reproduction number $\mathcal{R}_t$, i.e. the average number of secondary cases per primary case at time t.**

  $\mathcal{R}_t > 1$ ($\mathcal{R}_t < 1$) **suggests expansion (decline) of epidemic at time t.**

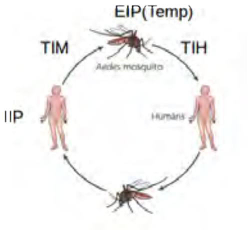
- **One can estimate $\mathcal{R}_t$ as**

$$\mathcal{R}_t = \frac{Y(t)}{\int_0^\infty Y(t-a)g(a)da}$$

  - $Y(t)$: **number of cases at time t**
  - g: **generation time distribution, i.e. probability distribution function of the time from infection of an individual to the infection of a secondary case**

Y **is obtained from epidemiological data, how to estimate the GT distribution for vectorial disease as dengue?**

Computing the Generation Time



Generation Time is the sum of 4 terms:

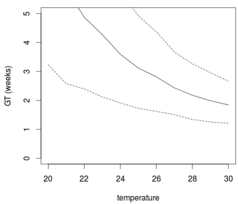
IIP Intrinsic Incubation Period **Time between human being infected and onset of symptoms (typically $4-7$ days)**

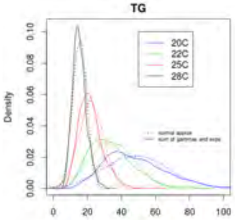
TIM Time to infect mosquito

EIP Extrinsic Incubation Period **Incubation period between viremic blood meal and mosquito becomes infectious ($\sim 8-12$ days)***

TIH Time to infect host

***: highly depends upon temperature**





Results:
dengue GT is $\sim$ normally distributed with temperature dependent mean and variance

Infodengue project

The Infodengue project (`info.dengue.mat.br`) aims to generate a near real-time warning system for cases of dengue transmission, at the level of the municipalities or neighborhoods, through quick access and processing of relevant data.

Alerts are generated through analysis of

- Epidemiological data obtained through Public Health notification systems
- Climatological data
- Social networks data

The exploitation of epidemiological data raises two challenges:

- delays in case notification
- absence of notification

Infodengue project

The Infodengue project (`info.dengue.mat.br`) aims to generate a near real-time warning system for cases of dengue transmission, at the level of the municipalities or neighborhoods, through quick access and processing of relevant data.

Alerts are generated through analysis of

- **Epidemiological data obtained through Public Health notification systems**
- **Climatological data**
- **Social networks data**

The exploitation of epidemiological data raises two challenges:

- **delays in case notification**
- **absence of notification**

Infodengue project

The Infodengue project (`info.dengue.mat.br`) aims to generate a near real-time warning system for cases of dengue transmission, at the level of the municipalities or neighborhoods, through quick access and processing of relevant data.

Alerts are generated through analysis of

- **Epidemiological data obtained through Public Health notification systems**
- **Climatological data**
- **Social networks data**

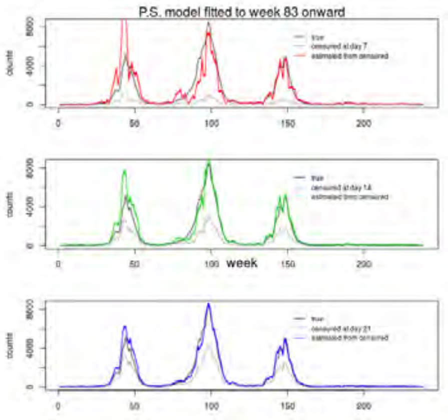
The exploitation of epidemiological data raises two challenges:

- **delays in case notification**
- **absence of notification**

Estimation of the actual cases from delayed data

Notifications are done with delay. . . which tends to be larger in case of outbreaks!

- **Prediction is therefore necessary.**

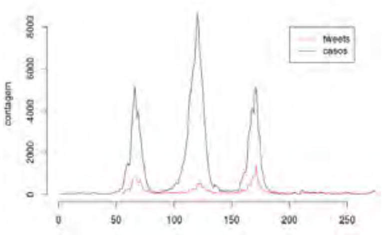


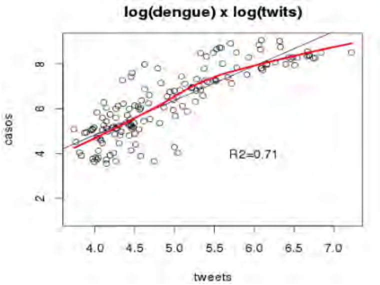
- **Cases notified after 1, 2 and 3 weeks (dotted)**
- **Estimations (colored)**
- **Ultimate value (black)**

Use of social networks when no or poor data are available

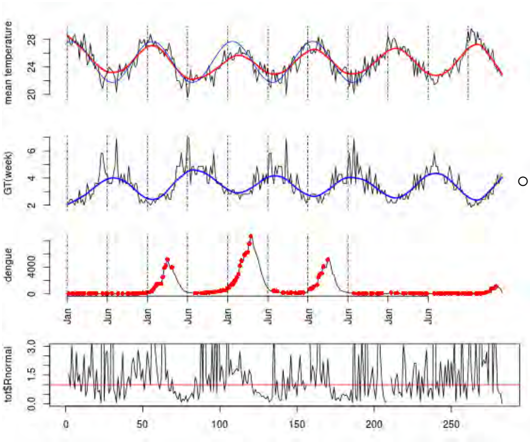
Especially in large urban centers, incidence of dengue goes with increased traffic on the social networks.

Related data — usually less precise in localization, but more reactive than the vigilance system — may be useful to estimate the number of new cases.





Near real-time estimation of $\mathcal{R}_t$

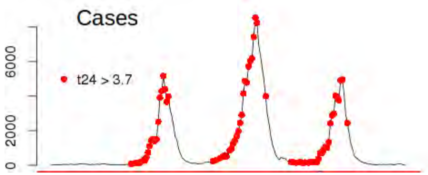


- Mean temperature
- Estimated mean generation time
- Estimated case number
- Estimated $\mathcal{R}_t$

(• indicates weeks when $\mathcal{R}_t > 1$ with confidence 95%)

Climatic effects and definition of alert levels

Statistical analysis shows that all outbreaks occur after mean value of 3.7 weeks with minimal temperature at least 24°C.

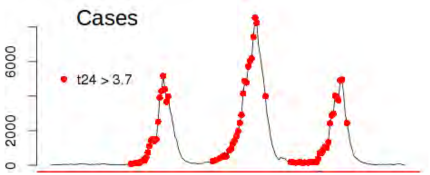


Alert levels are finally defined as follows:

- Unfavorable climate conditions, no cases
- Favorable climate conditions, sporadic cases with $\mathcal{R}_t < 1$
- Favorable climate conditions, $\mathcal{R}_t > 1$, prevalence $<$ 300/100,000 inhabitants
- Favorable climate conditions, $\mathcal{R}_t > 1$, prevalence $>$ 300/100,000 inhabitants

Climatic effects and definition of alert levels

Statistical analysis shows that all outbreaks occur after mean value of 3.7 weeks with minimal temperature at least 24°C.



Alert levels are finally defined as follows:

- **Unfavorable climate conditions, no cases**
- **Favorable climate conditions, sporadic cases with $\mathcal{R}_t < 1$**
- **Favorable climate conditions, $\mathcal{R}_t > 1$, prevalence $<$ 300/100,000 inhabitants**
- **Favorable climate conditions, $\mathcal{R}_t > 1$, prevalence $>$ 300/100,000 inhabitants**

Screenshots from the Infodengue site

Compartmental models of vectorial disease

Compartmental models of disease describe the evolution of

- **susceptibles S, capable of becoming infective;**
- **infectives I, capable of transmitting the disease to susceptibles;**
- **recovered R, permanently immune after healing.**

$$\frac{dS_H}{dt} = \mu_H N_H - \mu_H S_H - \beta_H S_H I_v$$
$$\frac{dI_H}{dt} = \beta_H S_H I_v - \mu_H I_H - \gamma_H I_H$$
$$\frac{dR_H}{dt} = \gamma_H I_H - \mu_H R_H$$
$$\frac{dS_v}{dt} = \mu_v N_v - \mu_v S_v - \beta_v S_v I_H$$
$$\frac{dI_v}{dt} = \beta_v S_v I_H - \mu_v I_v$$

N_H, N_v: **total populations**

μ_H, μ_v: **birth and death rates**

β_H, β_v: **transmission rates (function of biting rate, transmission probability, population size...)**

γ_H: **recovery rate**

Compartmental models of vectorial disease

Compartmental models of disease describe the evolution of

- **susceptibles S, capable of becoming infective;**
- **infectives I, capable of transmitting the disease to susceptibles;**
- **recovered R, permanently immune after healing.**

$$\frac{dS_H}{dt} = \mu_H N_H - \mu_H S_H - \beta_H S_H I_v$$
$$\frac{dI_H}{dt} = \beta_H S_H I_v - \mu_H I_H - \gamma_H I_H$$
$$\frac{dR_H}{dt} = \gamma_H I_H - \mu_H R_H$$
$$\frac{dS_v}{dt} = \mu_v N_v - \mu_v S_v - \beta_v S_v I_H$$
$$\frac{dI_v}{dt} = \beta_v S_v I_H - \mu_v I_v$$

N_H, N_v: **total populations**

μ_H, μ_v: **birth and death rates**

β_H, β_v: **transmission rates (function of biting rate, transmission probability, population size...)**

γ_H: **recovery rate**

Analysis of the compartmental model dynamics

Near the Disease Free Equilibrium (DFE), the effective reproduction number $\mathcal{R}_t$ is called basic reproduction number $\mathcal{R}_0$.

Theorem

The compartmental model has the following properties.

- *The value of $\mathcal{R}_0$ is* $\sqrt{\dfrac{\beta_H \beta_v}{(\mu_H + \gamma_H)\mu_v}}$
- *When $\mathcal{R}_0 < 1$, the DFE is globally asymptotically stable*
- *When $\mathcal{R}_0 > 1$, the DFE is unstable and a unique Endemic Equilibrium (EE) exists, which is almost-globally asymptotically stable.*

The previous model assumes that the “medium” is homogeneous — which is quite unrealistic for a large city. . .

Analysis of the compartmental model dynamics

Near the Disease Free Equilibrium (DFE), the effective reproduction number $\mathcal{R}_t$ is called basic reproduction number $\mathcal{R}_0$.

Theorem

The compartmental model has the following properties.

- *The value of $\mathcal{R}_0$ is* $\sqrt{\dfrac{\beta_H \beta_v}{(\mu_H + \gamma_H)\mu_v}}$
- *When $\mathcal{R}_0 < 1$, the DFE is globally asymptotically stable*
- *When $\mathcal{R}_0 > 1$, the DFE is unstable and a unique Endemic Equilibrium (EE) exists, which is almost-globally asymptotically stable.*

The previous model assumes that the "medium" is homogeneous — which is quite unrealistic for a large city...

Metapopulation models of vectorial disease

Patchy models have been introduced to describe urban (host) circulation without migration, consisting of n "homogeneous" patches, each one with host and vector populations $N_{H,i}$, $N_{v,i}$ for $i = 1, \ldots, n$.

$$\frac{dS_{H,i}}{dt} = \mu_H N_{H,i} - \mu_H S_{H,i} - S_{H,i} \sum_j \beta_{H,ij} I_{v,j}$$

$$\frac{dI_{H,i}}{dt} = S_{H,i} \sum_j \beta_{H,ij} I_{v,j} - \mu_H I_{H,i} - \gamma_H I_{H,i}$$

$$\frac{dR_{H,i}}{dt} = \gamma_H I_{H,i} - \mu_H R_{H,i}$$

$$\frac{dS_{v,j}}{dt} = \mu_v N_{v,j} - \mu_H S_{H,j} - S_{v,j} \sum_i \beta_{v,ij} I_{H,i}$$

$$\frac{dI_{v,j}}{dt} = S_{v,j} \sum_i \beta_{v,ij} I_{H,i} - \mu_v I_{v,j}$$

$N_{H,i}, N_{v,i}$: **total populations of patch** i

μ_H, μ_v: **birth/death rates**

$\beta_{H,ij}, \beta_{v,ij}$: **transmission rates between host from patch** i **and vector from path** j

Analysis of the patchy model dynamics

- **The behavior is qualitatively analogous:**
  - **if $\mathcal{R}_0^{\text{graph}} < 1$, the DFE is globally asymptotically stable;**
  - **if $\mathcal{R}_0^{\text{graph}} > 1$, the DFE is unstable, a unique EE exists and is almost-globally asymptotically stable.**
- **When $\mathcal{R}_0^{\text{graph}}$ fluctuates below and above 1 (due to seasonal changes), complicated oscillations appear.**

Now, to use such model for prediction, one should face important issues:

1. **How to quantify the intensity of the circulation between different neighborhoods of a given (large) city?**
2. **Given circulation pattern and climatic forecasting (and therefore the β's...), is it possible to estimate the number of susceptibles, and make short term outbreak predictions?**

Analysis of the patchy model dynamics

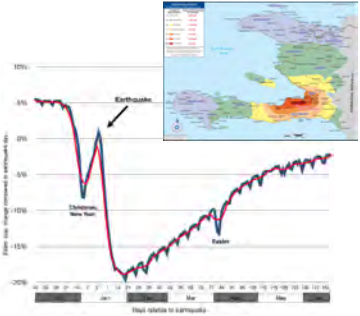
- **The behavior is qualitatively analogous:**
  - **if $\mathcal{R}_0^{\text{graph}} < 1$, the DFE is globally asymptotically stable;**
  - **if $\mathcal{R}_0^{\text{graph}} > 1$, the DFE is unstable, a unique EE exists and is almost-globally asymptotically stable.**
- **When $\mathcal{R}_0^{\text{graph}}$ fluctuates below and above 1 (due to seasonal changes), complicated oscillations appear.**

Now, to use such model for prediction, one should face important issues:

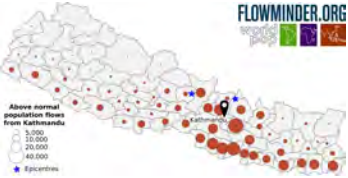
1. **How to quantify the intensity of the circulation between different neighborhoods of a given (large) city?**
2. **Given circulation pattern and climatic forecasting (and therefore the β's. . .), is it possible to estimate the number of susceptibles, and make short term outbreak predictions?**

Measuring human mobility through mobile phone data (1/2)

(Anonymized) mobile phone data offer the possibility to quantify population fluxes. This recent idea has been considered for quantifying large-scale population movements after disasters.



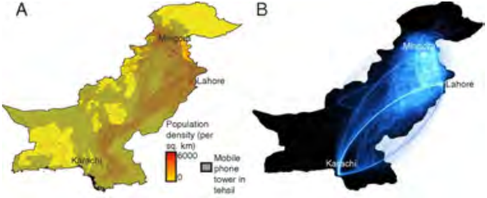
Estimated net changes of the Port-au-Prince population compared to the population on the 2010 earthquake day



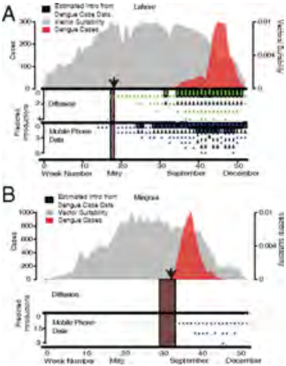
Anomalous flows from the Kathmandu valley, comparing the 10th-14th May with the 20th-24th April 2015

Measuring human mobility through mobile phone data (2/2)

It has also been used to predict geographic spread and timing of diseases in emerging locations.



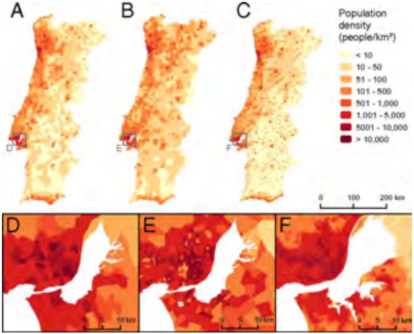
Human mobility dynamics in Pakistan.
(A) Population density (red, high; yellow, low) and mobile phone tower coverage (colored in gray).
(B) Top routes of travel between pairs of tehsils (counties).



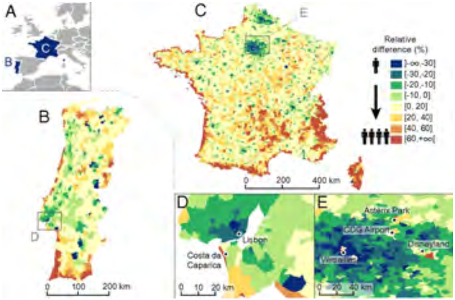
Mobility estimates predict timing of introduction from Karachi to (A) Lahore and (B) Mingora in 2013 dengue outbreak. Mobile phone data (blue boxplot), diffusion model (green boxplot), actual case data (red), estimated dengue suitability (gray).

Estimating urban circulation patterns

The mobile phone data are compatible with the space and time scales needed for dengue forecast in urban environment.



(A) Population density calculated from national census. (B) Population density estimated by the MP method. (C) Population density estimated by remotely-sensed and other geospatial data.



Relative difference in predicted population density between main holiday (Jul.-Aug.) and working period (Sep. to Jun.)

Sources and references

- **Key facts about dengue disease**
  - World Health Organization; Pan-American Health Organization
- **Vectorial distribution and evolution** • European Centre for Disease Prevention and Control
  - P. Reiter, Yellow fever and dengue: a threat to Europe?, Euro Surveill., 2010
- **Estimating the effective reproduction number from data**
  - J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, Proc. R. Soc. B, 2007 • M. Chan, M.A. Johansson, The incubation periods of dengue viruses, PLoS ONE, 2012
- **Nowcasting – the early warning system Infodengue in Rio de Janeiro**
  - `Infodengue` website
- **Towards forecasting** • Models • Data
  - K. Dietz, Transmission and control of arbovirus diseases, Proc. SIAM-SIMS Conf. 1974 • N.T.J. Bailey, The Mathematical Theory of Infectious Diseases, 1975 • A. Iggidr, G. Sallet, M. Souza, 2013, 2014 • R. Wilson et al., Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake, PLoS Currents: Disasters, 2016 • L. Bengtsson, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti, PLoS Medicine, 2011 • A. Wesolowski et al., Impact of human mobility on the emergence of dengue epidemics in Pakistan, PNAS, 2015 • P. Deville et al., Dynamic population mapping using mobile phone data PNAS, 2014