MODELING THE SPREAD OF XYLELLA FASTIDIOSA IN APULIA, ITALY

Github.com/LeonMengoni/Xylella-Fastidiosa

WHAT IS *XYLELLA FASTIDIOSA*?

- Bacterium transmitted by insect vectors feeding on **xylem sap**
- Causes **Olive Quick Decline Syndrome** (OQDS)
- Disease is **NOT** curable

WHERE DID IT COME FROM?

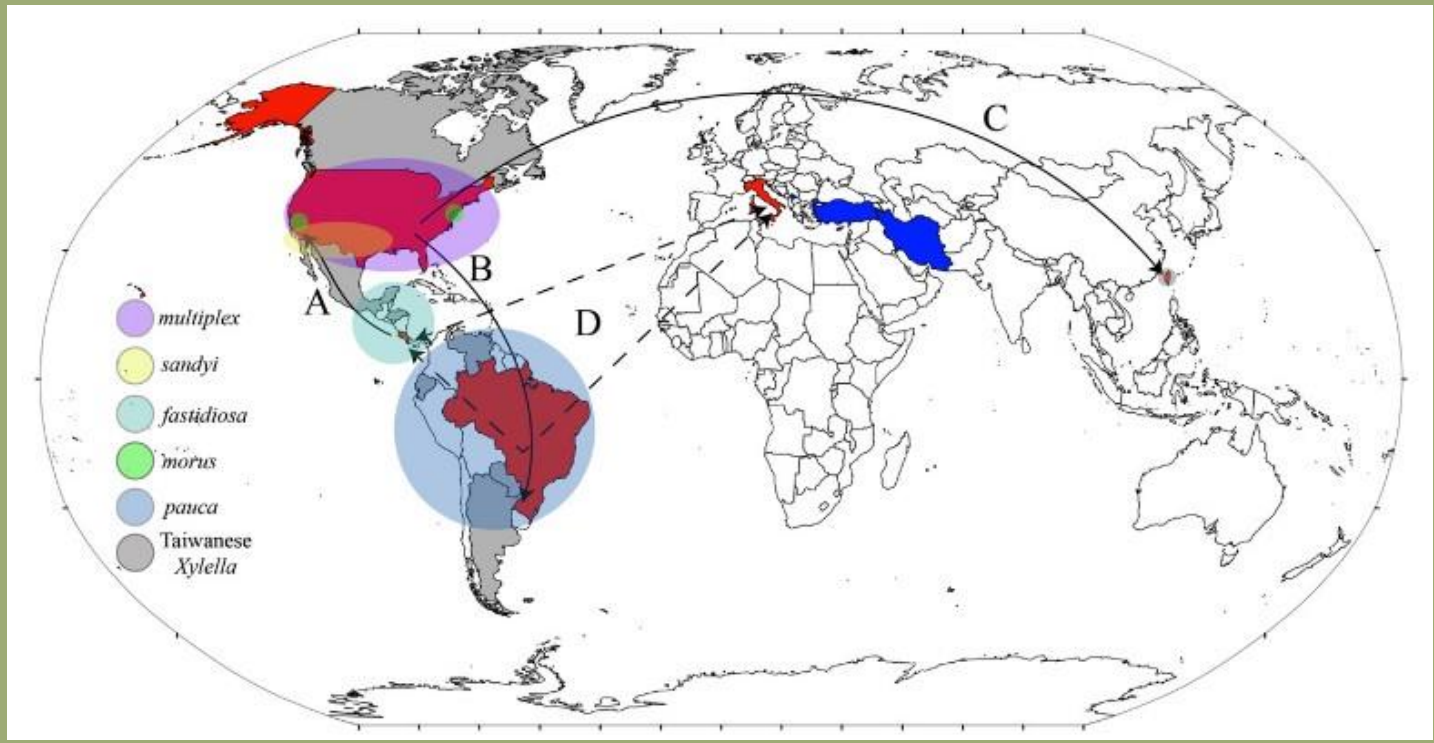
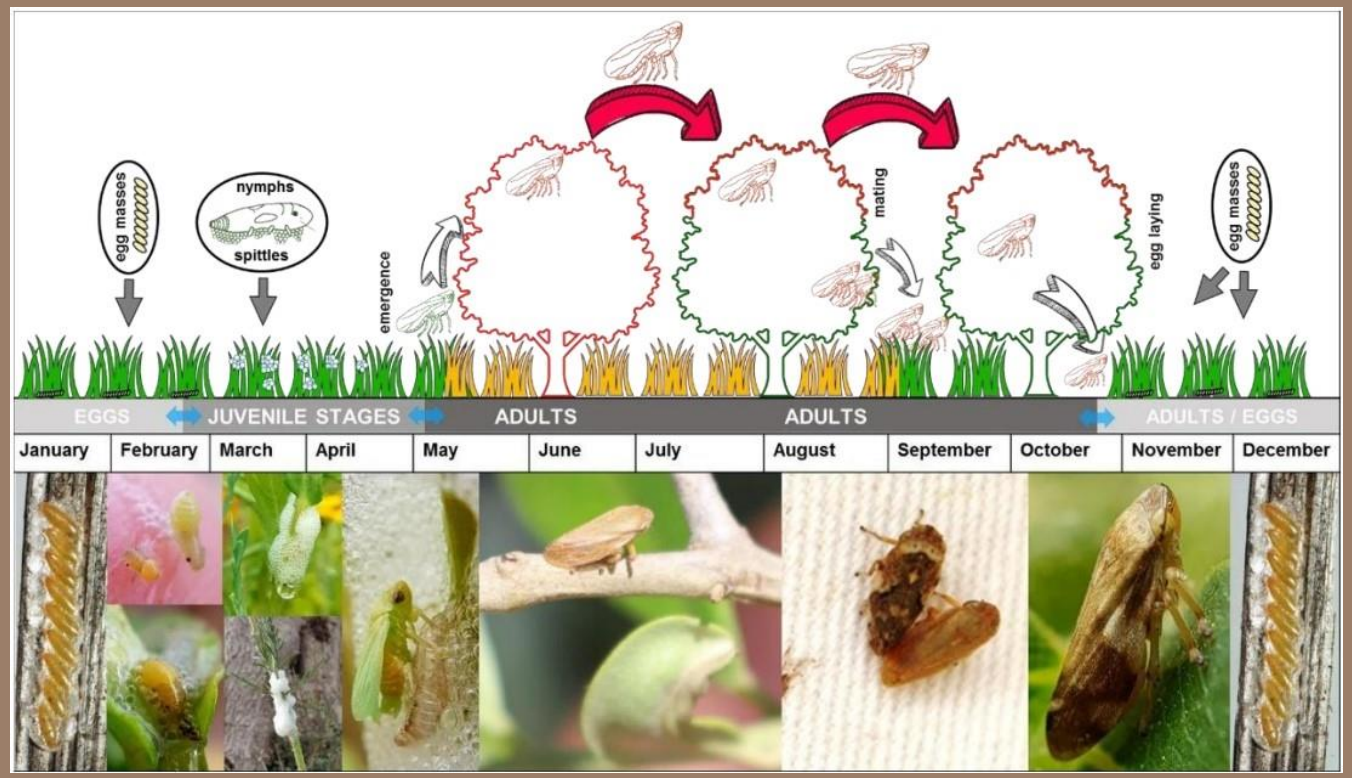
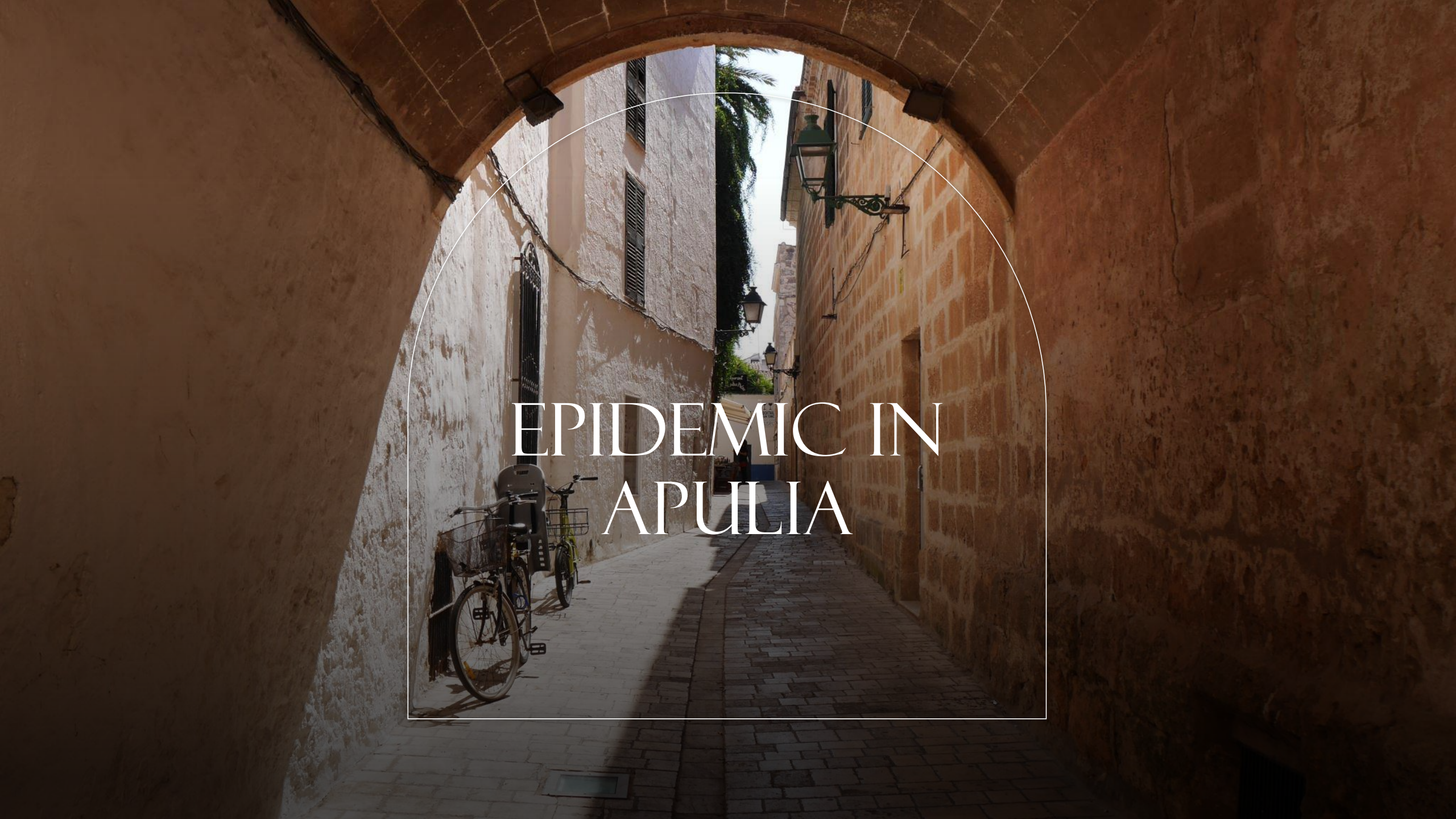


Image borrowed from ‘How do plant diseases caused by *Xylella Fastidiosa* emerge?’, Rodrigo P.P. Almeida and Leonard Nunney

HOW DOES IT SPREAD?



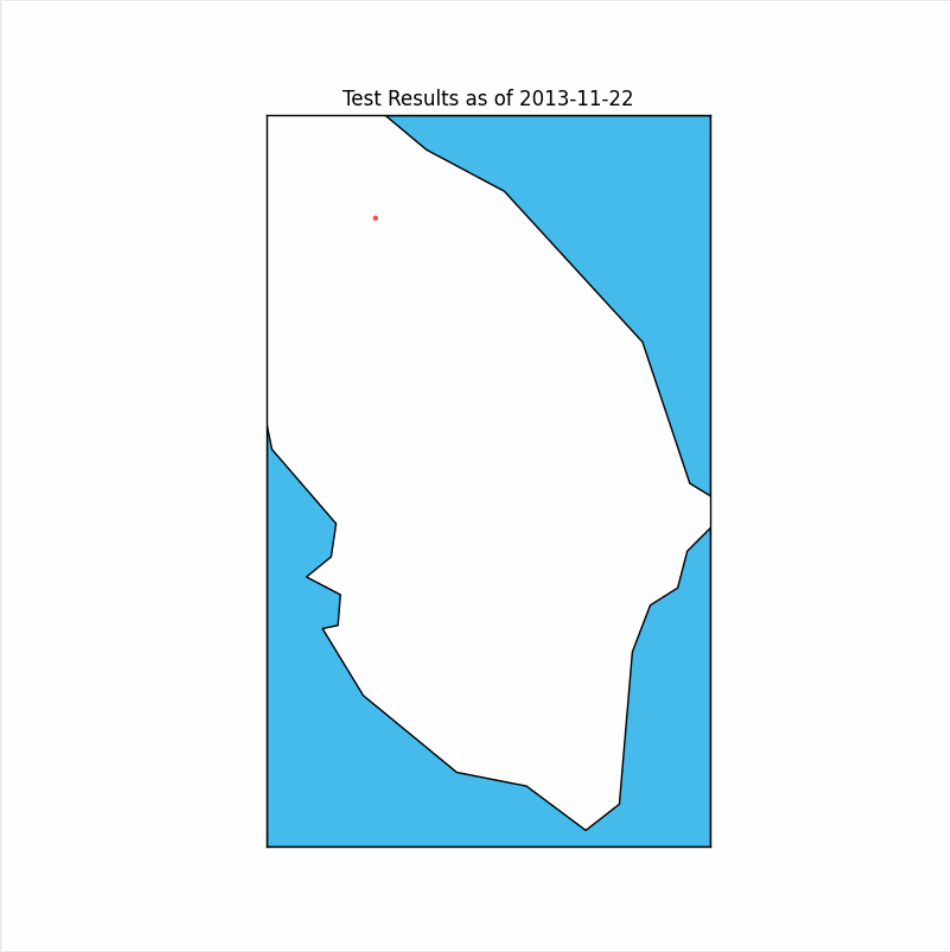
- *Philaneus Spumarius*
- 'Hopping' from weeds to trees
- Hitchhiking on cars and trucks

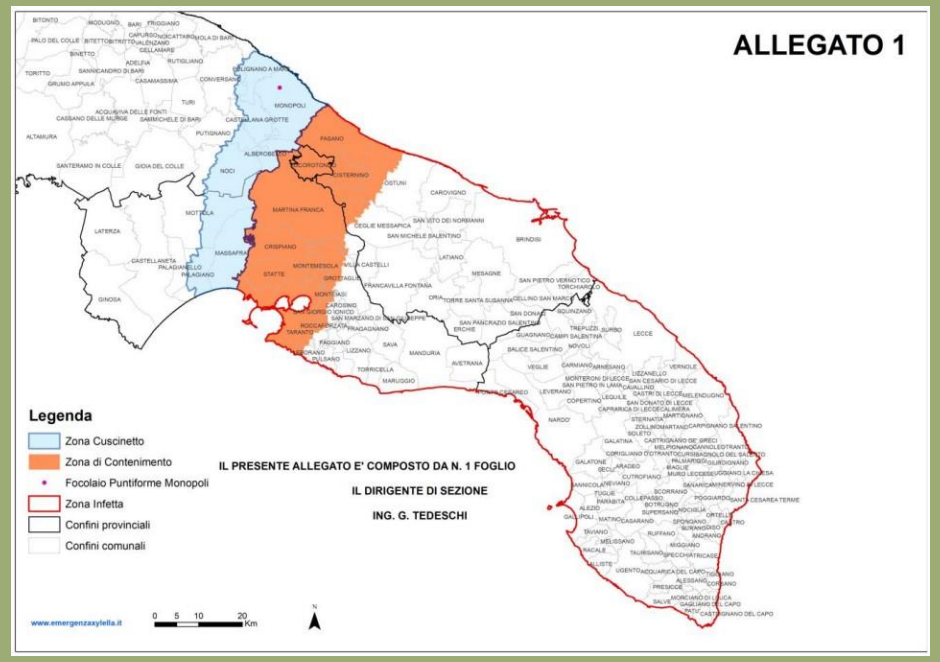


EPIDEMIC IN APULIA

INITIAL OUTBREAK

- First **detected** cases of *spp. pauca* in **2013**
- Entry-point: port of **Gallipoli**
- Protests and investigations → **slow response**
- Implementation of EU guidelines





EMERGENCY MEASURES

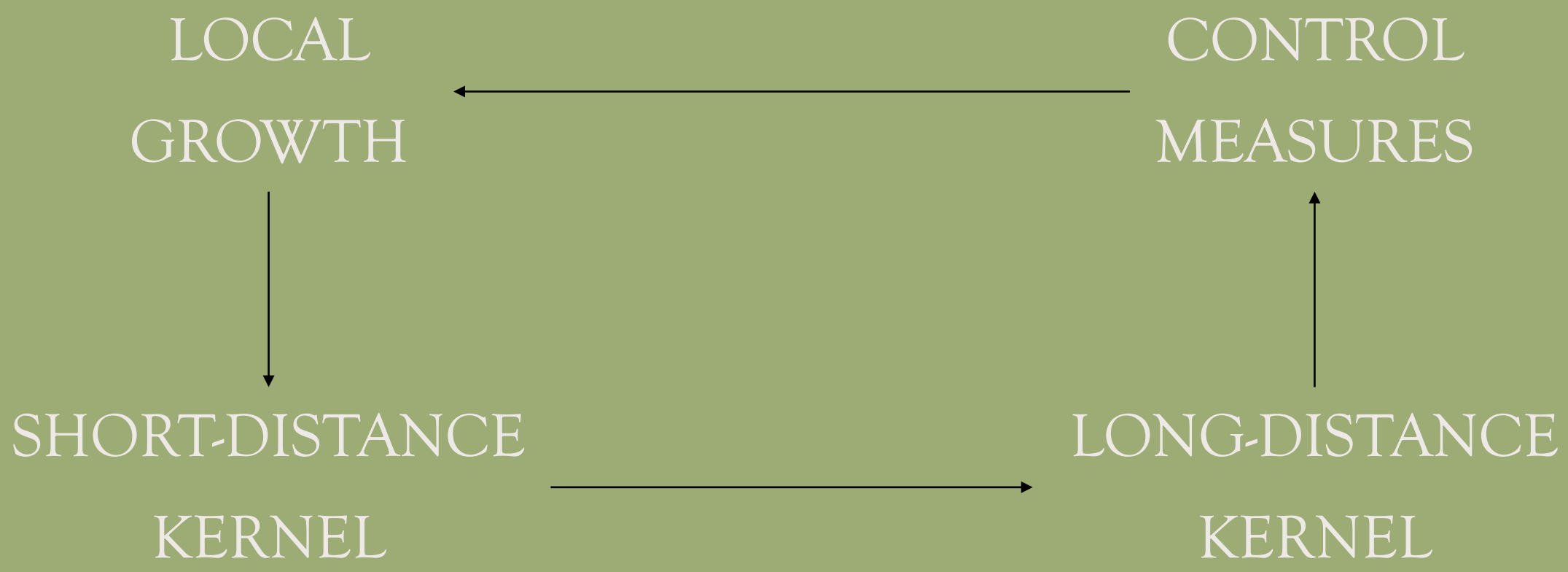
- **Infected zone** (IZ): disease cannot be eliminated
- **Containment zone** (CZ): last 20km of IZ
  - *eradicate infected trees and neighbors in 100m radius*
  - *vector control: weed removal, preventive ploughing*
- **Buffer zone** (BZ): 10km after CZ
  - *same measures as CZ*
  - *diversified planting of Xylella-resistant varieties*

CURRENT SITUATION

- **Coexistence** with Xylella
- More rapid testing and diagnostics
- Researching bacterium-resistant or –immune species of plants
- Olive oil production halved in 10 years

MODELING THE EPIDEMIC

MODEL BY WHITE ET AL. (2017)



1) LOCAL GROWTH

At every timestep t, model I_t^G:

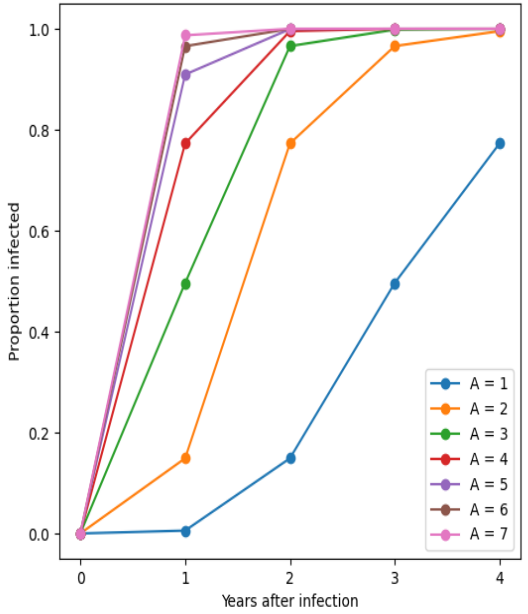
- Gompertz function: $I_t^G = Ke^{-Be^{-At}}$

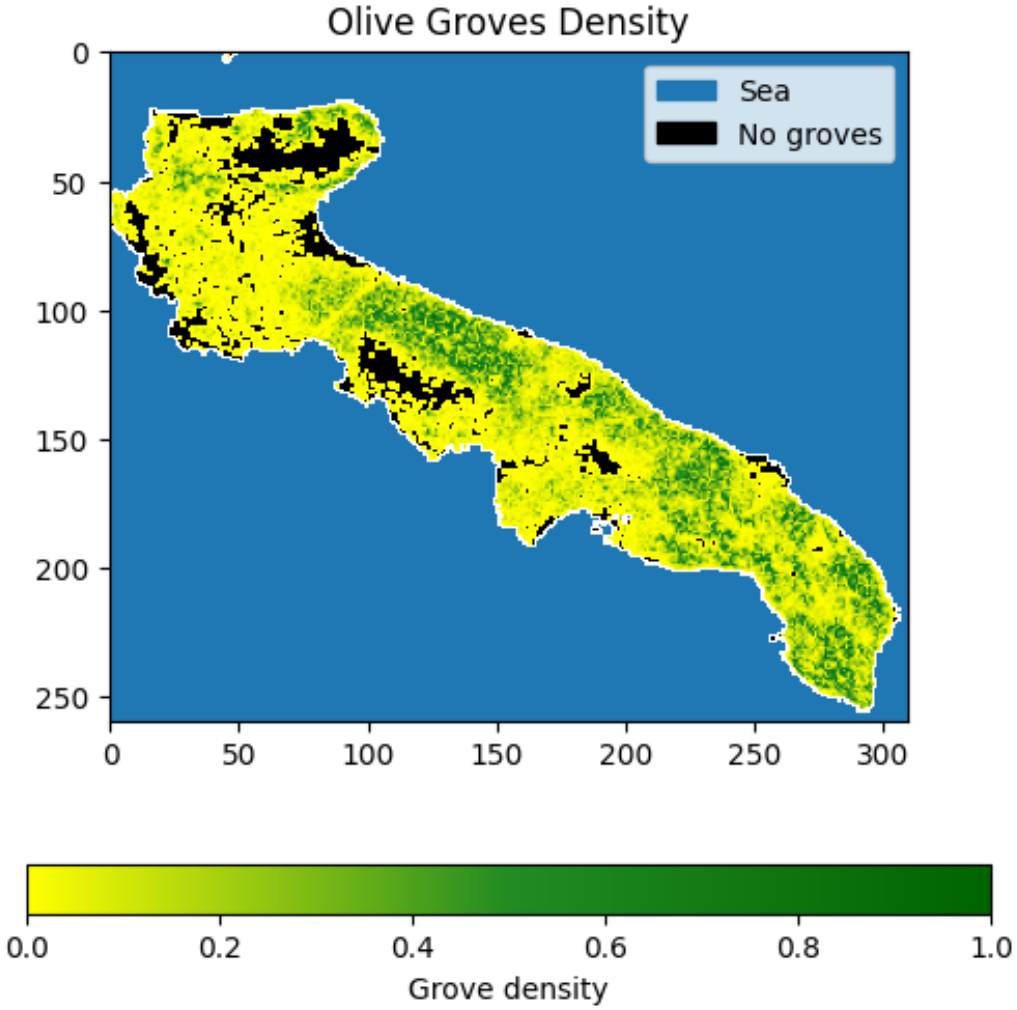
$$I_{t+1}^G(x,y) = K(x,y)\left(\frac{I_t(x,y)}{K(x,y)}\right)^{e^{-A}}$$

- Discretized on annual scale: $[t] = year$
- Olive grove density (per $1km^2$ grid cell):

  $d(x,y) \rightarrow I_t(x,y) \in [0, d(x,y)]$
- Carrying capacity: $K(x,y) = d(x,y) + a\big(1 - d(x,y)\big)$
- Proportion of non-olive groves:

  $a \in [0,1]$
- Incidence: $i_t(x,y) = \frac{I_t(x,y)}{d(x,y)}$

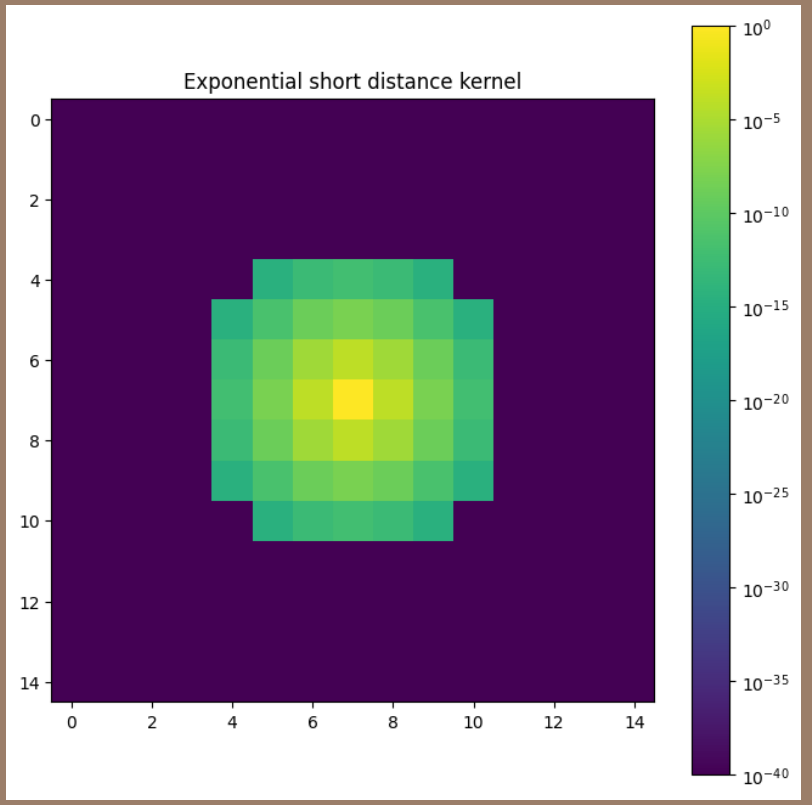




2) SHORT-DISTANCE KERNEL

At every timestep, model I_t^S:

- Exponential kernel: $\hat{k}_e(x,y) = e^{-\frac{(x^2+y^2)^{\frac{1}{2}}}{\beta}}, [\beta] = km$
- Gaussian kernel: $\hat{k}_g(x,y) = e^{-\frac{x^2+y^2}{2\beta^2}}$
- Convolve with $I_{t+1}^G \rightarrow I_{t+1}^S = \hat{k} * I_{t+1}^G$

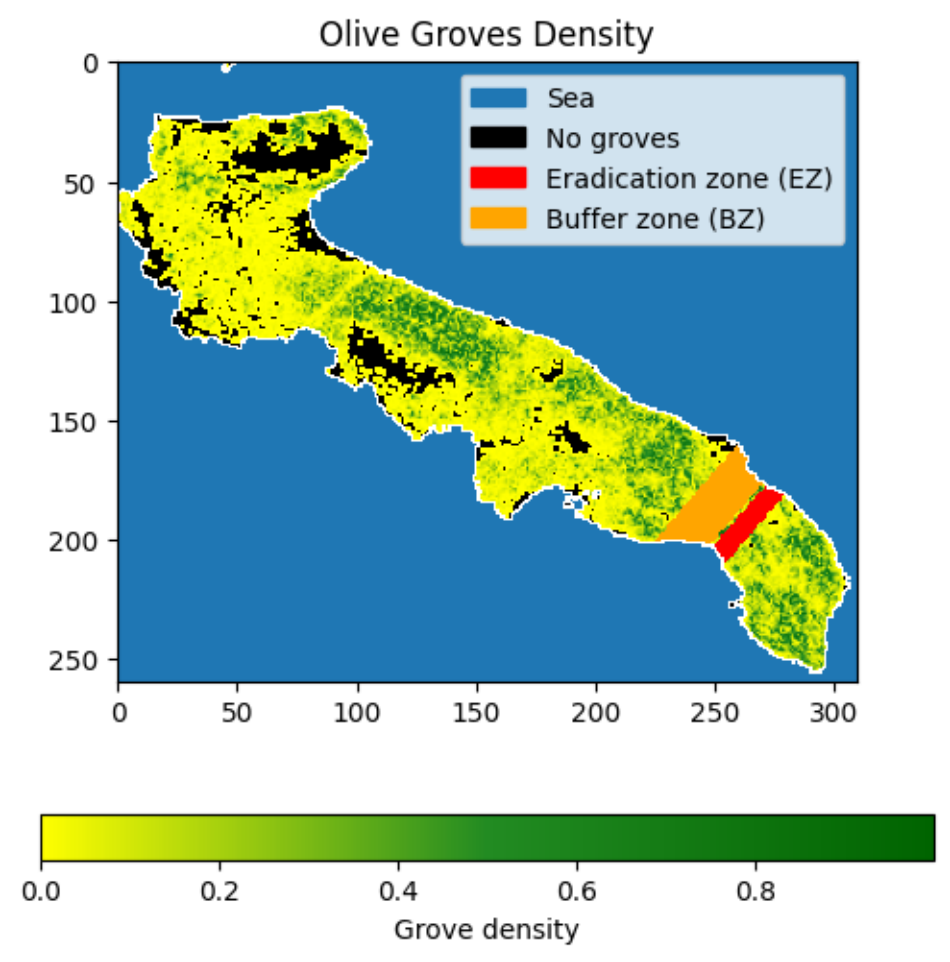


3) LONG-DISTANCE KERNEL

- Isotropic stochastic dispersal:
  - *sample* $\rho(x,y) \sim U[0,1]$
  - *if* $u(x,y) = \rho(x,y) I^S_{t+1}(x,y) > p\ (threshold\ probability)$:
    - *cell* (x,y) *generates* $M \leq M_{max}$ *random dispersers*
  - *every disperser* i *disperses a distance* $\vec{d}_i \sim N\left(0, (\sigma_x^2 = D, \sigma_y^2 = D)\right)$
  - *every destination cell* $(x',y') = (x,y) + \vec{d}$ *is further infected with*

$$\Delta I^L_{t+1}(x',y') = \left(d(x',y') - I^S_{t+1}(x',y')\right) e^{-B}$$

$$I^L_{t+1}(x',y') = I^S_{t+1}(x',y') + \Delta I^L_{t+1}(x',y')$$

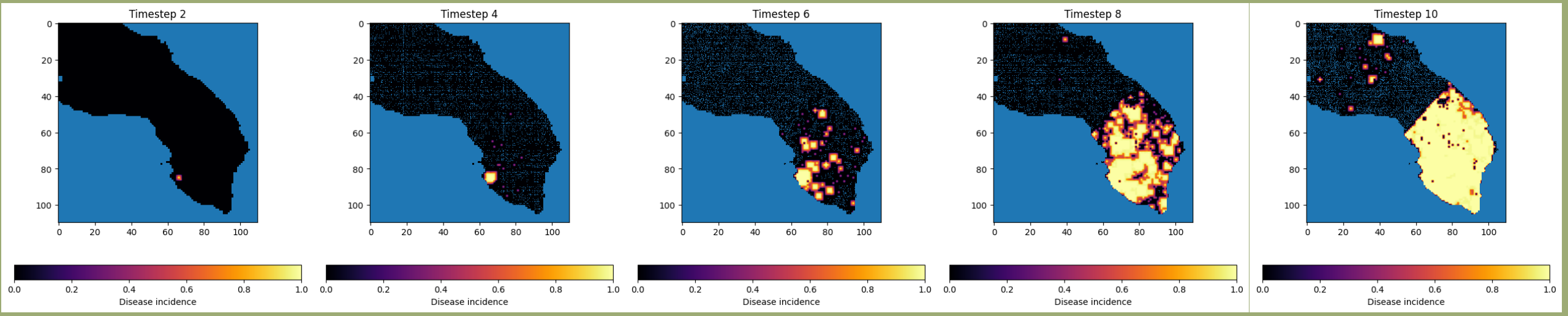


4) CONTROL MEASURES

- Eradication implemented only in containment zone (called *eradication zone*, **EZ**) and *buffer zone* (**BZ**)
- Assign a probability of infection detection to every cell: $p_{detect}(x,y) \sim U[0,1]$
- Define *surveillance efficiency* in both zones: $s_{EZ}, s_{BZ} \in [0,1]$ $(s_{EZ} > s_{BZ})$
- If s is greater than p_{detect} in a cell (x,y), eradicate all infected trees, without replacing them:

$$I^{C}_{t+1}(x,y) = \begin{cases} 0, & if\ p_{detect}(x,y) > s(x,y) \\ I^{L}_{t+1}(x,y), & otherwise \end{cases}$$

SIMULATION



$$I_t \rightarrow I_{t+1}^G \rightarrow I_{t+1}^S \rightarrow I_{t+1}^L \rightarrow I_{t+1}^C = I_{t+1}$$

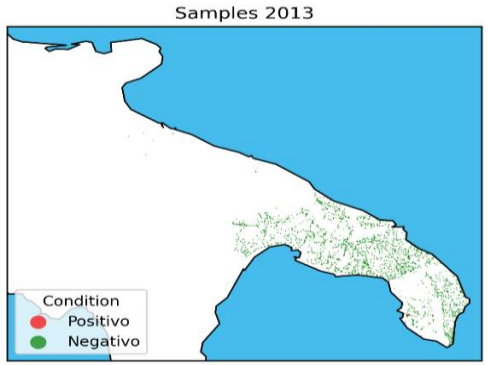
Default parameters from 2017 paper:

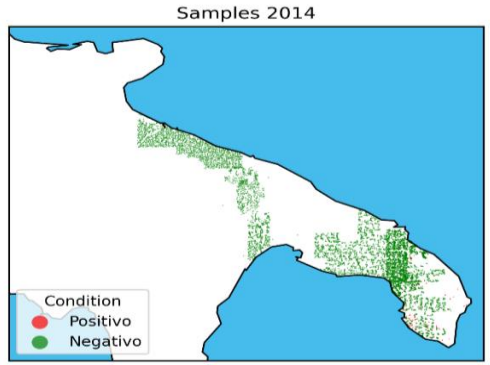
$$\{A = 3, B = 14.069, a = 0, \beta = 0.1, p = 0.2, M_{max} = 5, D = 20\}$$

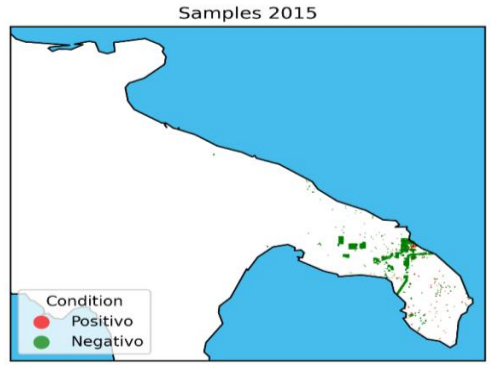
DATA ANALYSIS AND RESULTS

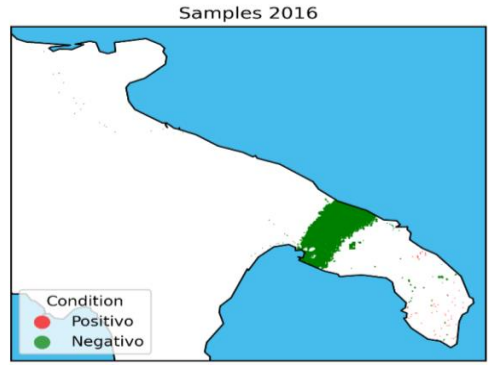
MONITORING DATA

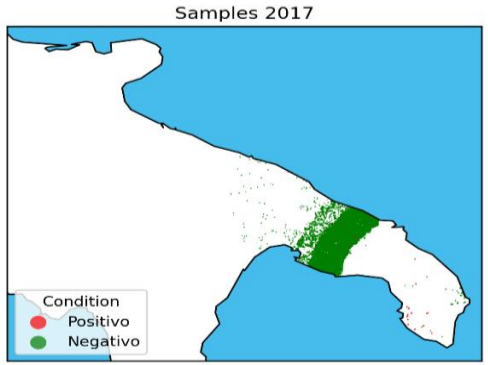
- Epicentre and date of seeding are unknown (hyp: Gallipoli, 2008)
- First tests only in 2013
- Samples are **heterogeneous in time and space**:
  - *number of samples taken every year varies a lot*
  - *Spatial distribution of samples varies (e.g. after 2016 most samples taken in control zone)*

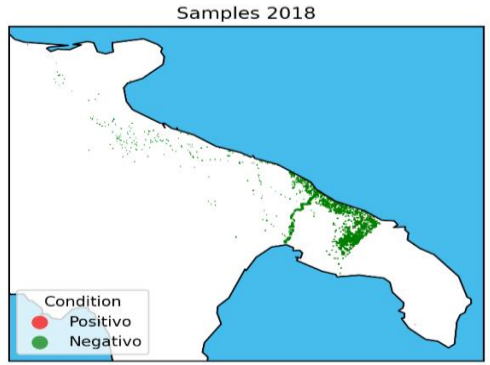

Samples 2013
Condition
Positivo
Negativo

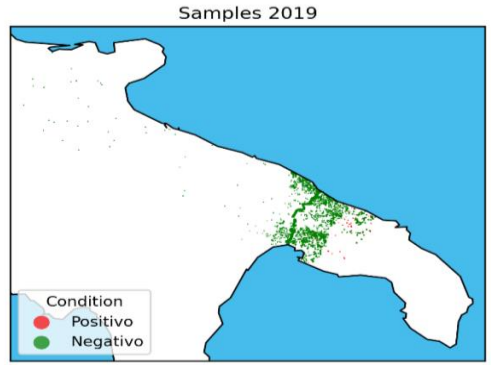

Samples 2014
Condition
Positivo
Negativo

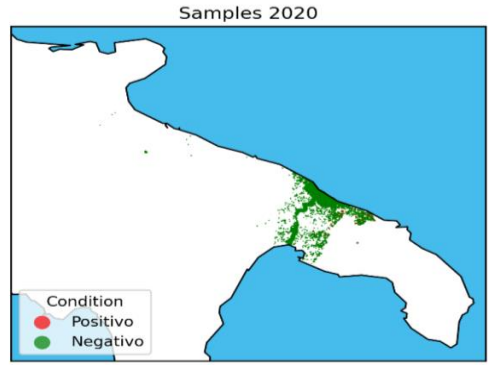

Samples 2015
Condition
Positivo
Negativo

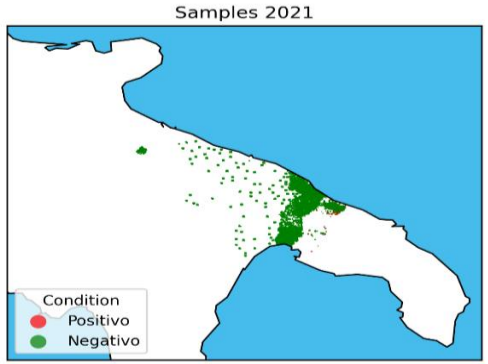

Samples 2016
Condition
Positivo
Negativo

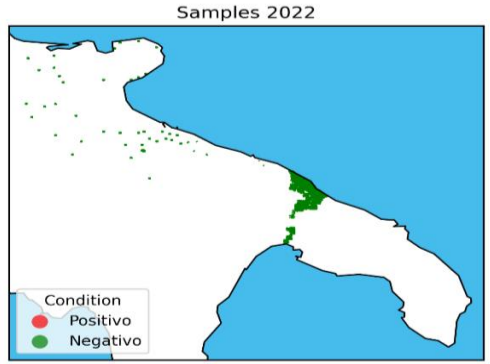

Samples 2017
Condition
Positivo
Negativo

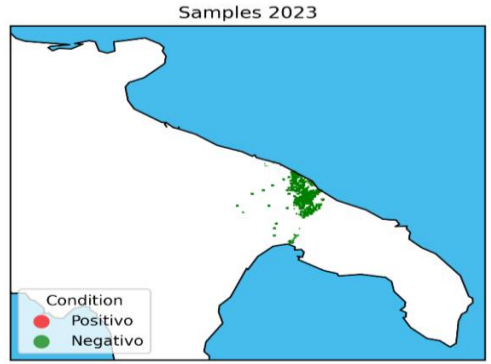

Samples 2018
Condition
Positivo
Negativo


Samples 2019
Condition
Positivo
Negativo


Samples 2020
Condition
Positivo
Negativo


Samples 2021
Condition
Positivo
Negativo


Samples 2022
Condition
Positivo
Negativo


Samples 2023
Condition
Positivo
Negativo

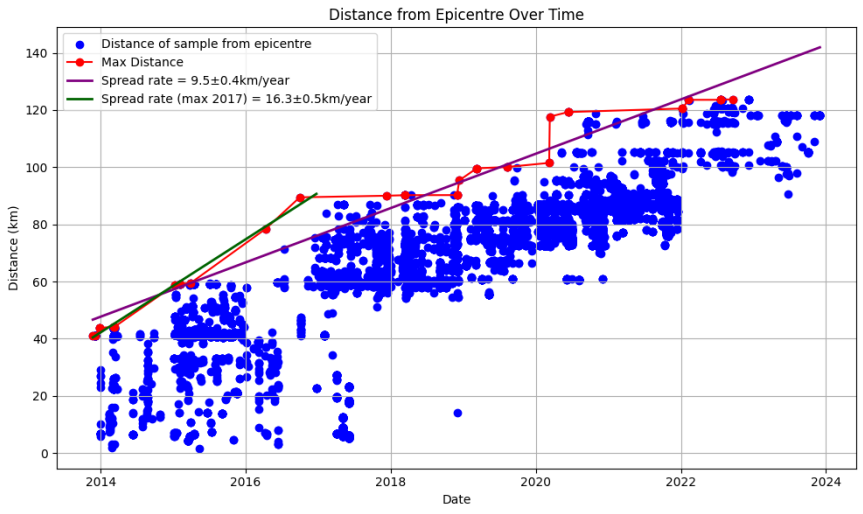
1) ESTIMATION OF EPIDEMIC FRONT SPEED

FROM MONITORING DATA!

A) Linear regression of maximum distances

Track evolution of maximum distance from epicentre over time

$$c = 9.5 \pm 0.4 km/year$$
$$c_{init} = 16.3 \pm 0.5 km/year$$



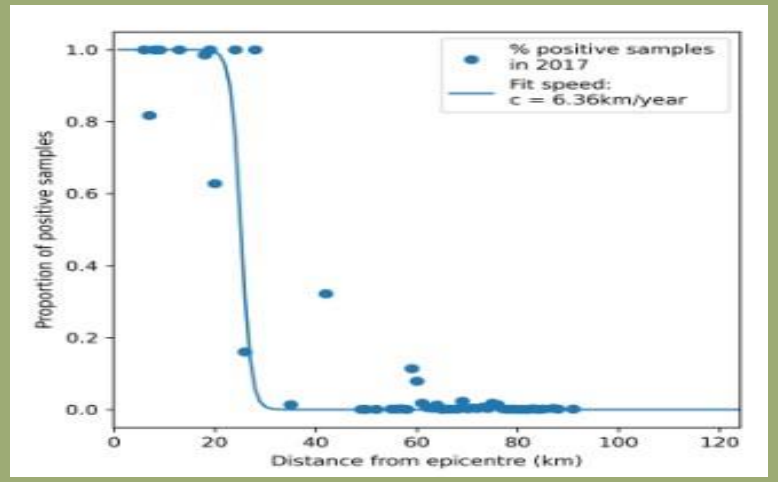
B) Kottelenberg method

Estimate shape of the invasion front: logistic function + binomial distribution

Data grouped in distance classes (x) from epicentre at time t (years from 2013)

$$p = p(x,t) = \frac{1}{1 + e^{x-(x_0+ct)}}, L(pos_d, n_d; p) = Bin(pos_d, n_d; p)$$

$$\bar{c}(2014-2018) = 6.5 km/year$$



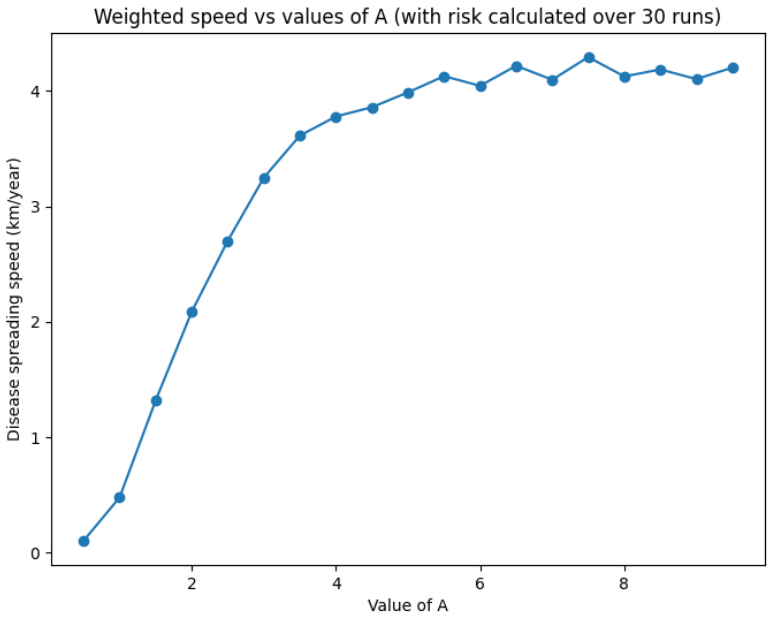
2) MODEL SENSITIVITY ANALYSIS

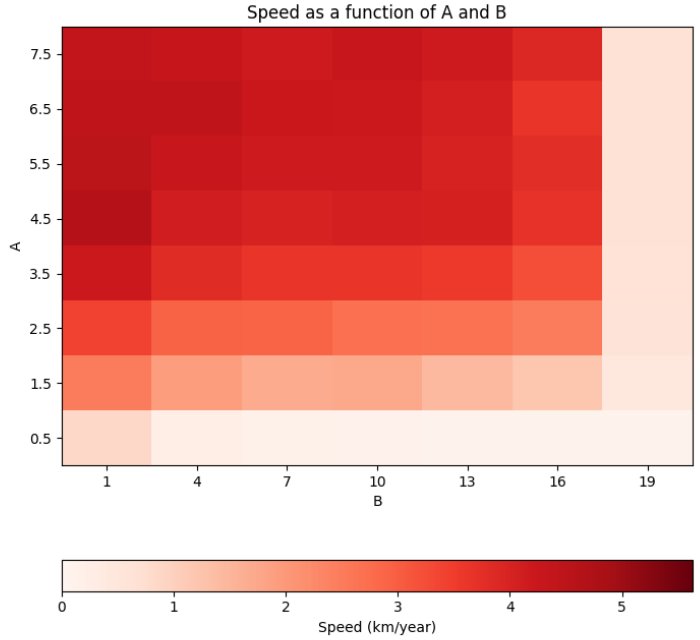
- Calculate of model disease spreading speed $\boldsymbol{c_{sim}}$:
  - *define* ***risk*** $R(x, y; t)$ *as average incidence over N simulation runs at time t in cell (x,y)*
  - *get average distance of cells from the epicentre* $(d_O(x, y; t))$*, weighted by their risk*
  - *perform linear regression to obtain* c_{sim}

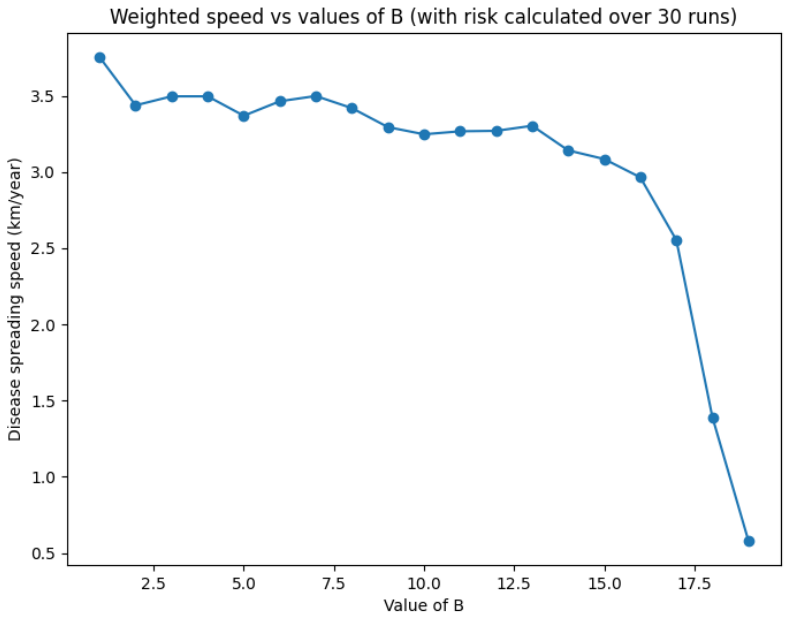
$$c_{sim} = \frac{1}{|\{infected\ cells\}|} \sum_{\substack{(x,y)\in \\ \{infected\ cells\}}} \frac{R(x, y; t) d_O(x, y; t)}{t}$$

- Parameters to explore: $\{A, B, \beta, M_{max}, D\}$

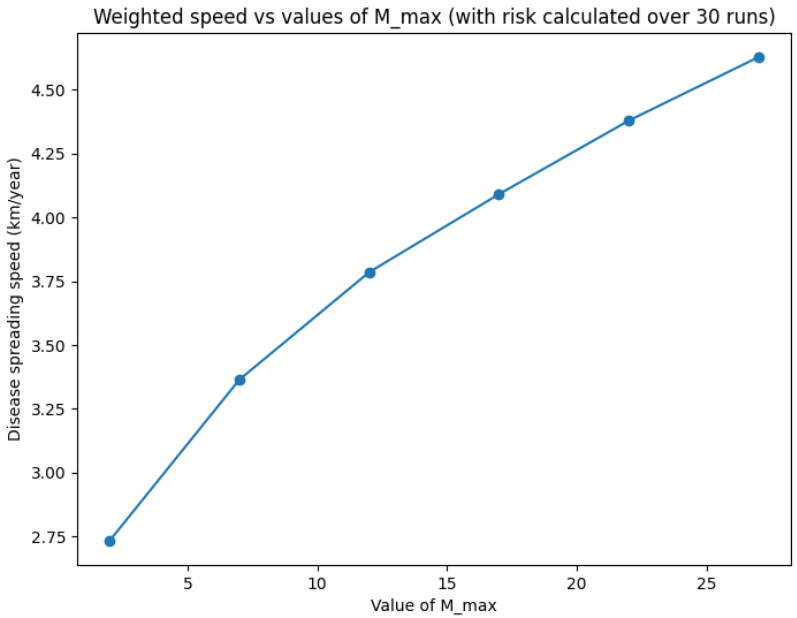
c_{sim} vs A, B

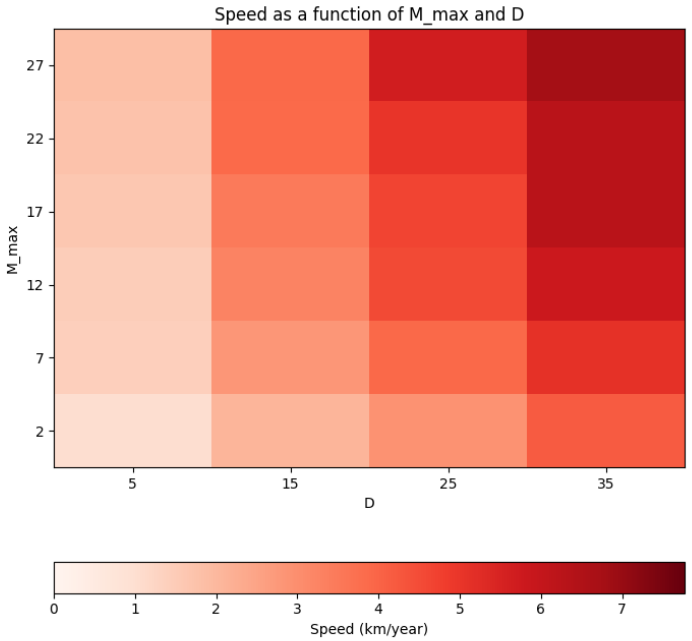

Weighted speed vs values of A (with risk calculated over 30 runs)
Disease spreading speed (km/year)
Value of A
0
1
2
3
4
2
4
6
8

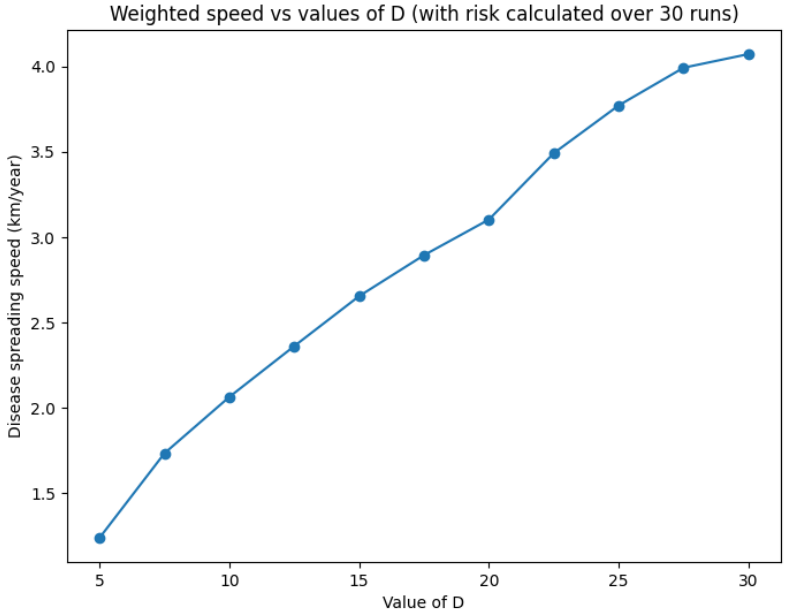

Speed as a function of A and B
A
B
Speed (km/year)
7.5
6.5
5.5
4.5
3.5
2.5
1.5
0.5
1
4
7
10
13
16
19
0
1
2
3
4
5


Weighted speed vs values of B (with risk calculated over 30 runs)
Disease spreading speed (km/year)
Value of B
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2.5
5.0
7.5
10.0
12.5
15.0
17.5

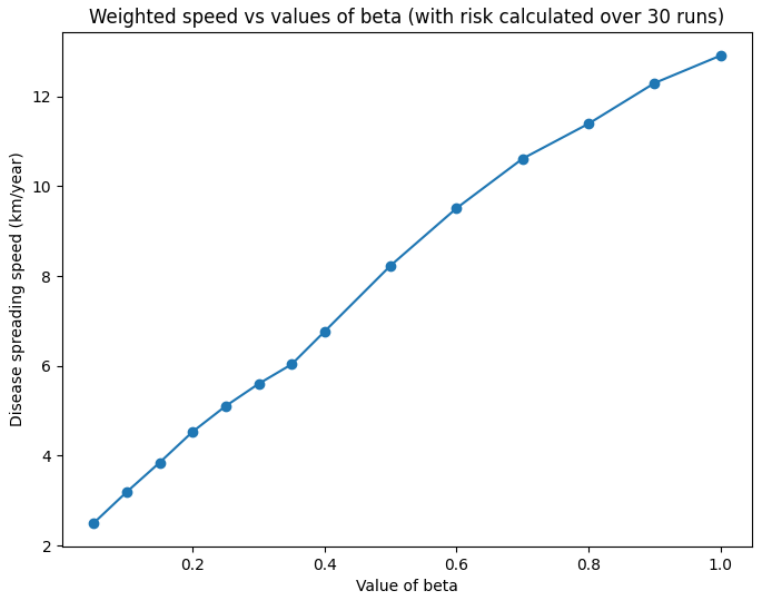
c_{sim} vs M_{max}, D







c_{sim} vs β



CONCLUSIONS ON SENSITIVITY ANALYSIS

- Rate of growth A up to $4/5$ $(year^{-1})$
- B (related to density of infected) is not relevant
- Mean short-distance dispersal β has the greatest effect
- M_{max} determinant only for large increases
- Long-distance jump variance D has also effect

A (≤ 5), β, D are the most relevant parameters

3) MODEL PARAMETER SEARCH

- Search for model parameters that give speed closest to $\boldsymbol{\bar{c} = 6.5 km/year}$
- Brutal grid search
- The parameters $\{A, \beta, D\}$ (with fixed $B = 14.069$ and $M_{max} = 5$) that minimize

$$f(A, \beta, D) = (c_{sim}(A, \beta, D) - \bar{c})^2$$

are $\boldsymbol{A = 3.0, \beta = 0.2, D = 35.}$

- These parameters give a speed of $c_{opt} = 6.40 km/year$

CONCLUSIONS

- Monitoring data is reliable but **incomplete**: estimating epidemiological parameters is difficult and requires in-depth analysis
- **Kottelenberg method** for calculating the disease spreading speed gives a more dependable estimate
- Spread model can be **simplified** as some of its parameters do not influence the spread that much
- Further avenues for analysis:
  - *varying control measures*
  - *introducing an incubation period*