

Modeling of HLB

(and H5N1 Qnets)

ZeD Lab - UChicago

Nicholas Sizemore

Univ. of Florida

2021-04-14

About

- PhD candidate, Mathematics, University of Florida
- Data analyst, Dept. of Emergency Medicine, University of Florida
- Lecturer, Dept. of Mathematics, Western Carolina University

Research Interests

- Applied/computational problems - theory and practice, esp:
 - Stochastic processes and modeling
 - ML/data science theory and interdisciplinary applications
 - Computationally driven mathematical theory

Stochastic modeling of HLB

HLB

Huanglongbing (黃龍病), aka citrus greening disease

- Worldwide vector-borne disease affecting citrus
- Devastating bacterial infection: infected trees produce unusable green fruit, die early, and are economically unviable
- Vectored by *D. Citri* (psyllids); transmits to most commercial citrus cultivars (including ornamental orange jasmine)



HLB in Florida

- spread rapidly statewide due to pre-existing vector population
- caused major disruption to the industry

Economic consequences

- 49% reduction in citrus production
- thousands of jobs
- $\approx \$1$ billion per year estimated impact



Timeline of key events:

- 1997/1998: *Invasion*
 - Psyllids arrive to SW Florida (Palm Beach County)
- Early 1999: *Dispersal*
 - Psyllids continue dispersing urban corridor of SW Florida, reaching Homestead
- \leq 1999: *Disease arrival*
 - Small number of CLas-positive plants arrive to Homestead
 - Psyllids acquire CLas in Homestead
- Early 2000s: *Major dispersal*
 - Widespread, statewide dispersal of uninfected psyllids
 - Initial spread of CLas
 - heavy infestation of commercial groves
- 2005: *Detection*
 - First finding of CLas in Florida
- By 2009: *Devastation*
 - Widespread visible symptoms of infection and large blocks of citrus removed from production statewide

Objectives

To help mitigate risk in other locales, eg. CA, TX, AZ, need a more cohesive understanding of what happened (relative significance of factors driving the spread).

Challenges

- high-dimensional system acting over large and varied spatial and temporal scales
- many underlying processes which are individually complex
- sparse, irregularly collected data (historical)
 - not intended for rigorous analysis
- gaps in biological understanding
- undocumented management practices
- and many others...

Goal

a comprehensive, realistic model with a small number of biologically-plausible assumptions that qualitatively generates the spread

The Model

Overview

Structure/scope

- Stochastic; spatially explicit; discrete time/space
- Captures all major relevant processes statewide
 - From vector arrival in 1997/1998 until heavy loss of productive crop in \approx 2009

Resolution

- 0.25×0.25 mi (40 acre) grid overlaying the state; 1-acre resolution in commercial groves
- Fundamental time unit: days

Philosophy

Make a limited number of biologically defensible assumptions to describe the underlying processes (and consistent with the known timeline)

Dispersal Mechanisms

Some key processes captured by the model:

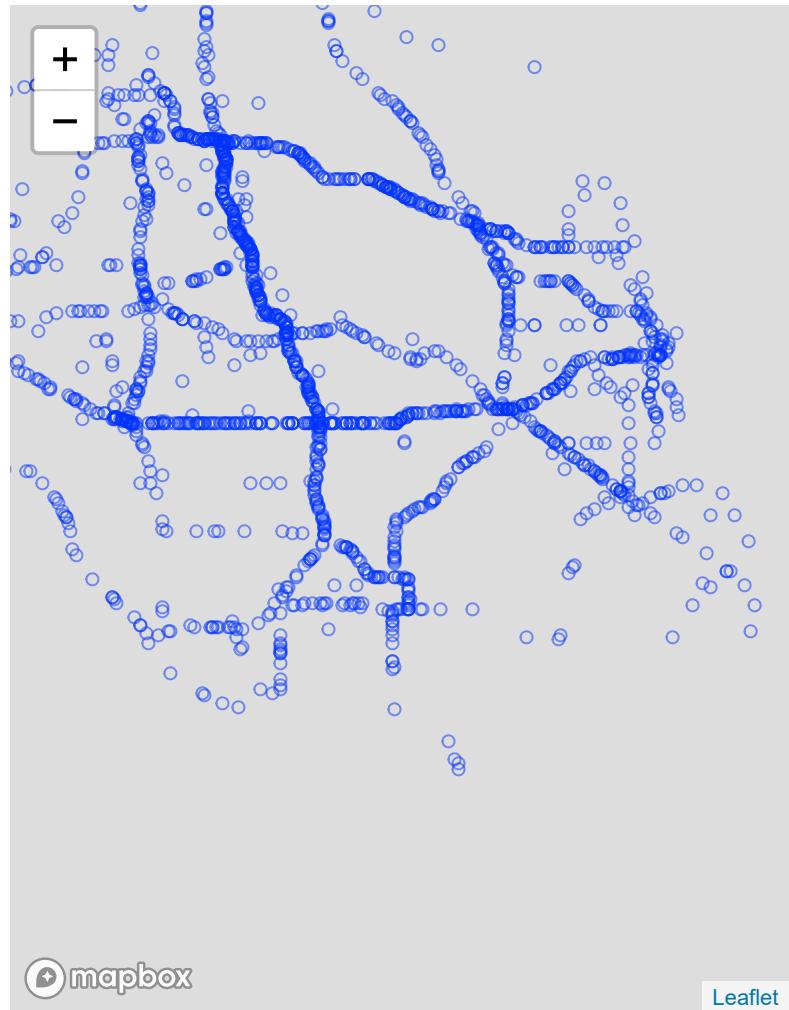
Mechanism	Implementation Details
Psyllid demography	Seasonal (monthly) variation based on location and historical temperature data
"Natural" psyllid movement	Convolution with local kernel; separate dispersal patterns in urban areas and commercial groves
Sales of citrus plants and hedges statewide to retail stores and then to consumers	Fixed store locations; two-stage/layered Poisson process based on regional population
Dispersal of psyllids via truck traffic to commercial processing plants	Harvesting process based on citrus density; routes sampled uniformly along paths to facilities

Tech Stack

- At these spatial scales (> 3 million grid cells), a significant part of the model are the convolutions used for dispersal daily
 - vector populations are stacked as tensors and we call TensorFlow for hardware accelerated performance
- On the front-end, the base model is called from R due to its excellent interface to high-level geospatial libraries

Samples

Examples of truck dispersal:



Sample Output

Infected population after 10+ years of simulation:



Observations

The model suggests that:

- Natural vector movement generates significant (but not statewide) dispersal
 - Long-range natural dispersal occurs, but is not significant
- Human-driven factors (sales of plants, trucks) expedited the spread and enabled access to regions otherwise inaccessible (northern regions in particular)
- Traditional vector control mechanisms (eg. pesticides) have limited utility in containing the spread

H5N1 Quasinet: Visualizing with Clustermaps

Objectives

For purposes of exploring H5N1 hemagglutinin sequences (sourced from [NCBI](#)) in a Qnet framework, we:

- Construct a Q-net
- Look at the Q-distance matrix
- Visualize using `seaborn.clustermap()`

H5N1 experiment

Took 200 recently collected H5N1 hemagglutinin nucleotide sequences

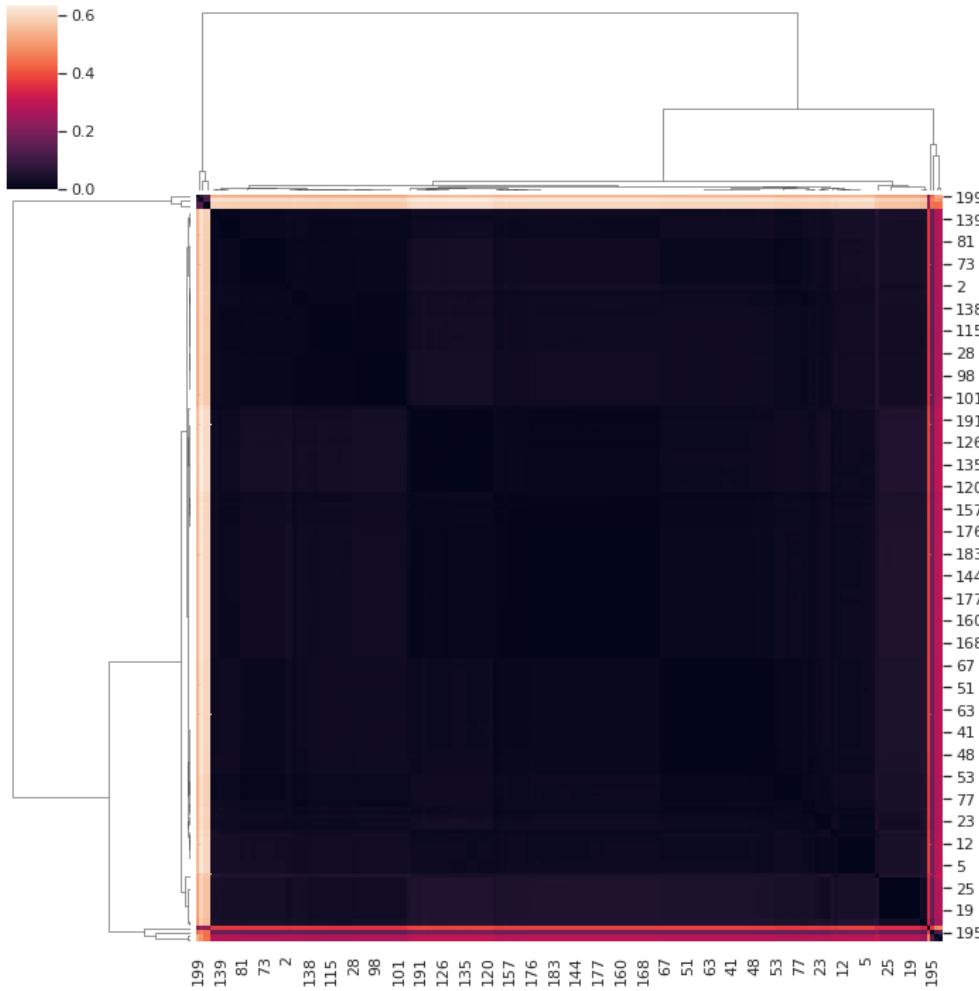
- Hosts: *Gallus gallus* (chickens) sequences collected in Asia
- Aligned, kept all features
- Generated Qnet using `quasinet.qnet.Qnet()`
- Got distance matrix using `quasinet.qnet.qdistance_matrix()`

To compare, also took a wider cohort of 200 H5N1 protein sequences from:

- Any species
- All geographic regions

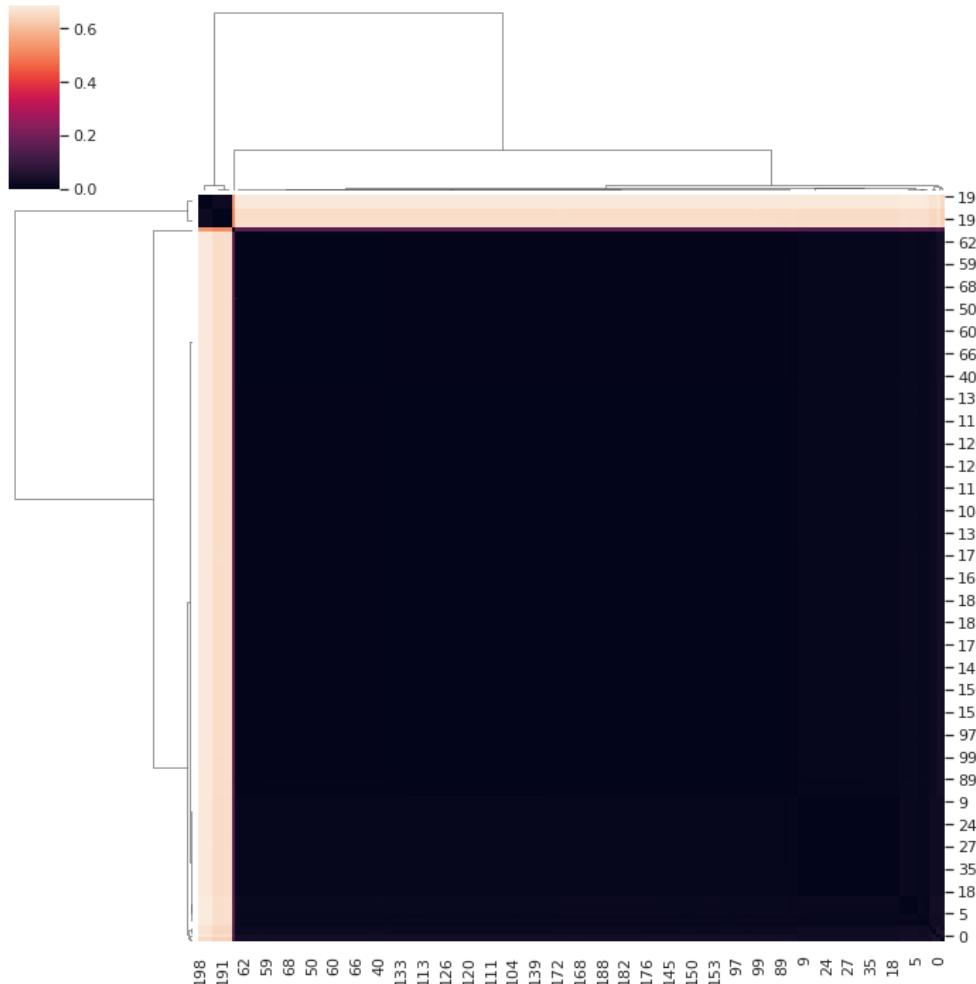
Gallus gallus-hosted clustermap

- Vast majority of sequences are tightly clustered wrt q-distance



Multi-species protein clustermap

- Similar results, though a bit more variation:



Thanks!