A Multiscale Stochastic Model of Statewide HLB Spread in Florida

Nicholas Sizemore
Dept. of Mathematics - Univ. of Florida
July 7, 2021

Overview

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

Some known, some informed conjecture:

- Early 1998: **Invasion**
  - Uninfected psyllids arrive to SE Florida (Palm Beach County)
- 1998-99: **Dispersal**
  - Uninfected psyllids continue dispersing urban corridor of SW Florida, reaching Homestead
- $\leq$ early 2000s: **Disease arrival**
  - Small number of *C*Las-positive plants arrive to Homestead
- Early 2000s: **Disease acquisition**
  - Psyllids acquire *C*Las in Homestead
  - Widespread, statewide dispersal of uninfected psyllids (heavy infestation of commercial groves)
  - Initial spread of *C*Las

Timeline of key events

- Mid 2000s: **Disease spread**
  - Statewide dispersal of infection faciliated by plant sales and truck traffic
- 2005: **Detection**
  - First finding of *C*Las in Florida
- By 2009: **Devastation**
  - Widespread visible symptoms of infection and large blocks of citrus removed from production statewide

Objectives

To 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...

Objectives

- This makes precise quantitative model validation/parameter estimation unfeasible
- But the timeline offers constraints on the set of possible models

Goal

- Comprehensive, realistic model with a small number of biologically-plausible assumptions that reproduces the spread

The Model

Overview

Philosophy

- Make a limited number of biologically defensible assumptions (i.e. consistent with the known timeline) to describe the underlying processes

Structure/scope

- Stochastic; spatially explicit; discrete time/space
- Captures all major relevant processes statewide
    - From vector arrival in 1998 until major loss of productive crop (through $\approx$ 2010)

Overview

Spatial Resolution

- 0.25×0.25 mi (40 acre) grid overlaying the state; 1-acre resolution in commercial groves
- Partition statewide grid:
    1. urban/developed areas
    2. citrus areas (eg. commercial groves)
    3. areas with no host material/inhospitable to psyllids (vectors)
- Fundamental time unit: days

Model components

Key processes captured by the model include:

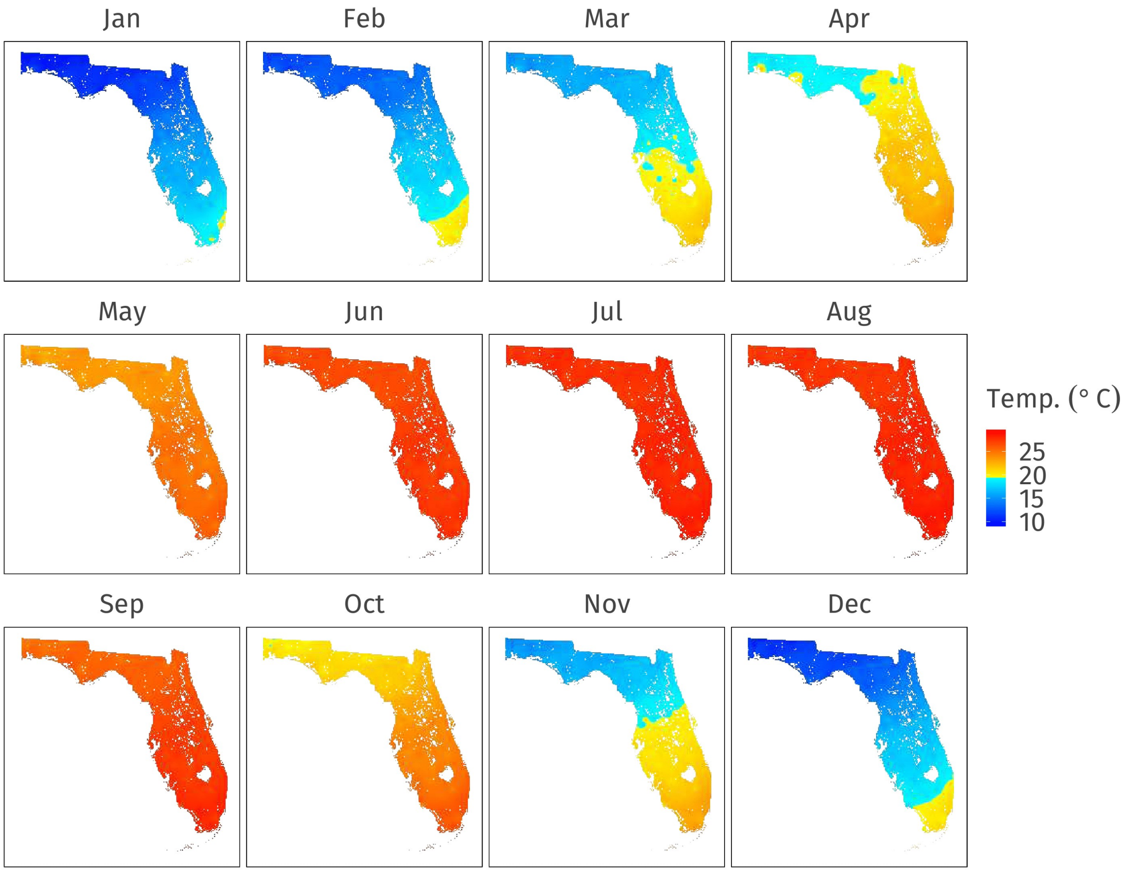
Mechanism	Implementation Details
Psyllid demography	Seasonal (monthly) variation based on location and historical temperature data
"Natural" psyllid movement	Convolution with local kernel; separate dispersal patterns in urban areas and commercial groves
Sales of citrus plants and hedges statewide to retail stores and consumers	Fixed store locations; two-stage/layered Poisson process based on regional population; store-specific sales distributions

Model components

Mechanism	Implementation Details
Dispersal of psyllids via trucks to commercial processing plants	Seasonal harvesting process based on citrus density; routes sampled uniformly along paths to facilities
Infection to removal from production	Time to visible symptoms + Time to removal

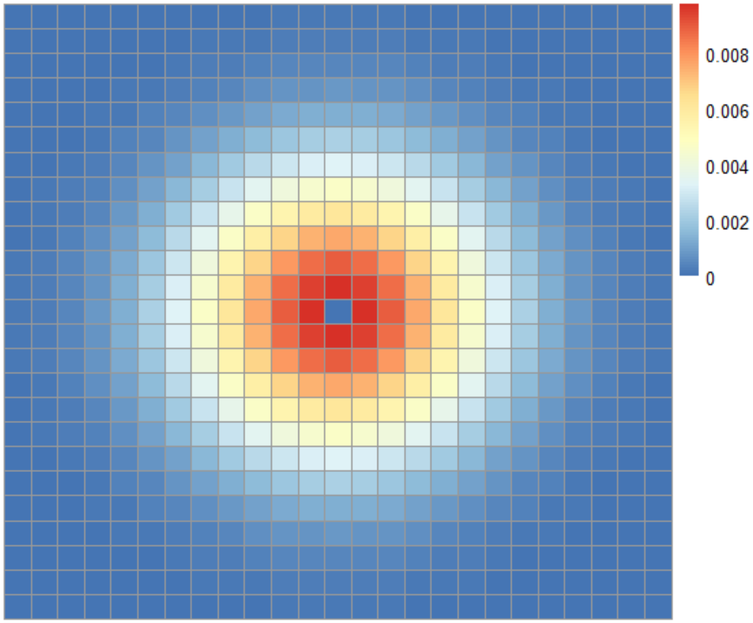
Demography

Parameters vary spatially according to the partition and are adjusted monthly based on temperature

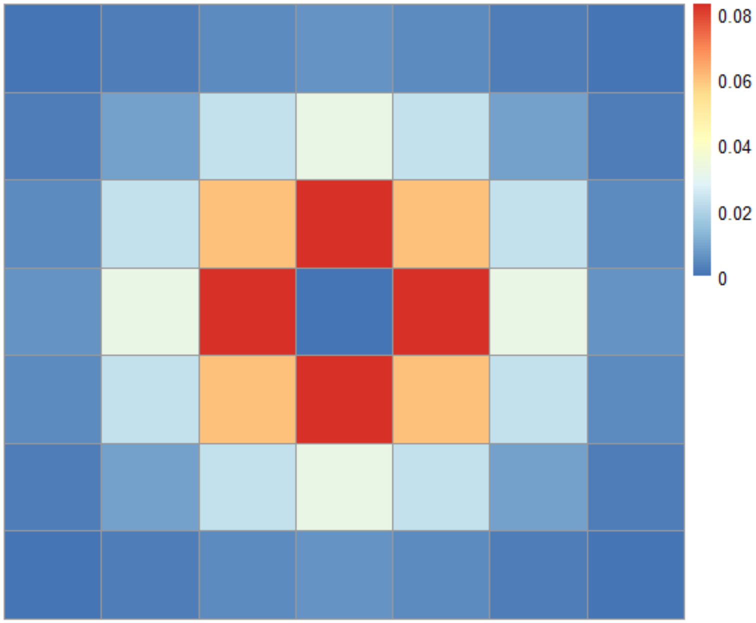


Dispersal kernel

Separate dispersal kernels in urban and citrus areas:



Urban

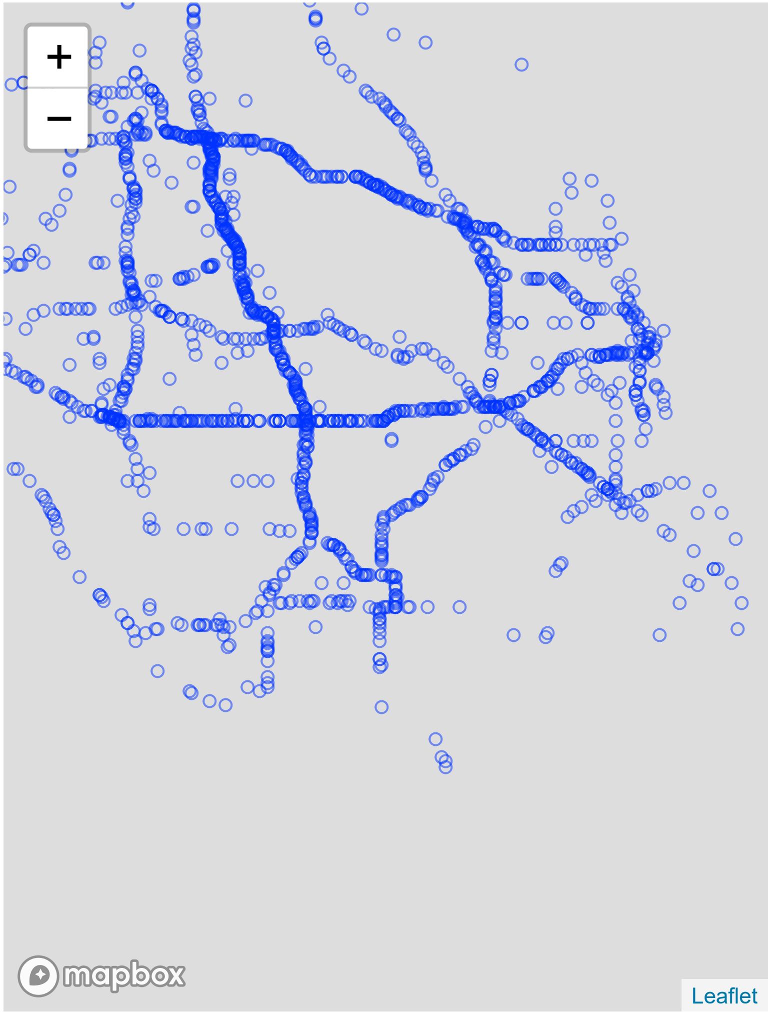


Citrus

Distributional assumptions

Process/Time	Type/Distribution	Description
$S(t)$	Poisson process (arrivals $\sim \mathrm{Pois}(\lambda)$)	Count of sales from Homestead to retail stores
$W(t)$	Poisson process (arrivals $\sim \mathrm{Pois}(\lambda_w)$)	Count of long-range wind dispersals
$S_r(t)$	Poisson process (arrivals $\sim \mathrm{Pois}(\lambda_r)$)	Count of sales from store r to consumers
$V_{(x,y)}$	Waiting time $\sim \mathrm{Gamma}(a(x,y), s(x,y))$	Time to visible symptoms at (x,y)
$R_{(x,y)}$	Waiting time $\sim \mathrm{Exp}(\lambda_r)$	Time to removal from production at (x,y)

Truck Dispersal Sampling

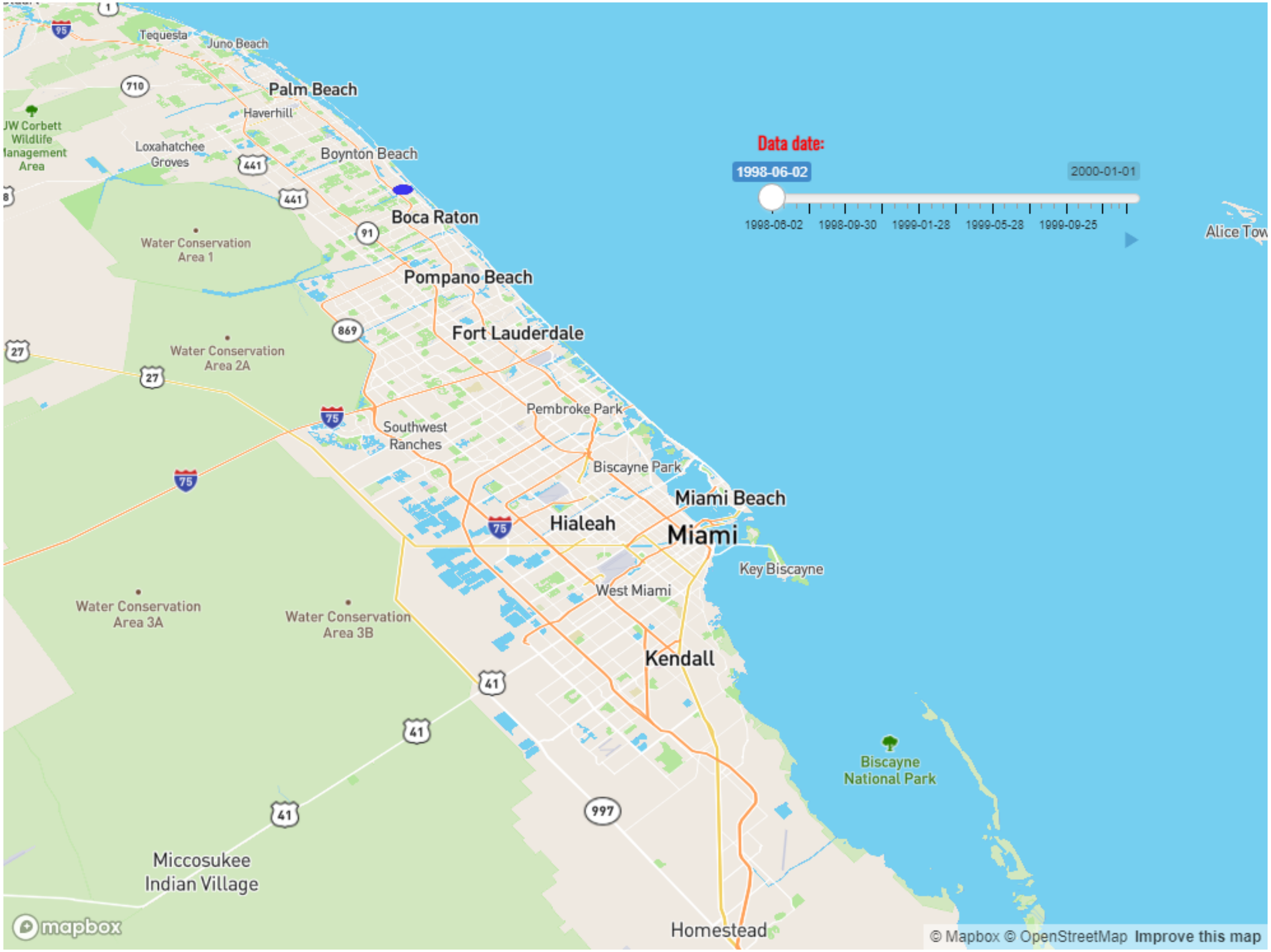


Tech + Programming Stack

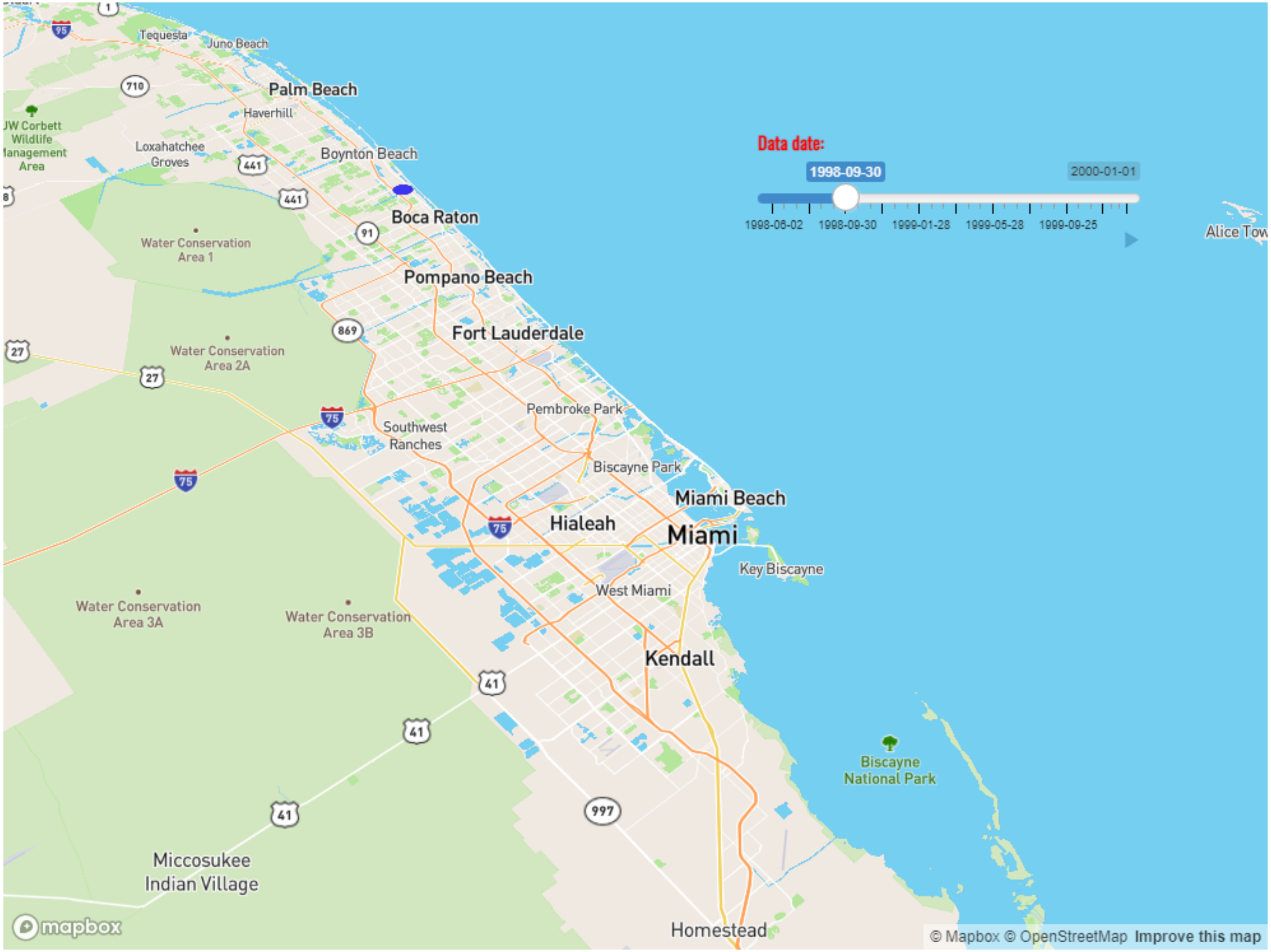
- At these spatial scales (> 3 million grid cells), a significant part of the model's computational expense are the convolutions used for dispersal daily
  - vector populations are stacked as tensors and we call TensorFlow for hardware accelerated performance
- On the front-end, the base model is called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

Model Assessments

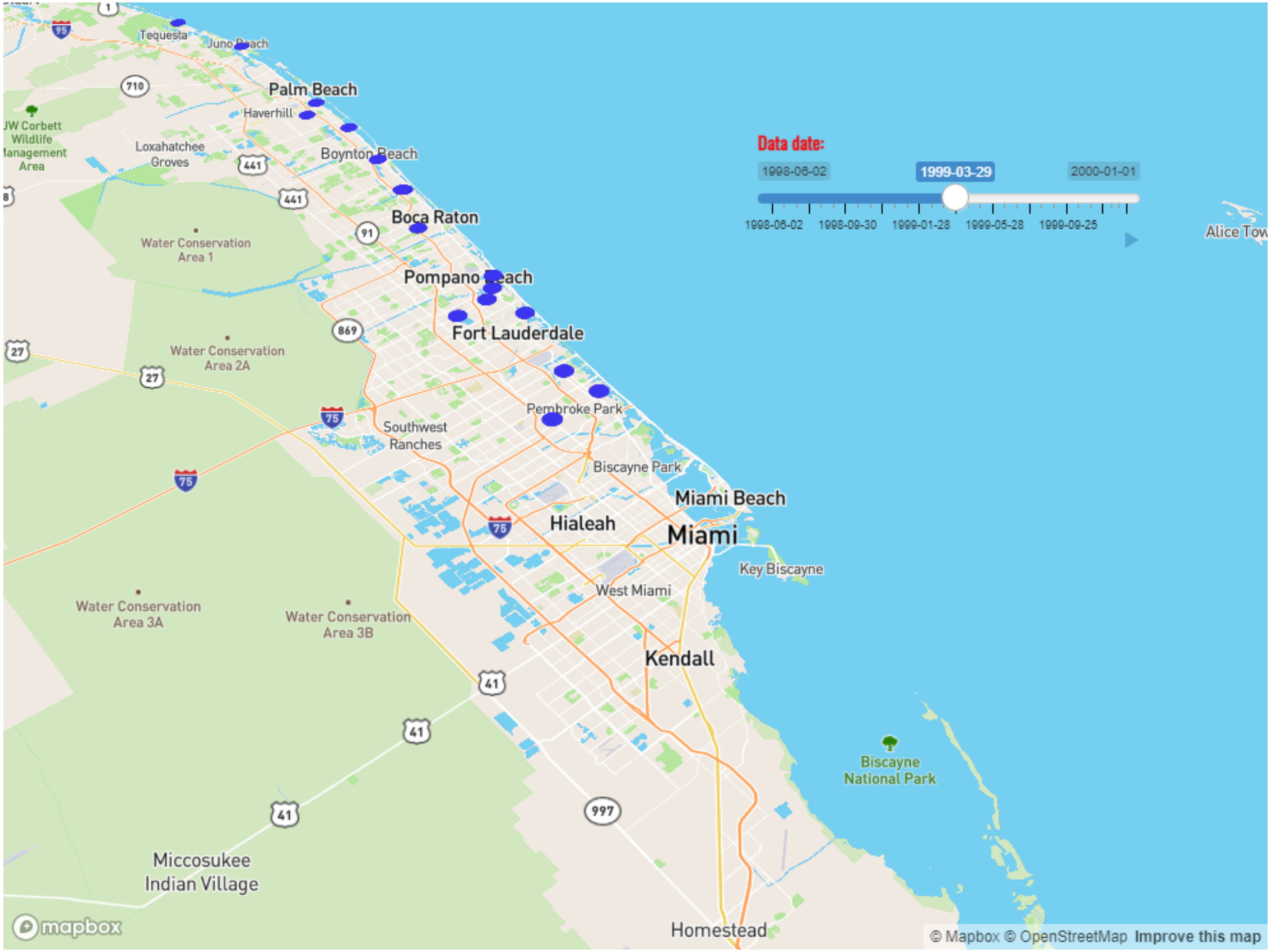
Initial Invasion - DPI data set - Jun 2, '98



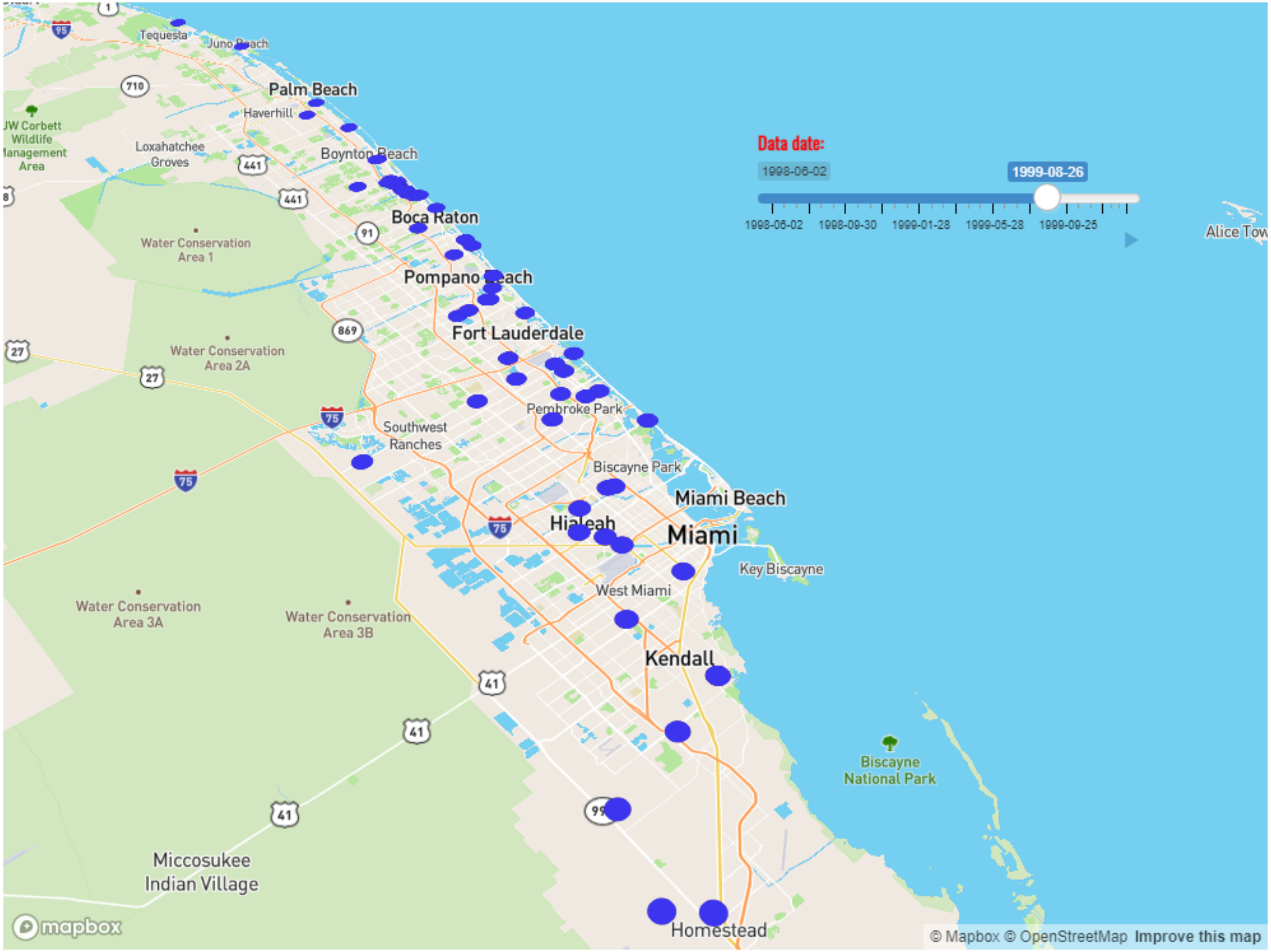
Initial Invasion - DPI data set - Sep 30, '98:



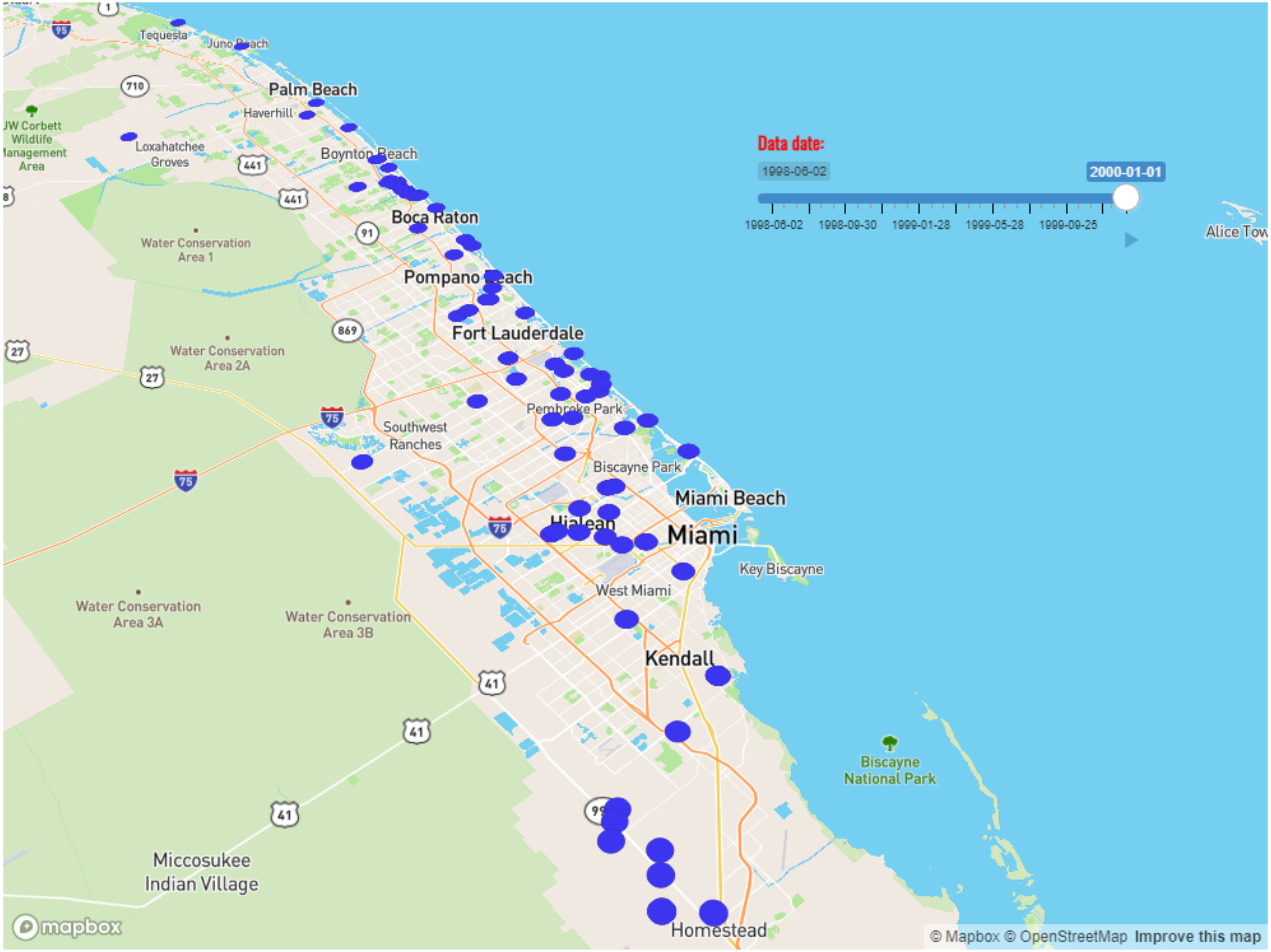
Initial Invasion - DPI data set - Mar 29, '99



Initial Invasion - DPI data - Aug 26, '99



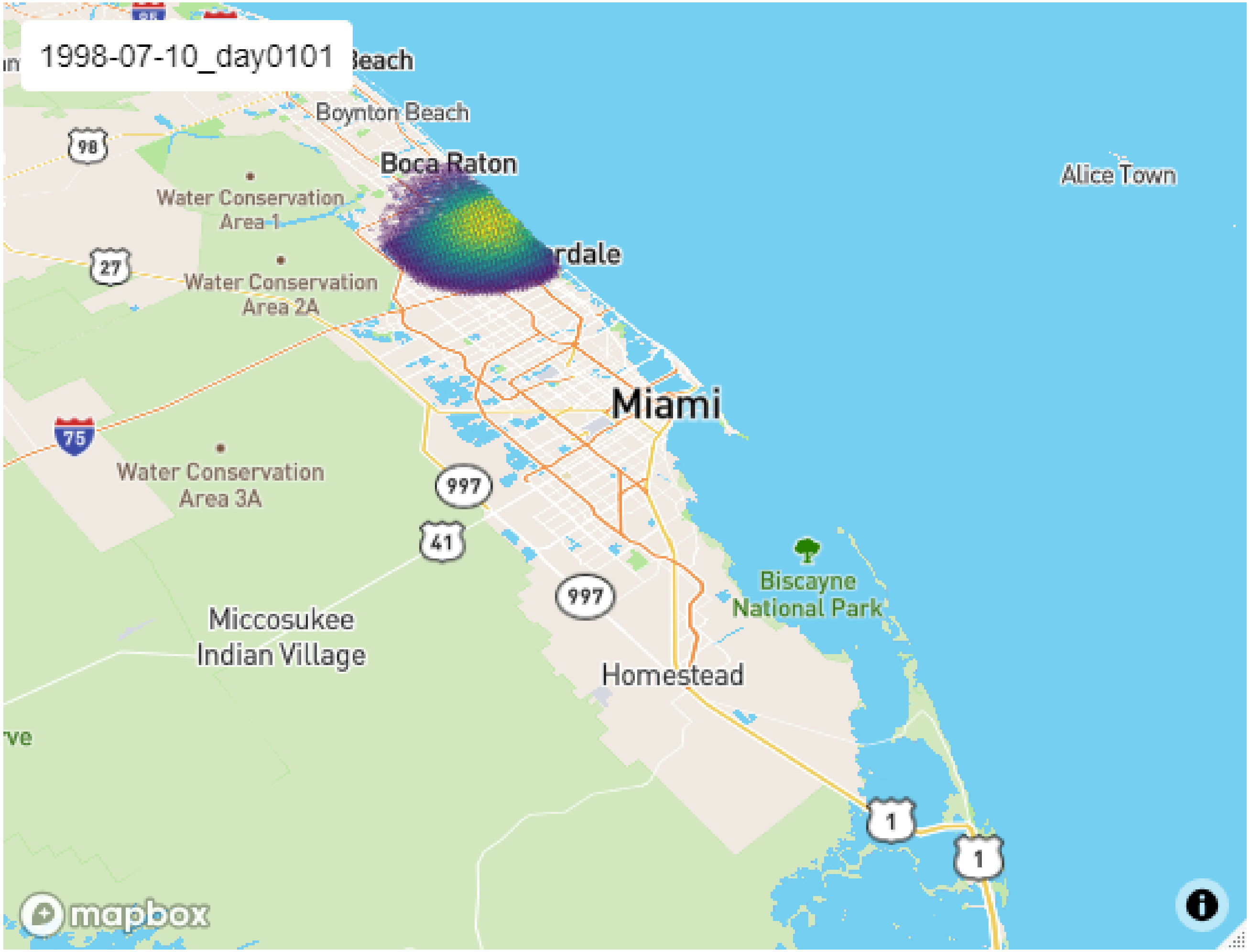
Initial Invasion - DPI data - Jan 1, '00



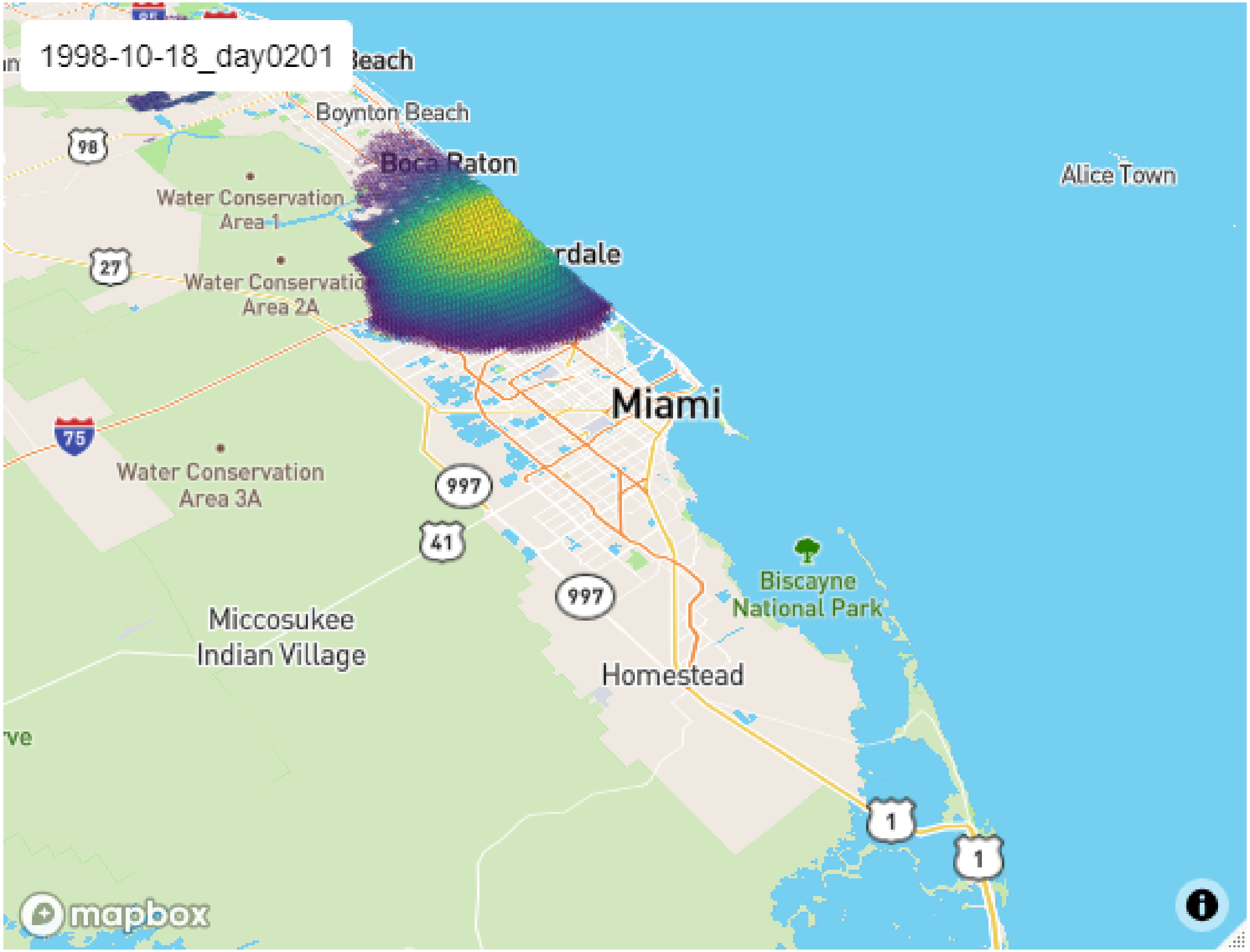
Summary of Features

- After detection in June 1998, significant dispersal along the urban corridor was apparent by March 1999.
- By August 1999, psyllids have reached Homestead, FL.
- Throughout this time period, density of samples continues to increase along the corridor.
- By 1999 or 2000, uninfected psyllids have exited the corridor north and invaded commercial groves.

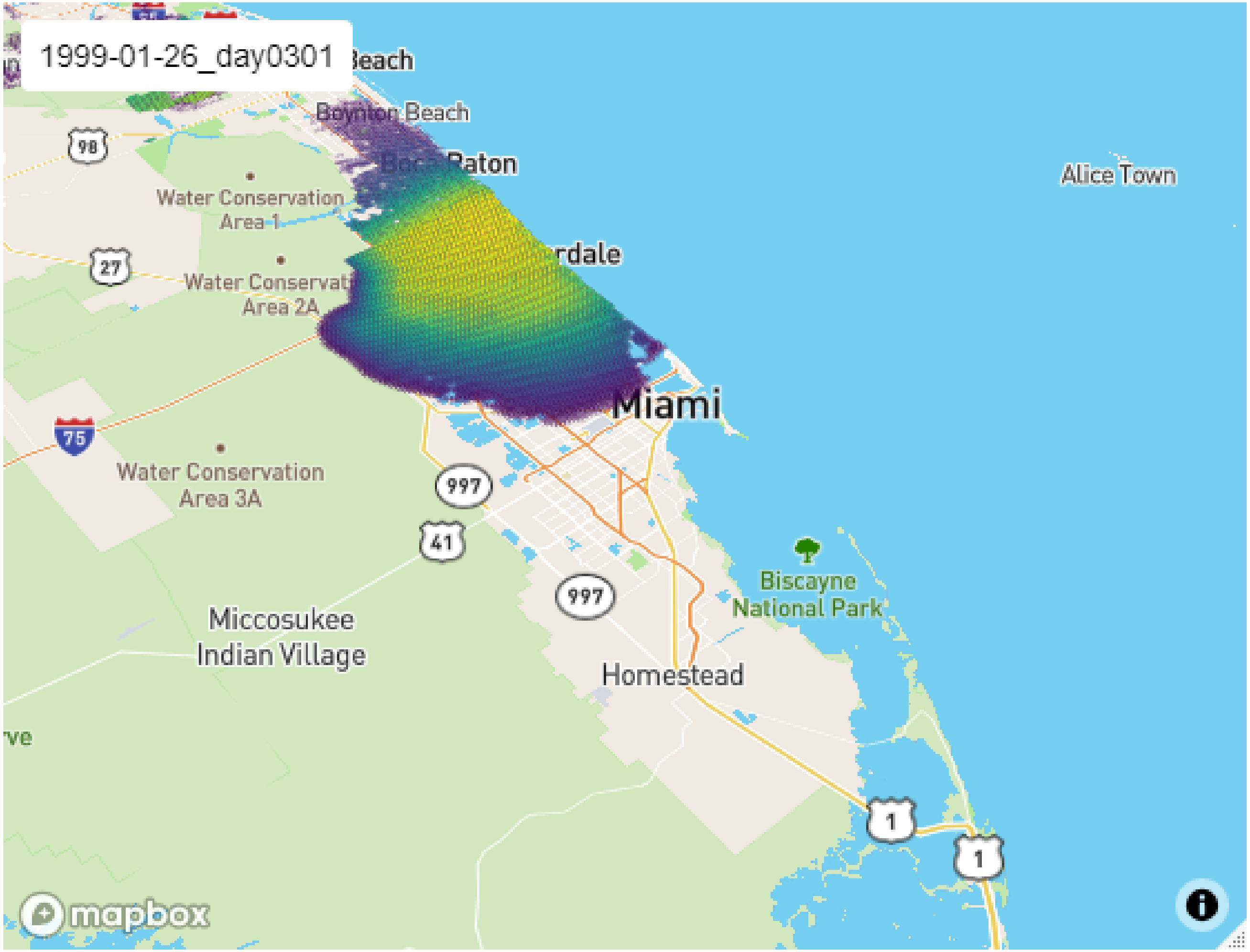
Initial Invasion - Model output - Jul '98



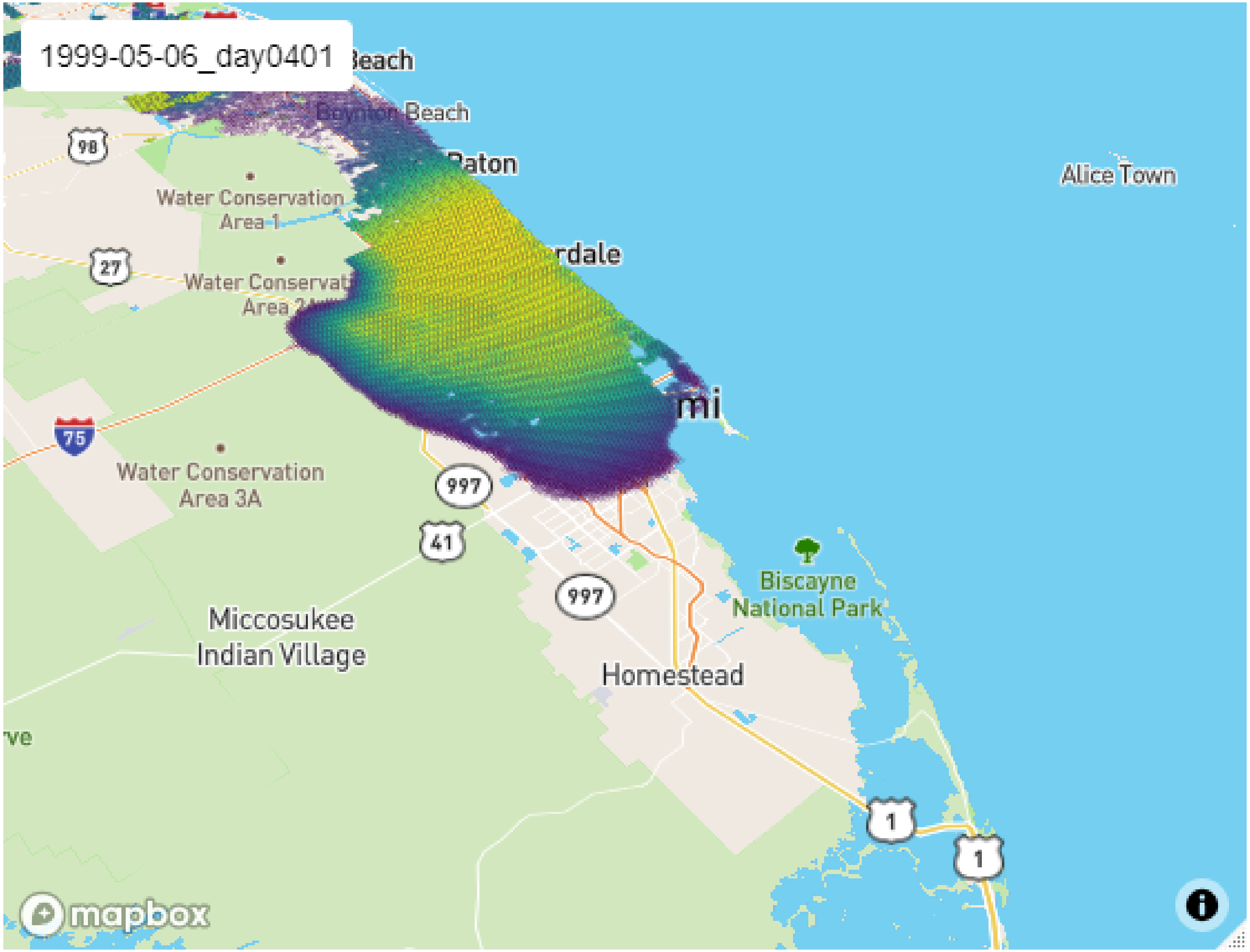
Initial Invasion - Model output - Oct '98



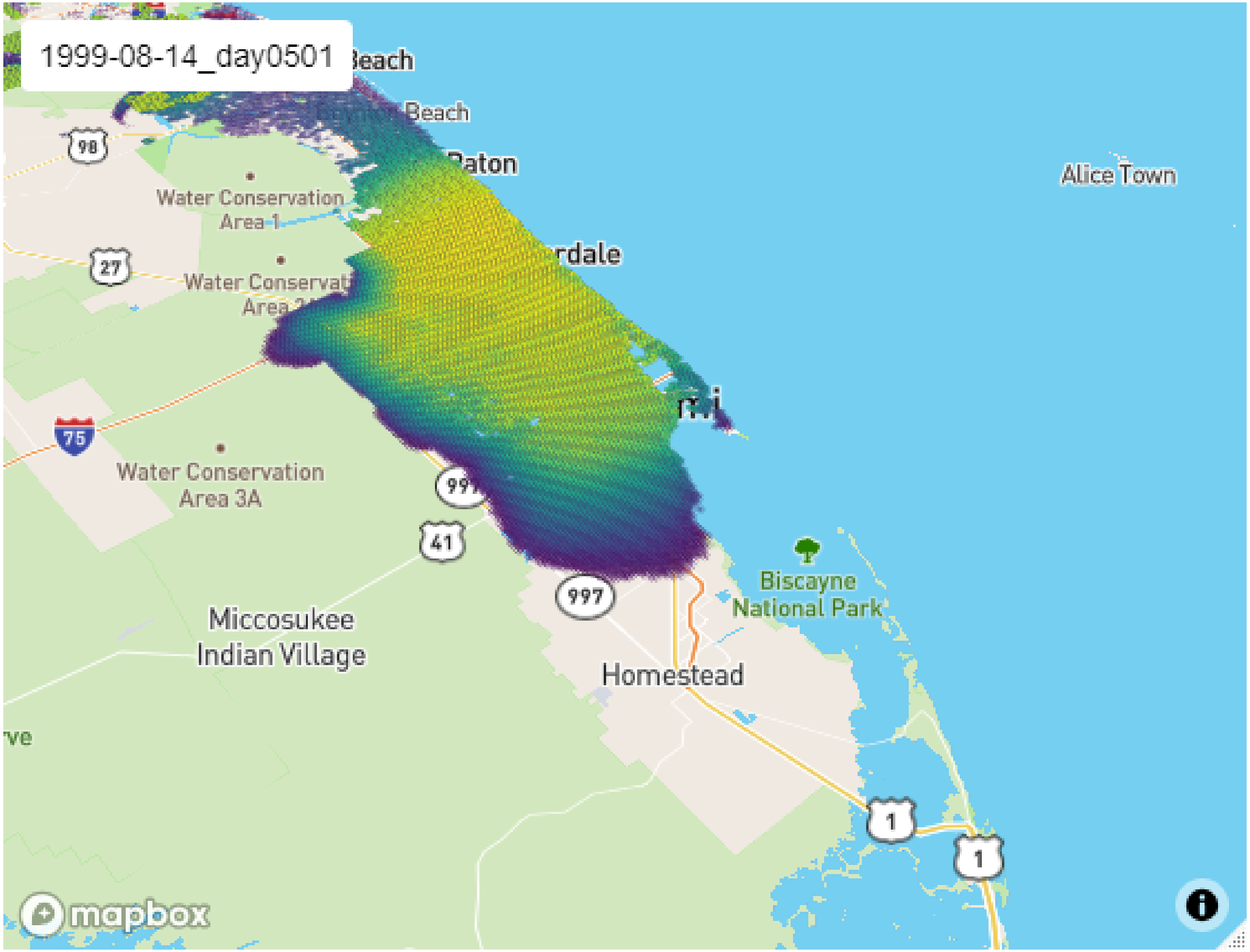
Initial Invasion - Model output - Jan '99



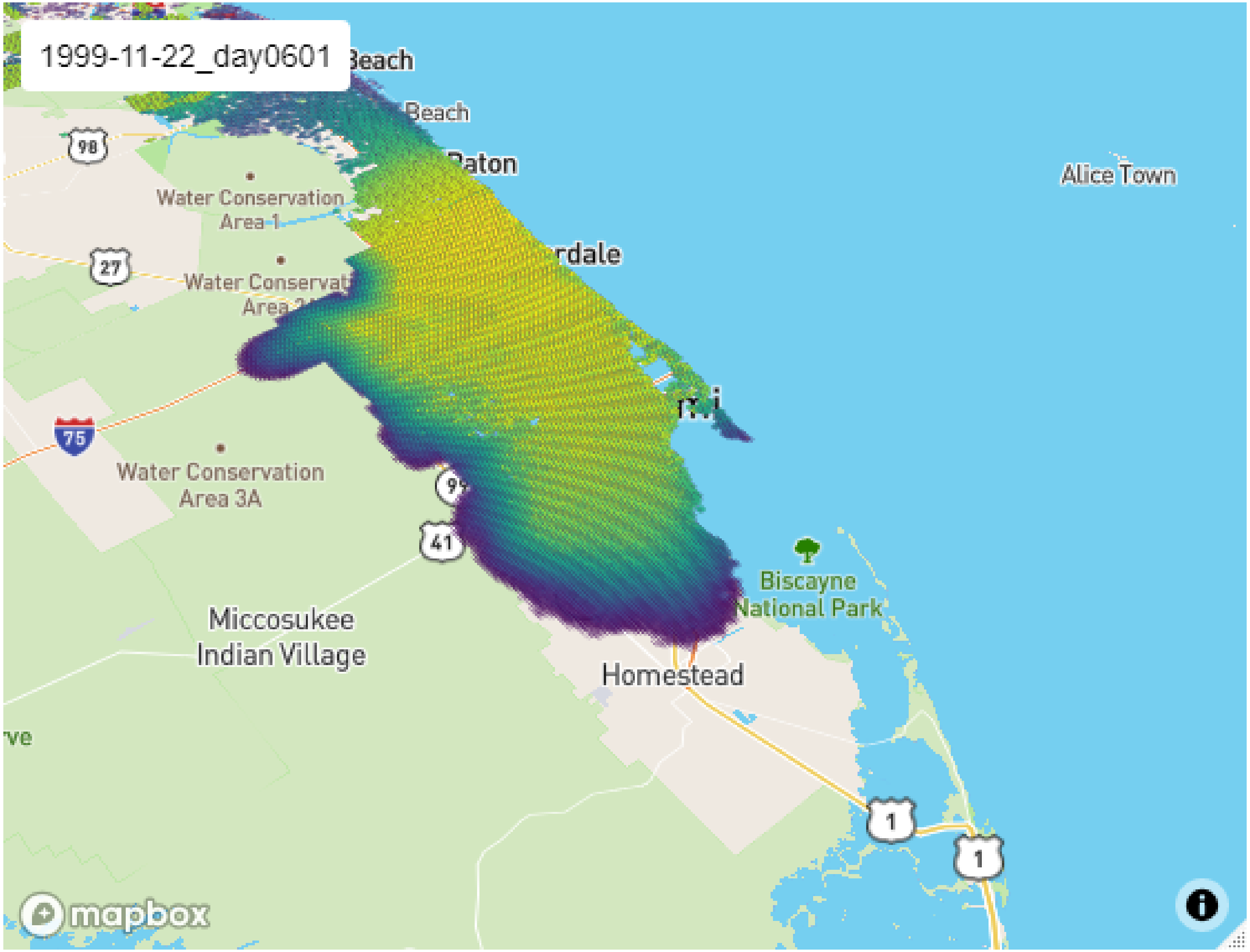
Initial Invasion - Model output - May '99



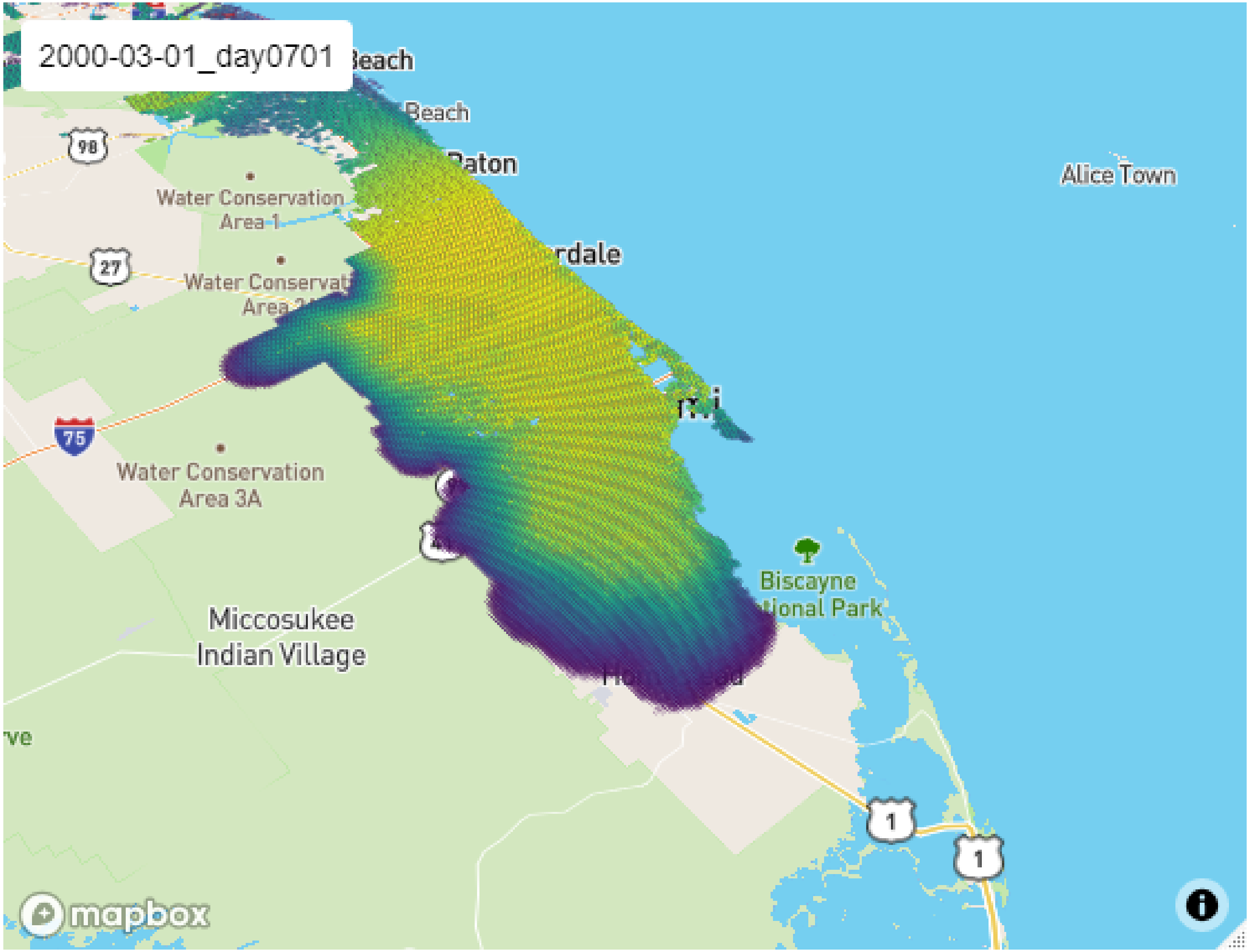
Initial Invasion - Model output - Aug '99



Initial Invasion - Model output - Nov '99



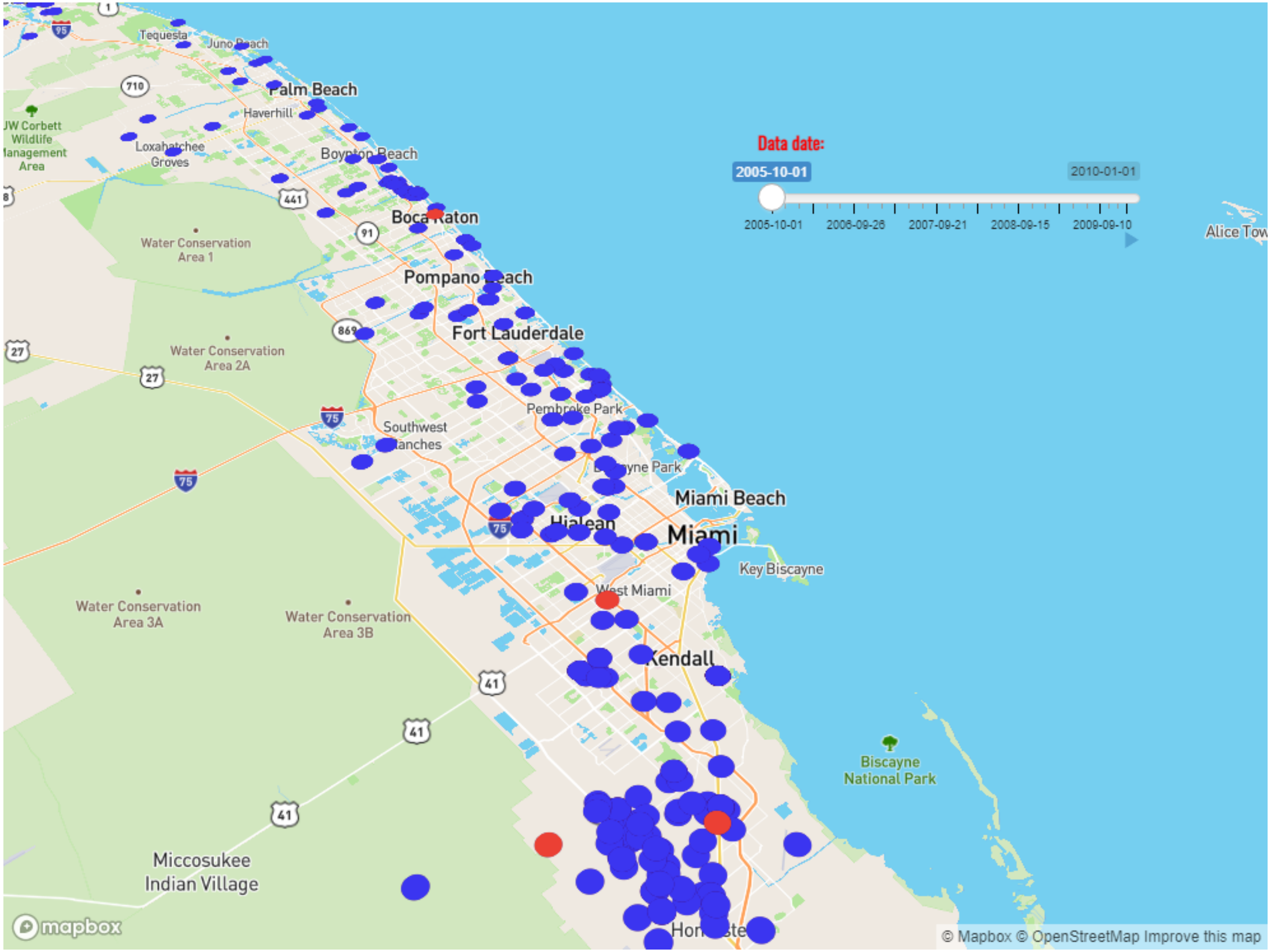
Initial Invasion - Model output - Mar '00



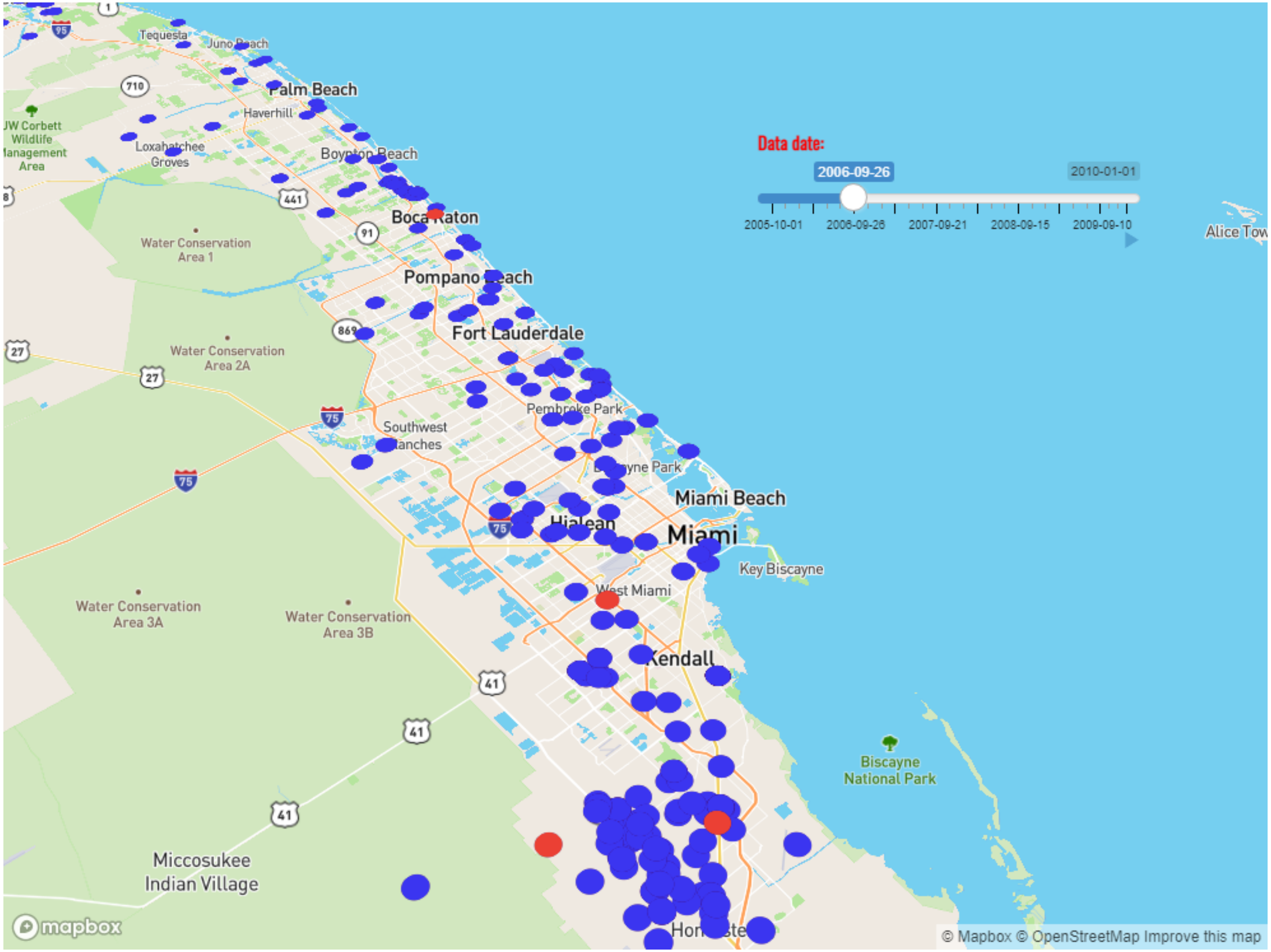
Model vs. Empirical Features

- Range of first few dispersal steps is excellent match
- Model reaches Homestead by early 2000 (slightly later than data)
- Model generates smoother dispersal pattern (by construction)
- Both model and data show uninfected psyllids exiting the urban corridor and entering commercial groves by this point

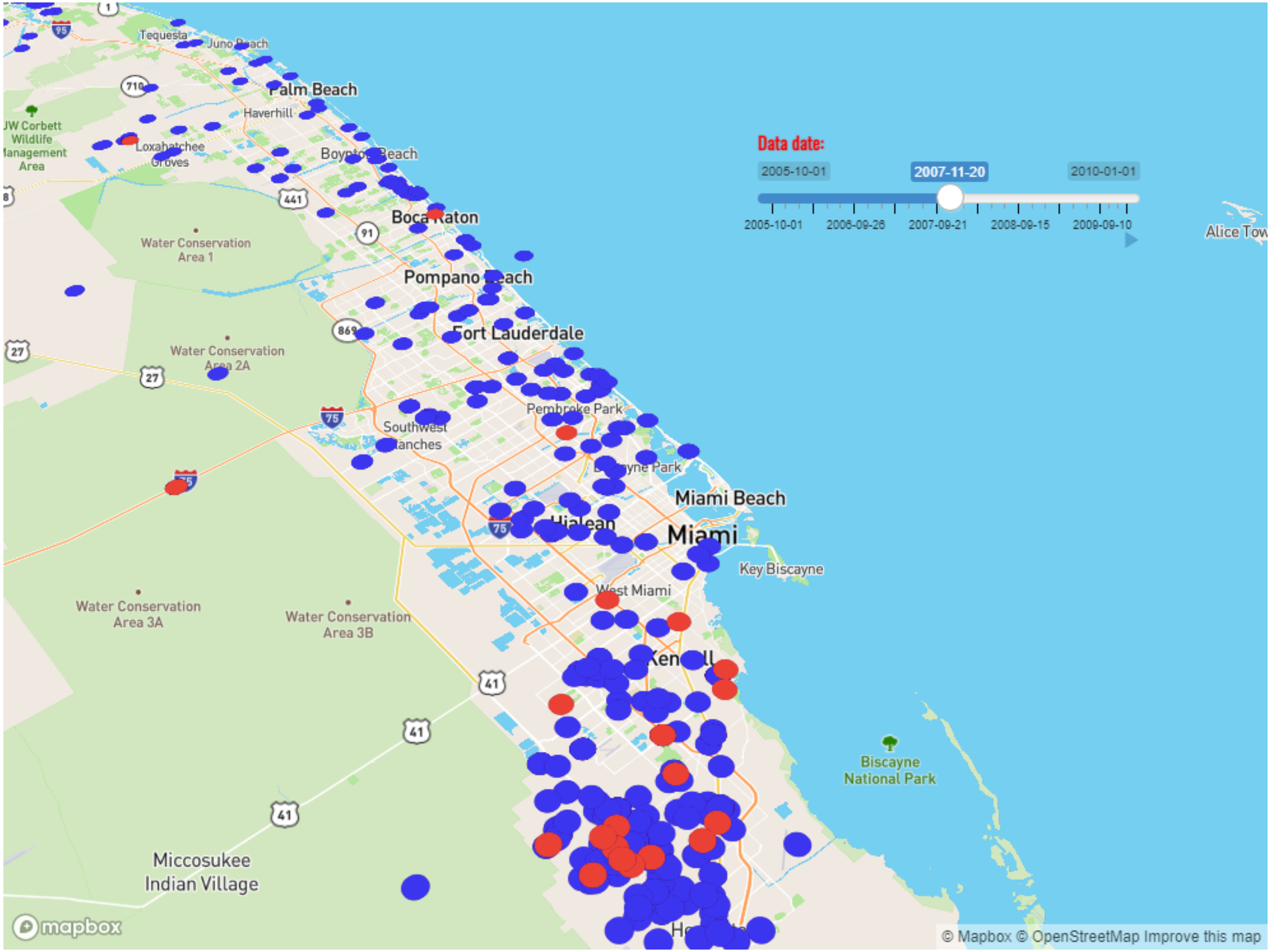
Northward Migration - DPI data - Oct 1, '05



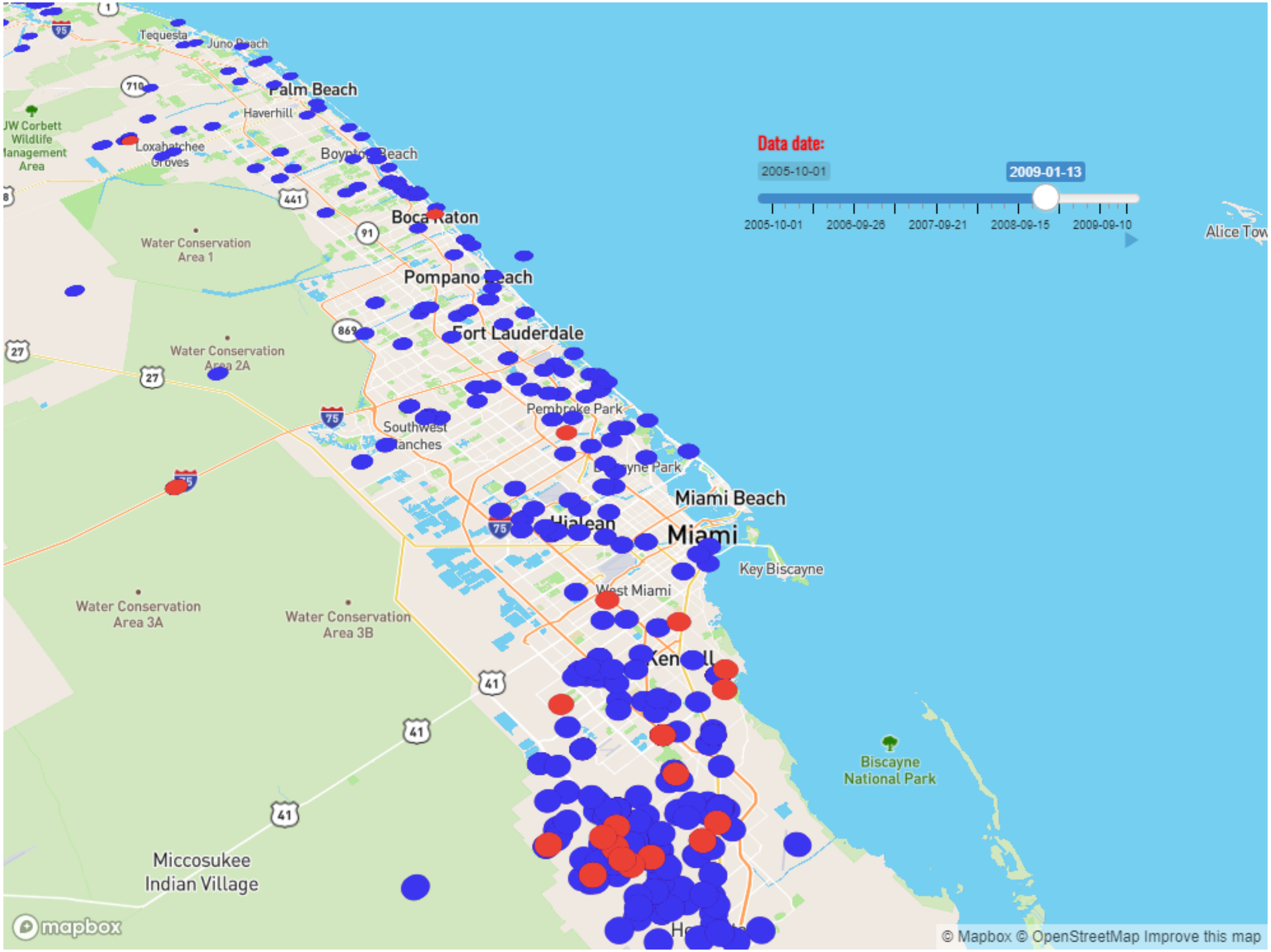
Northward Migration - DPI data - Sep 26, '06



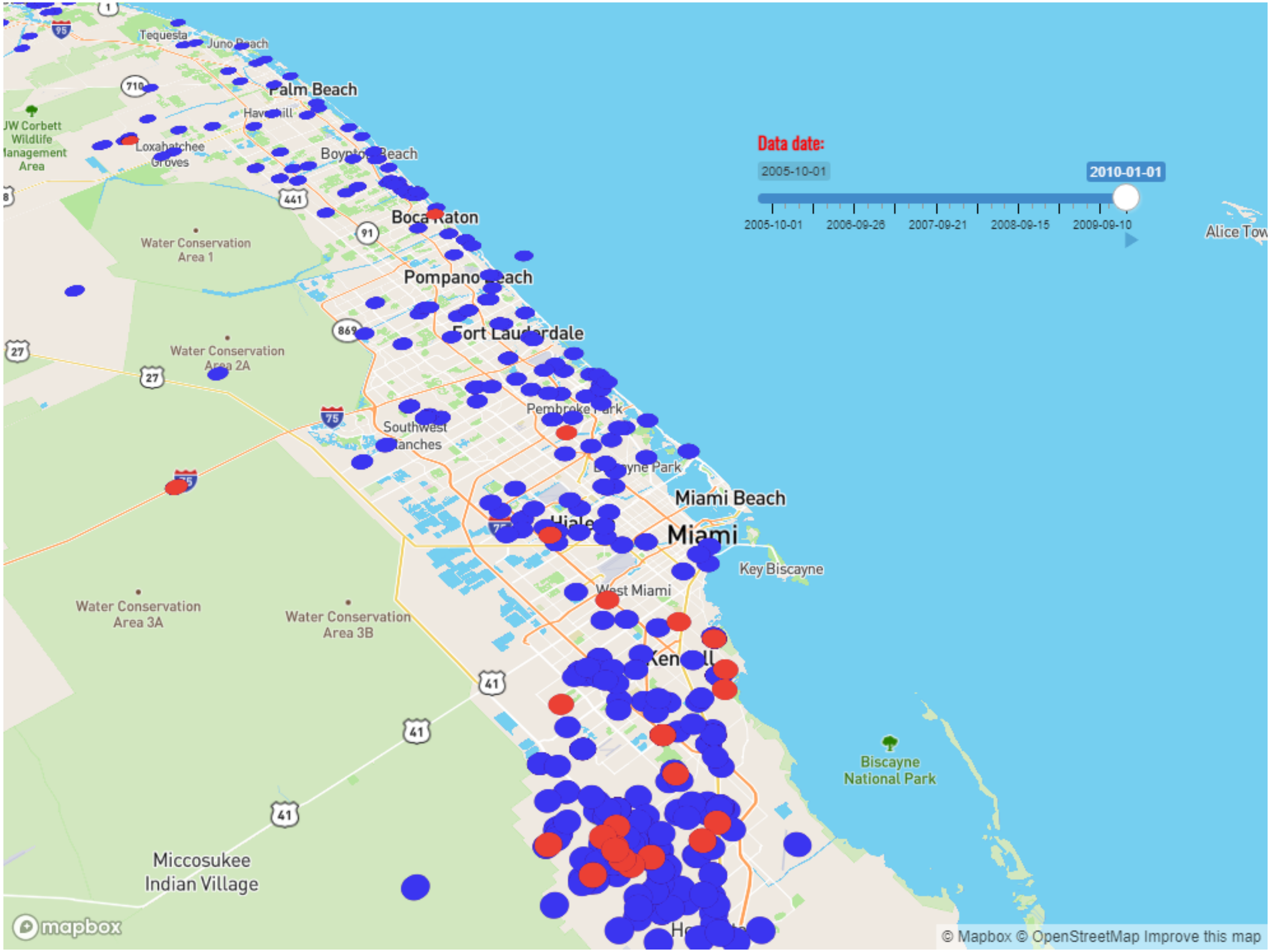
Northward Migration - DPI data - Nov 20, '07



Northward Migration - DPI data - Jan 13, '09

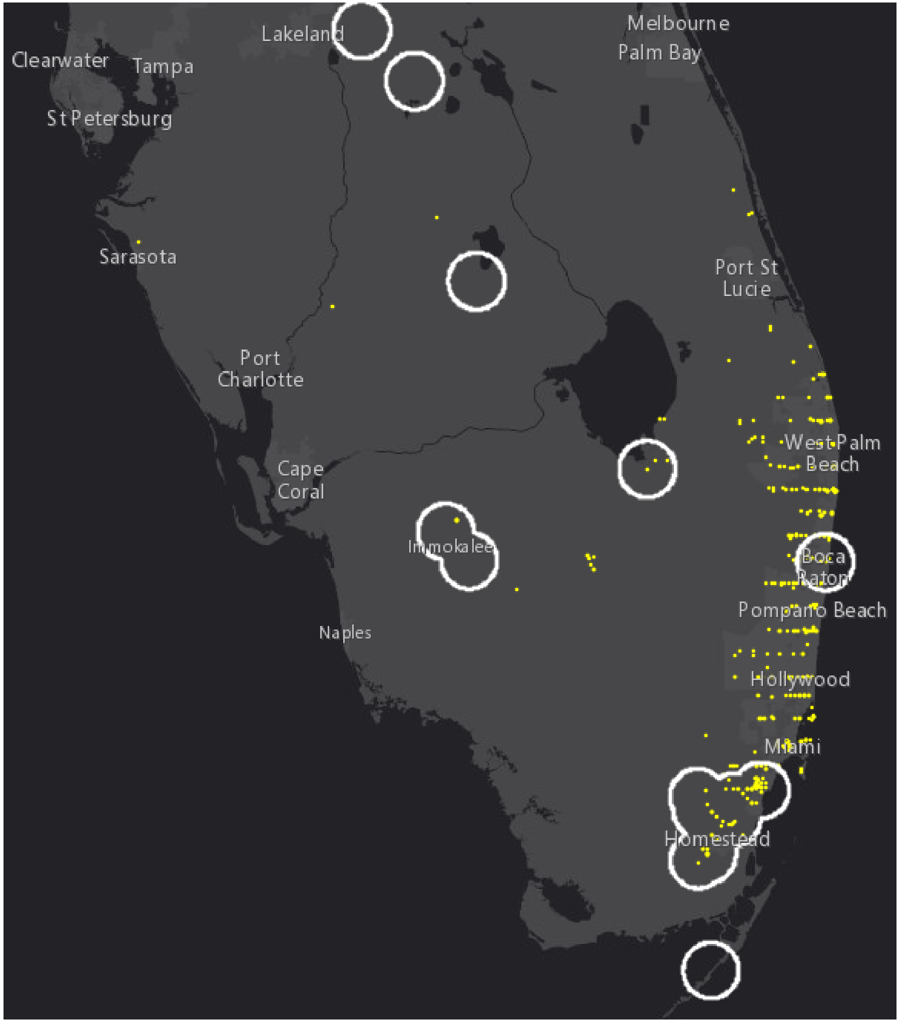


Northward Migration - DPI data set - Jan 1, '10

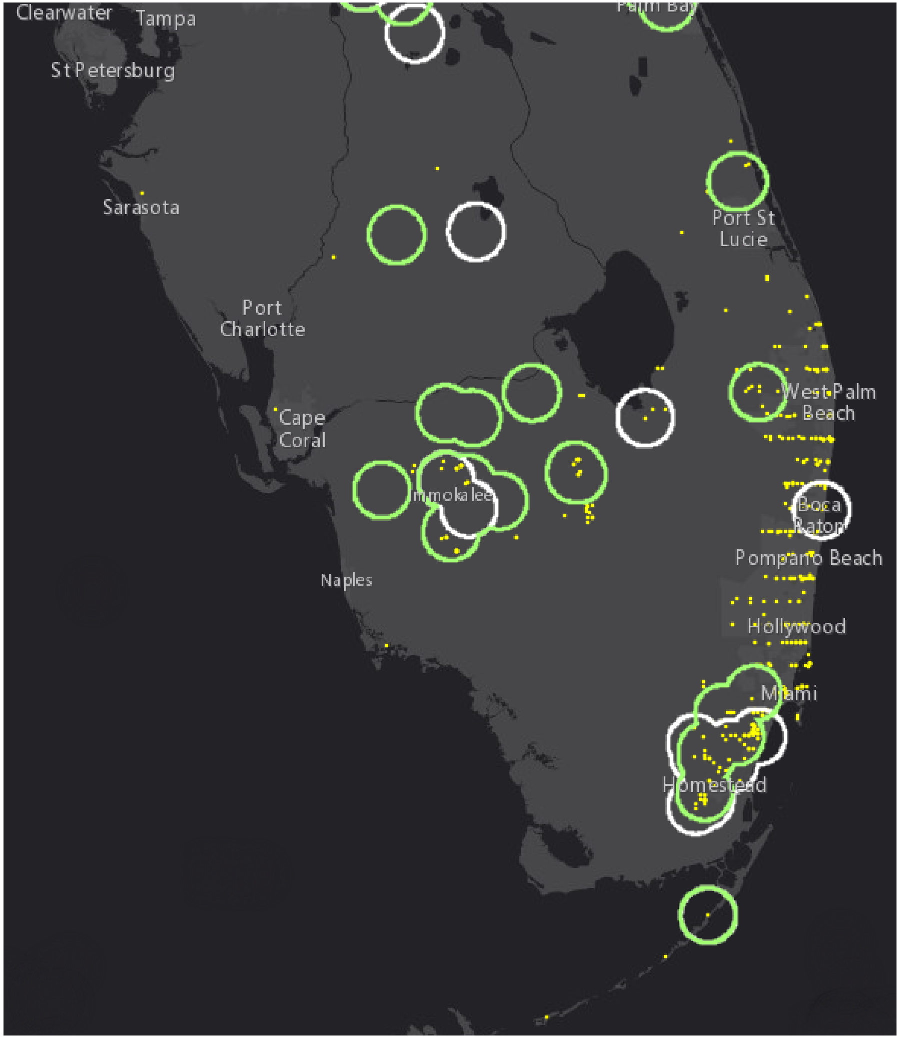


Northward Migration - Halbert Maps - Sep '05

Northward Migration - Halbert Maps - Dec '05



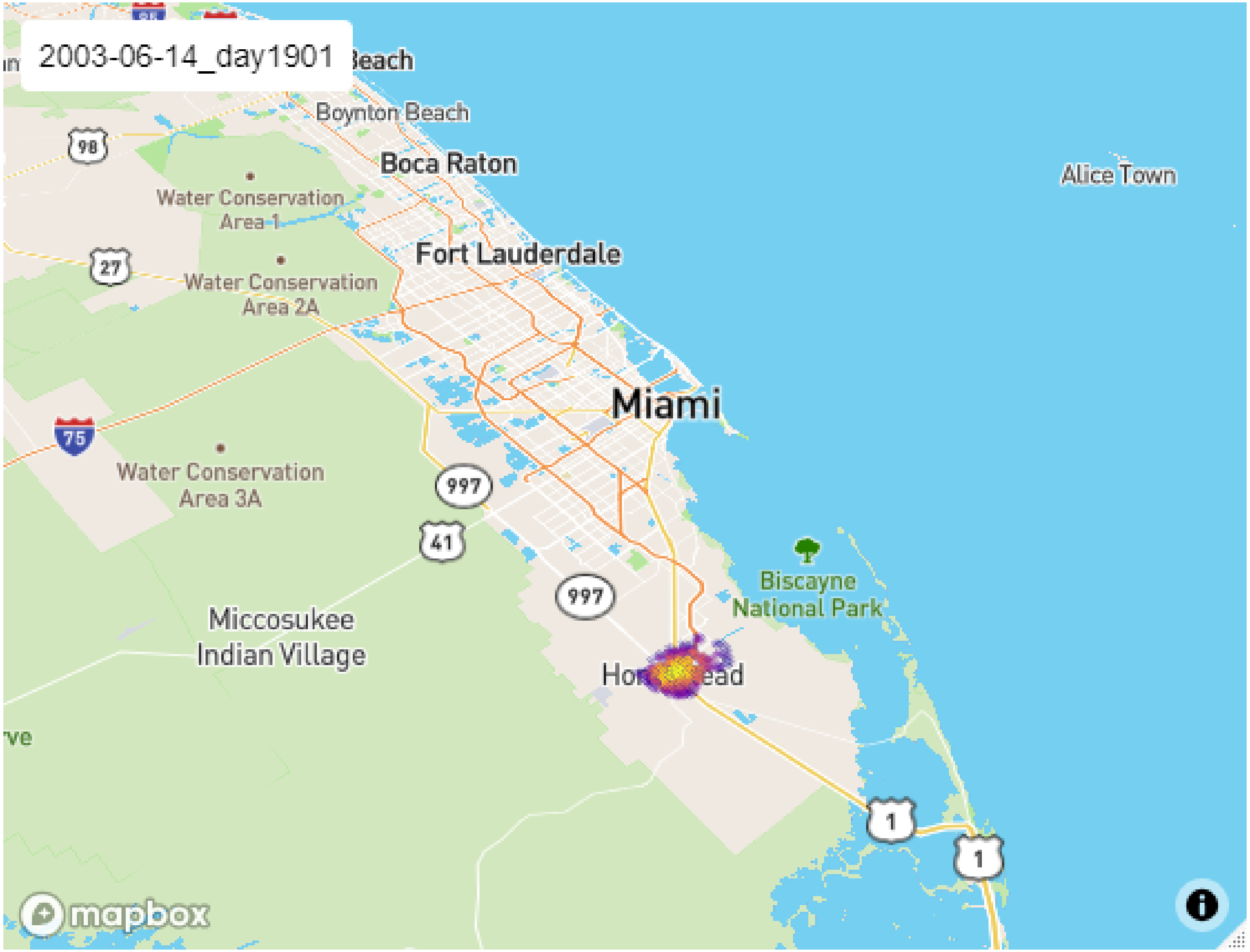
Northward Migration - Halbert Maps - Dec '06



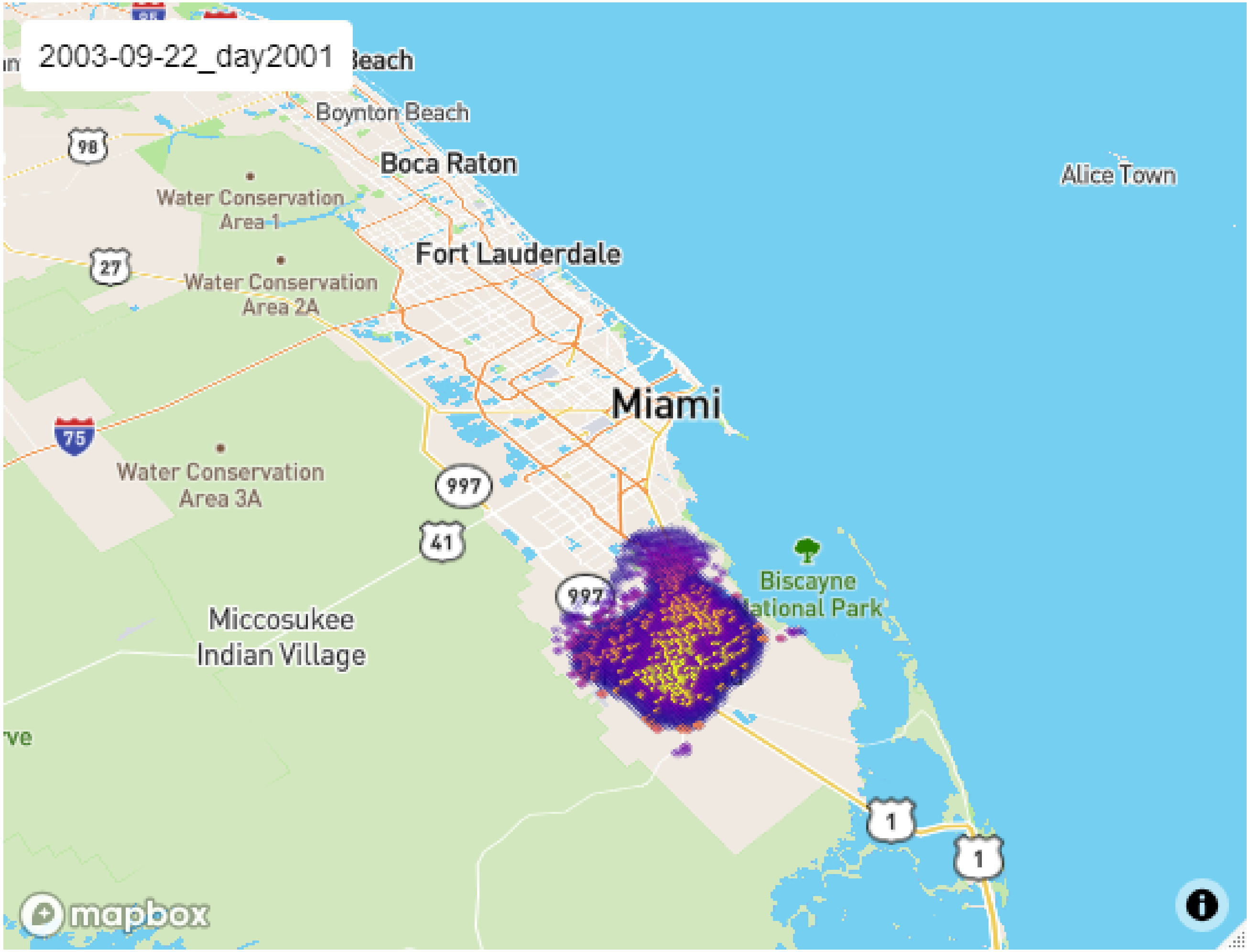
Summary of Features

- Overall extent of locations observed to have ACPs present increases throughout this period.
- Earliest samples of infected psyllids are observed in the Homestead area (October 2005).
- Density of infection increases in Homestead, and increasing numbers of samples are found farther north.
- Density of samples continues to increase along the corridor.
- Infection detected throughout the corridor by the end of 2005 (actual spread likely before as we already see infection in groves)

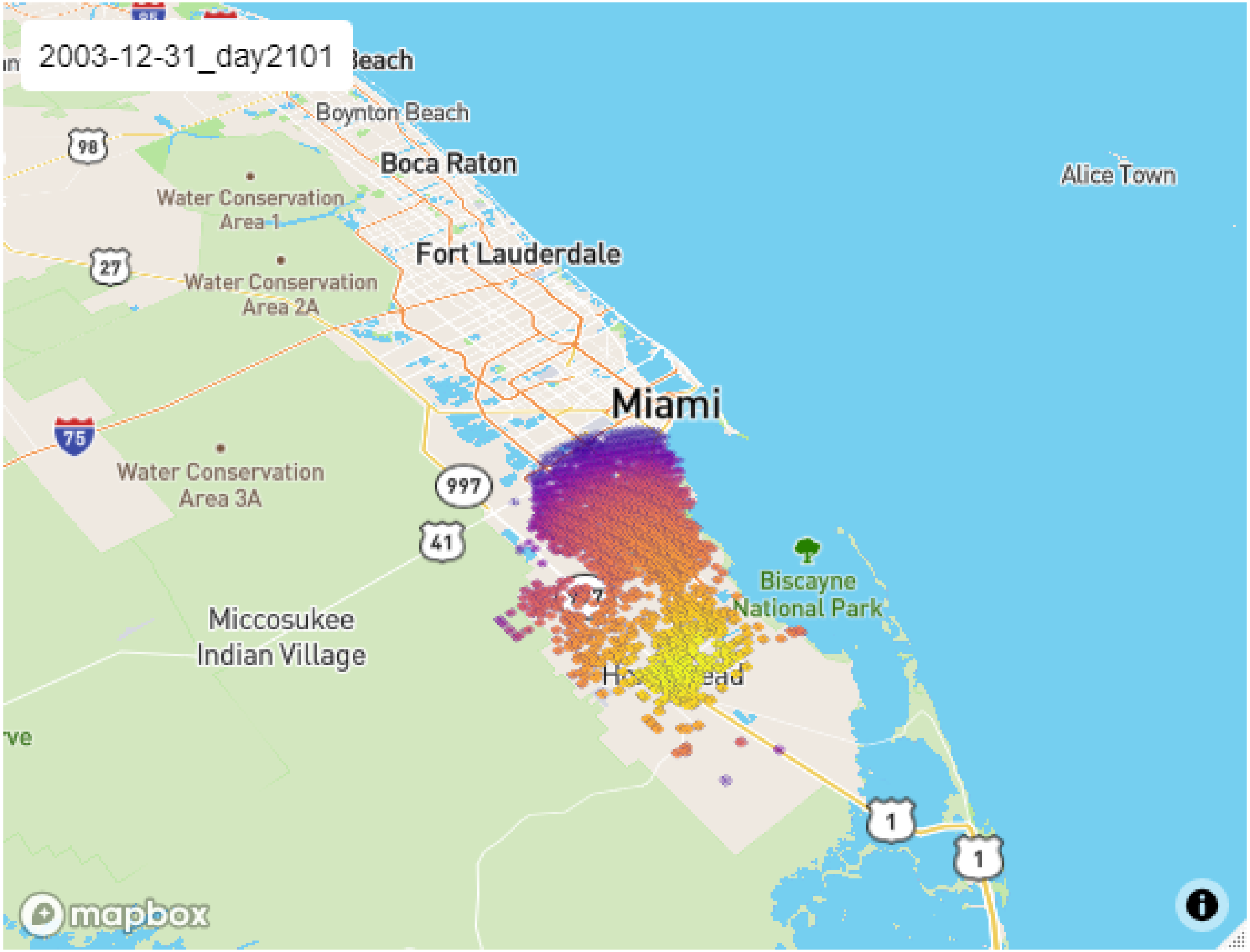
Northward Migration - Model output - Jun '03



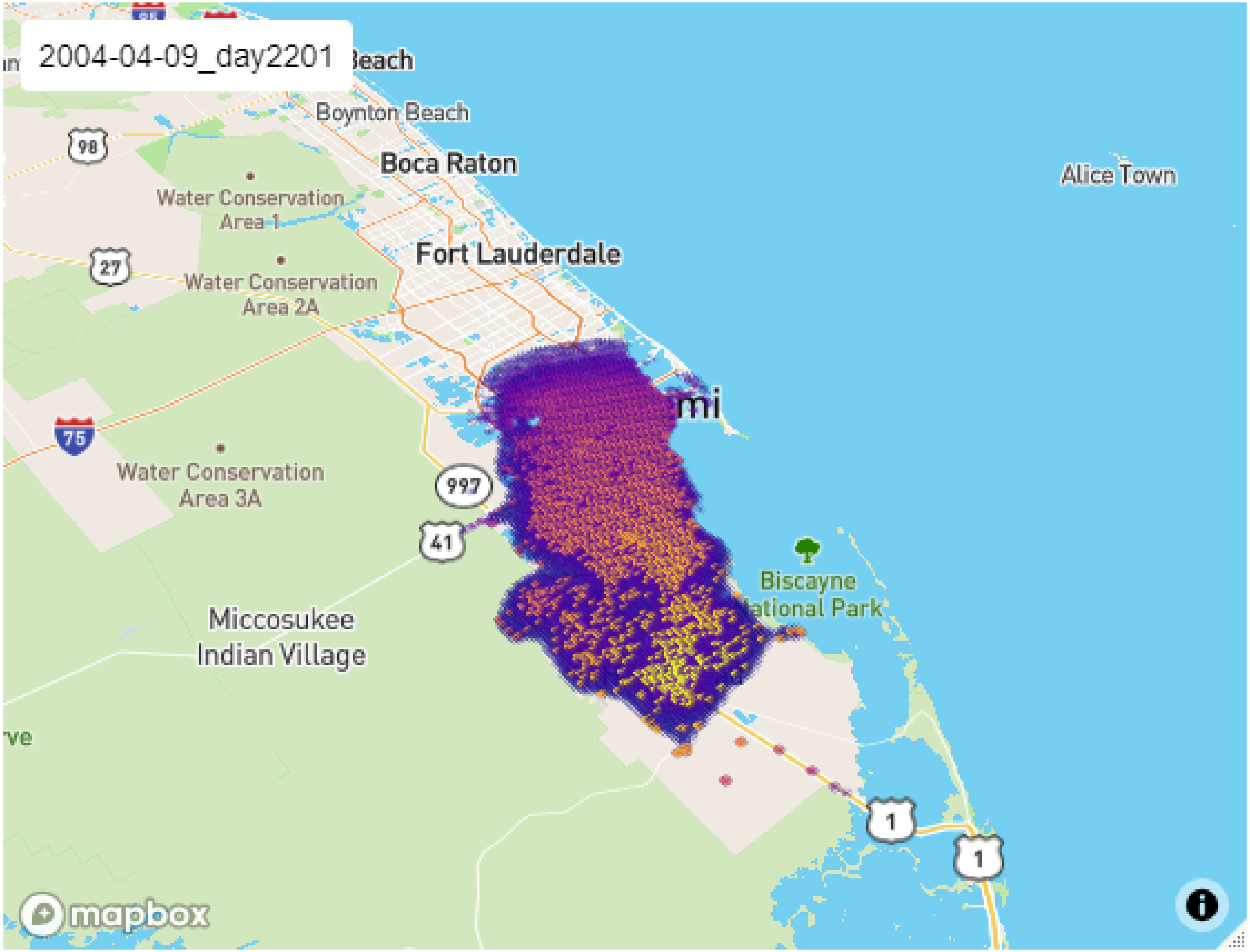
Northward Migration - Model output - Sep '03



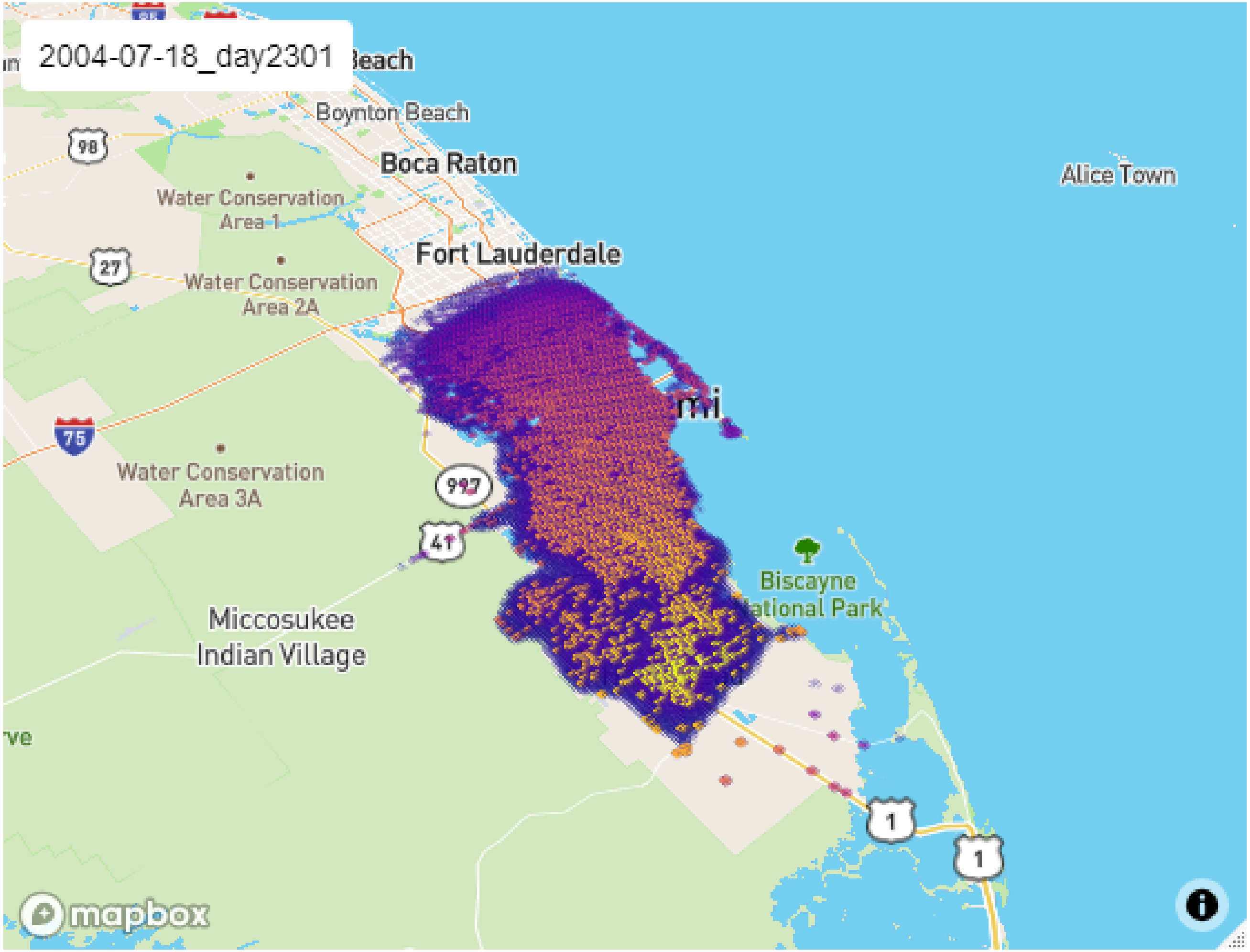
Northward Migration - Model output - Dec '03



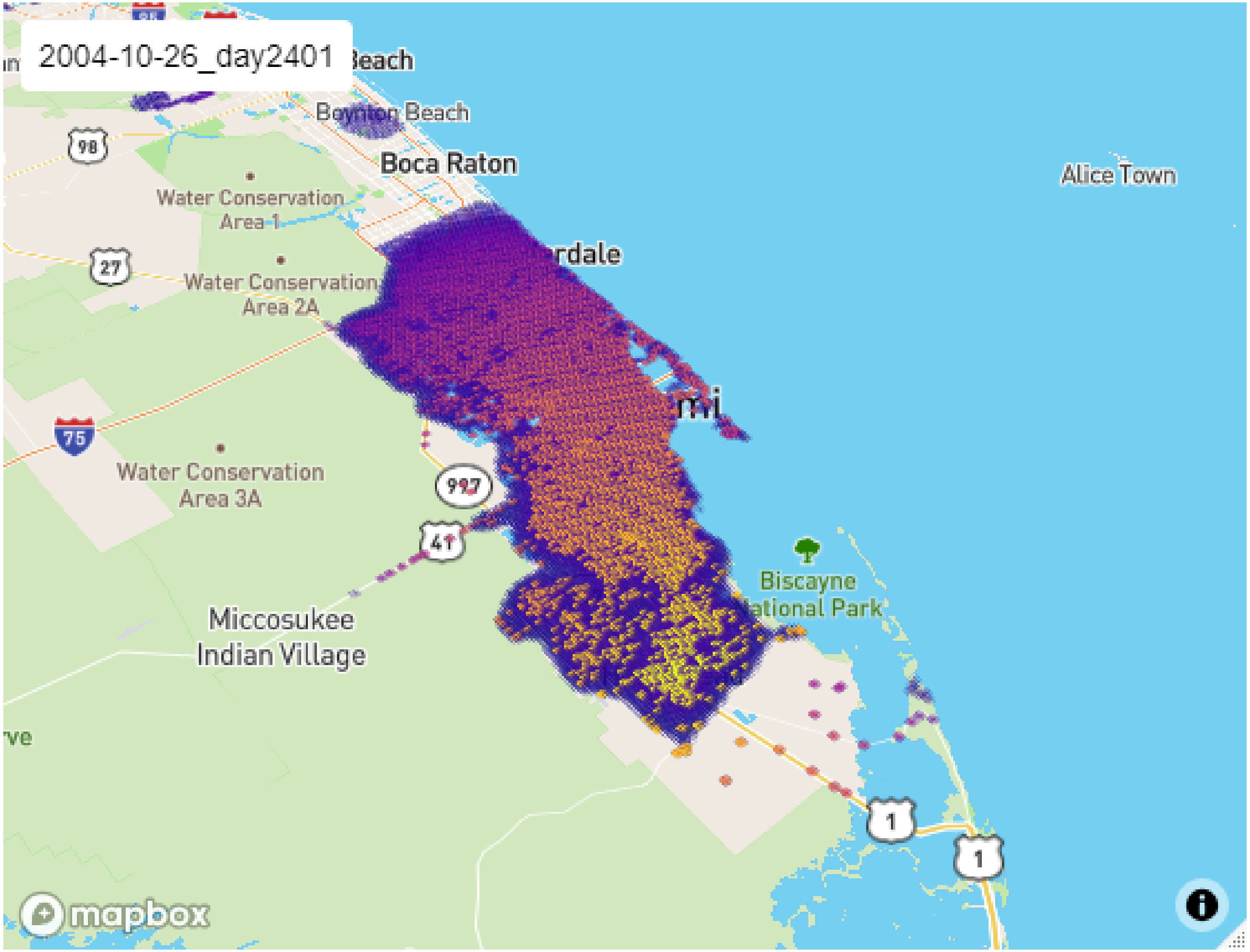
Northward Migration - Model output - Apr '04



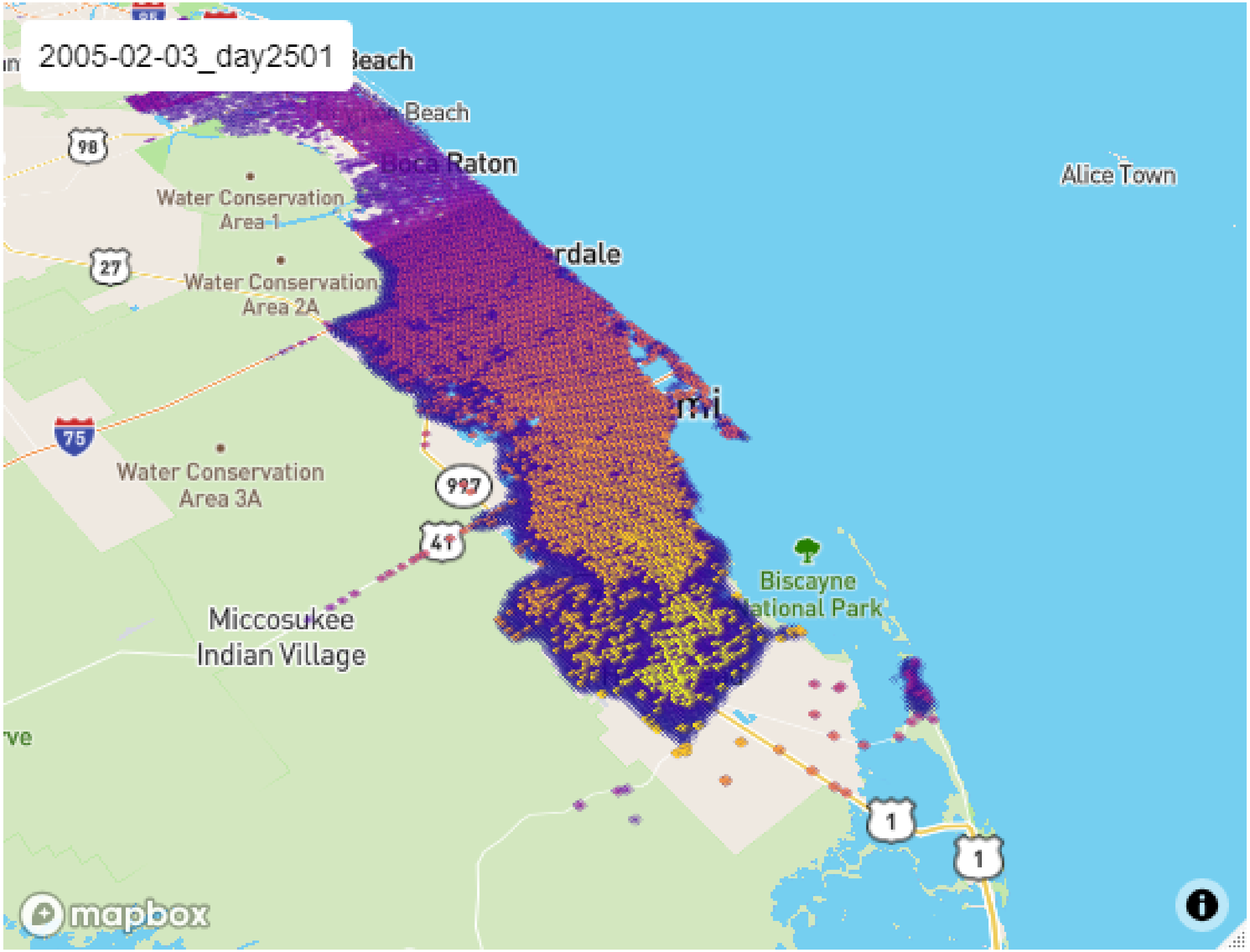
Northward Migration - Model output - Jul '04



Northward Migration - Model output - Oct '04



Northward Migration - Model output - Feb '05

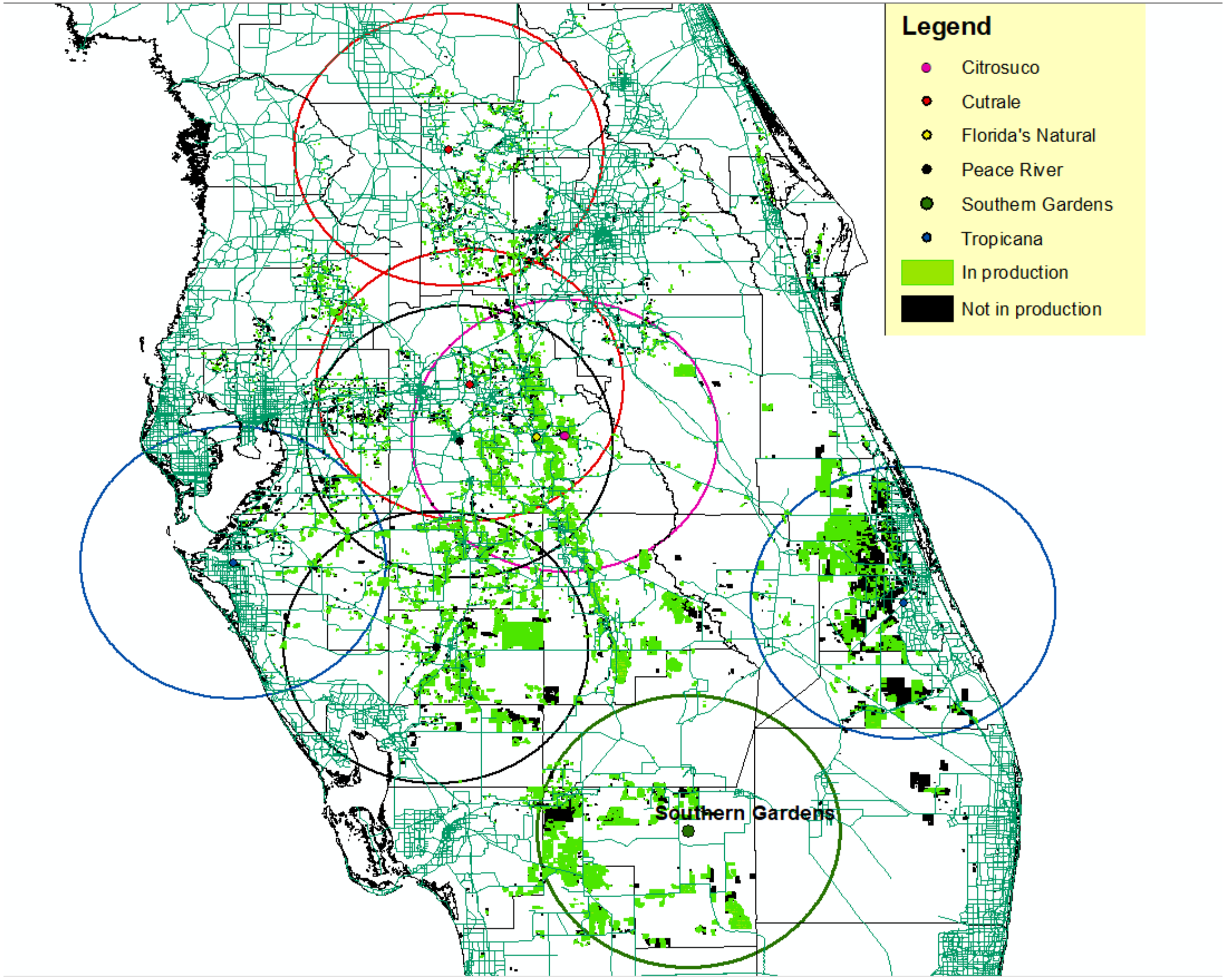


Model vs. Empirical Features

- Initial infection disperses from Homestead northward through the urban corridor
- Model generates infection throughout the corridor in early 2005
- Model's dispersal pattern is ahead of the data by several months, consistent with a "lagging detection" scenario
- Unclear when infection was acquired at Homestead
    - likely unknowable but model presents a plausible scenario

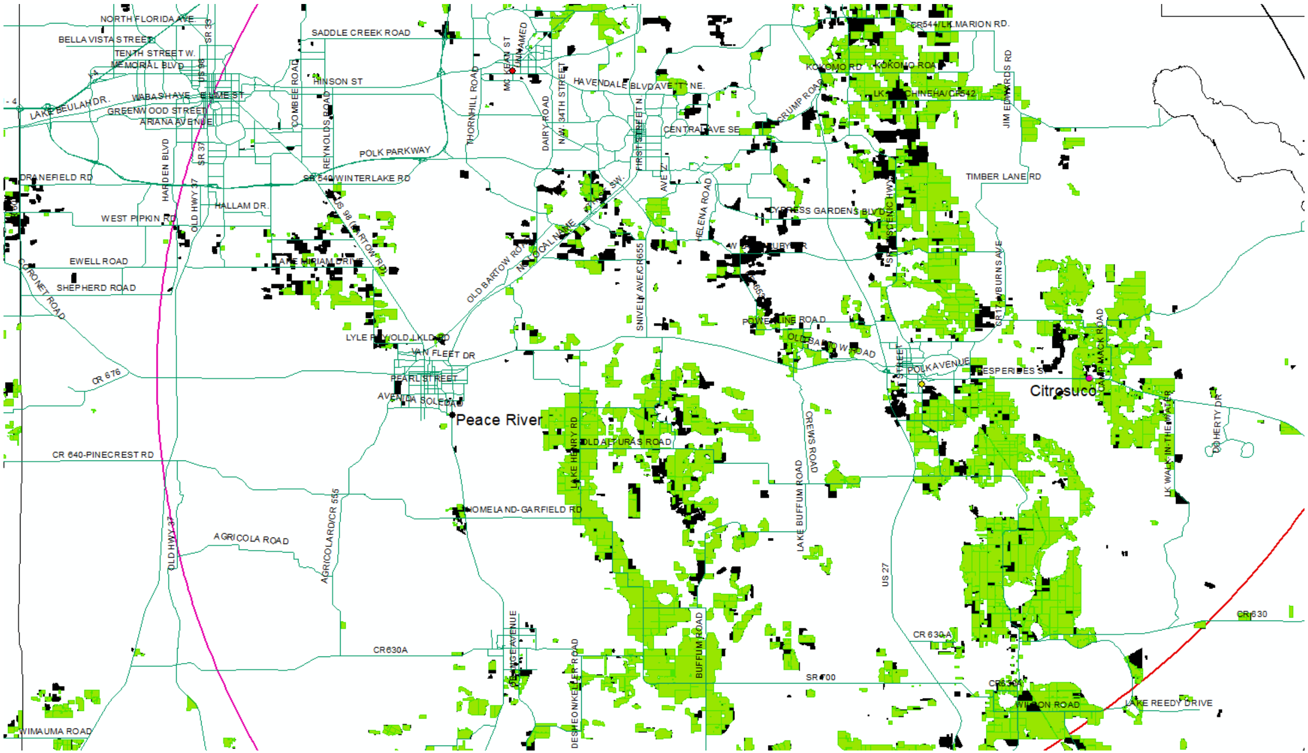
Commerical Groves - NASS Maps

Blocks of commercial citrus removed statewide, 2009:



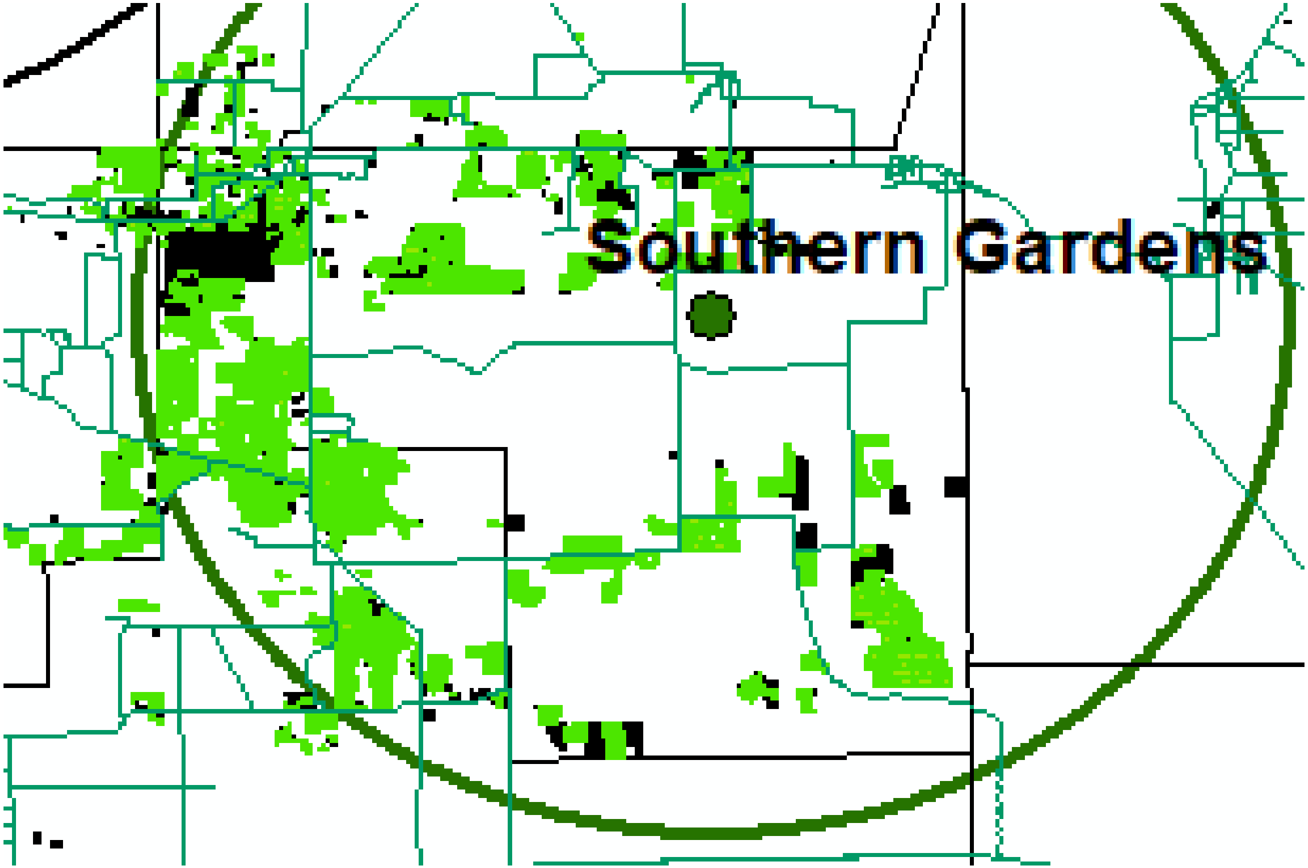
Commerical Groves - NASS Maps

Blocks of commercial citrus removed (centered east of Bartow, just south of Alturas), 2009:



Commerical Groves - NASS Maps

Blocks of commercial citrus removed near Southern Gardens, 2009:

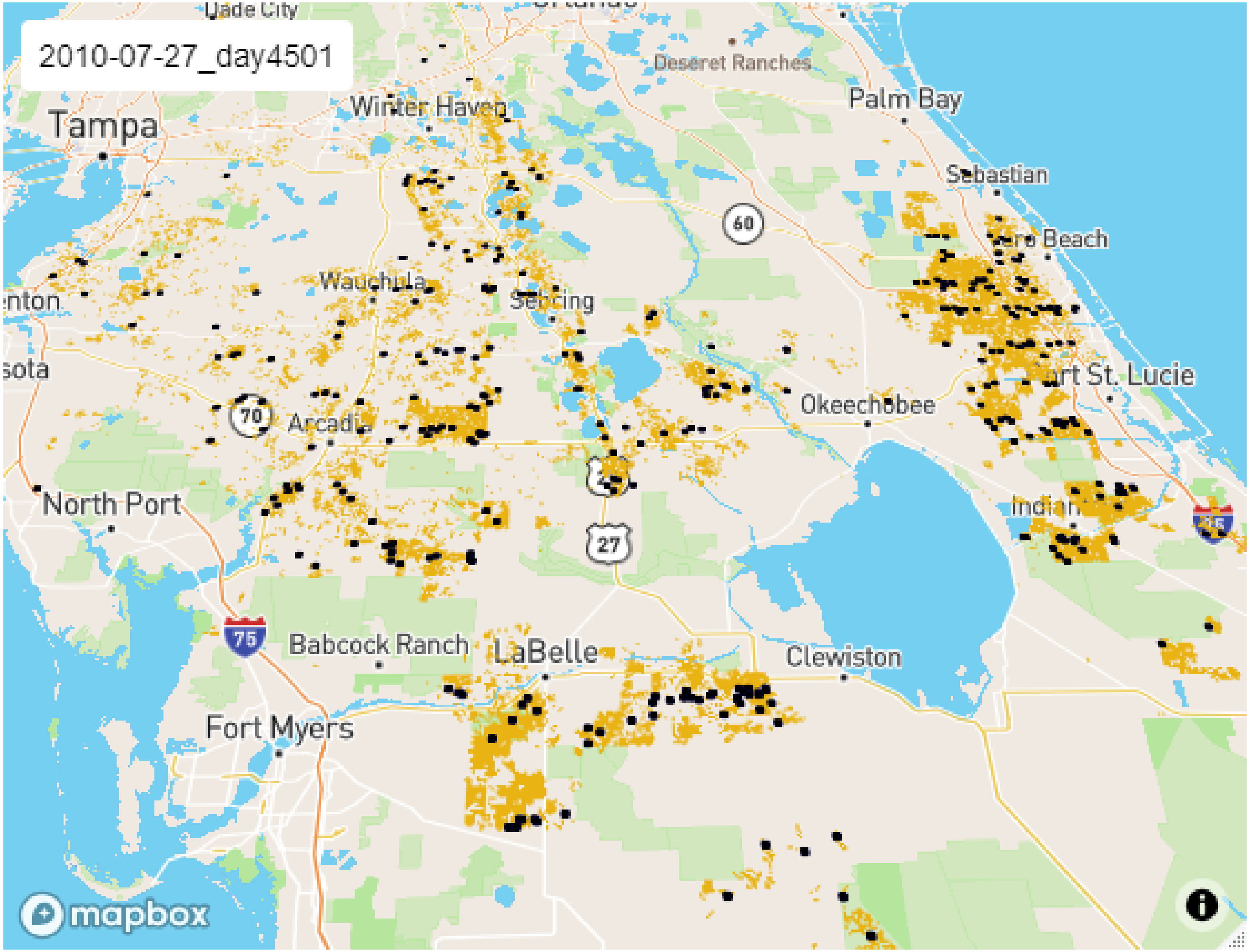


Summary of Features

- Broad spatial extent of removed citrus
    - essentially all of the citrus-growing region affected
- Heterogeneous removal pattern
    - some areas with heavy removal; others minimal
    - spatial correlation in removal (likely due to common ownership/managment)
- note: no time sequence of removal

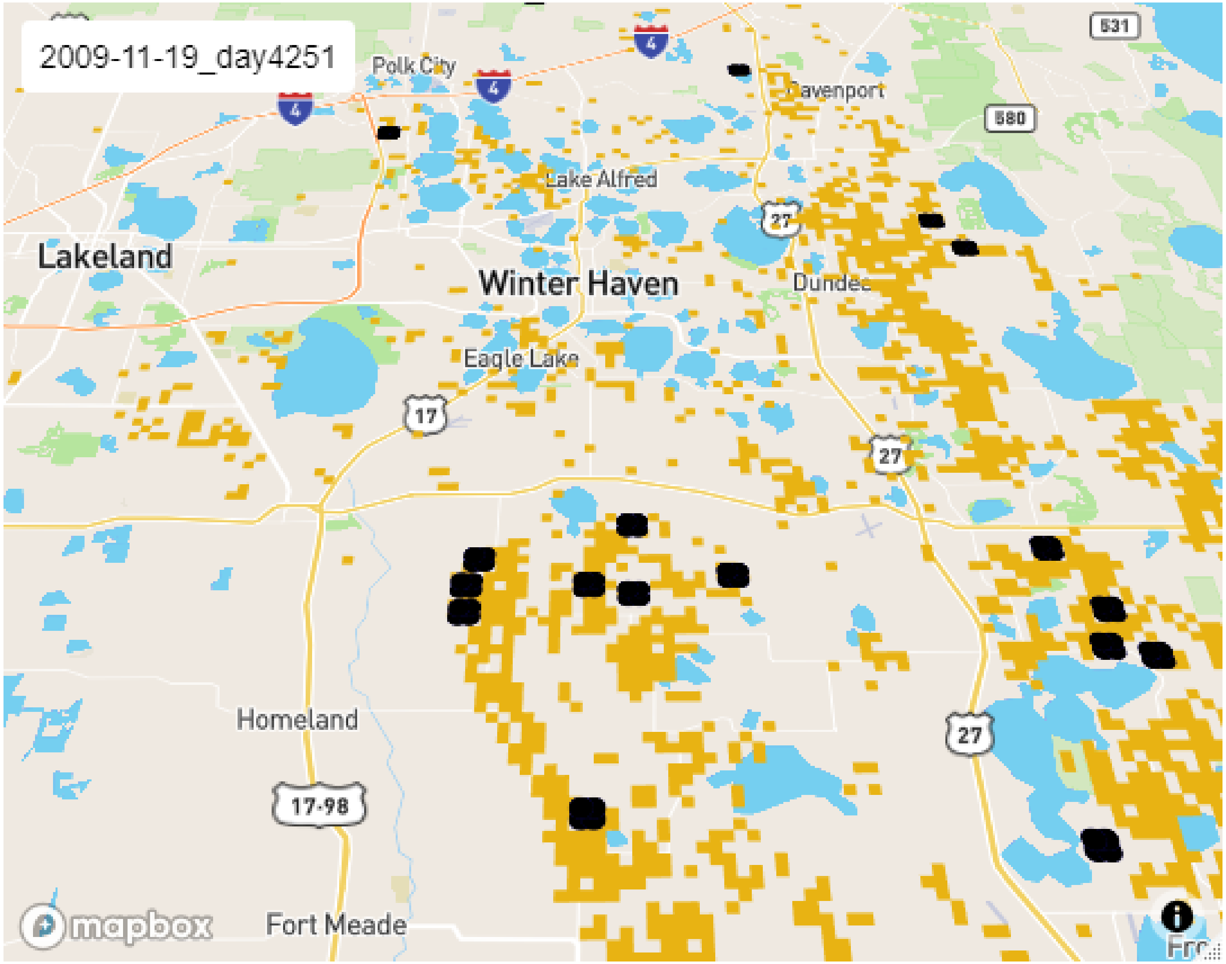
Commerical Groves - Model output

Statewide removal of citrus:



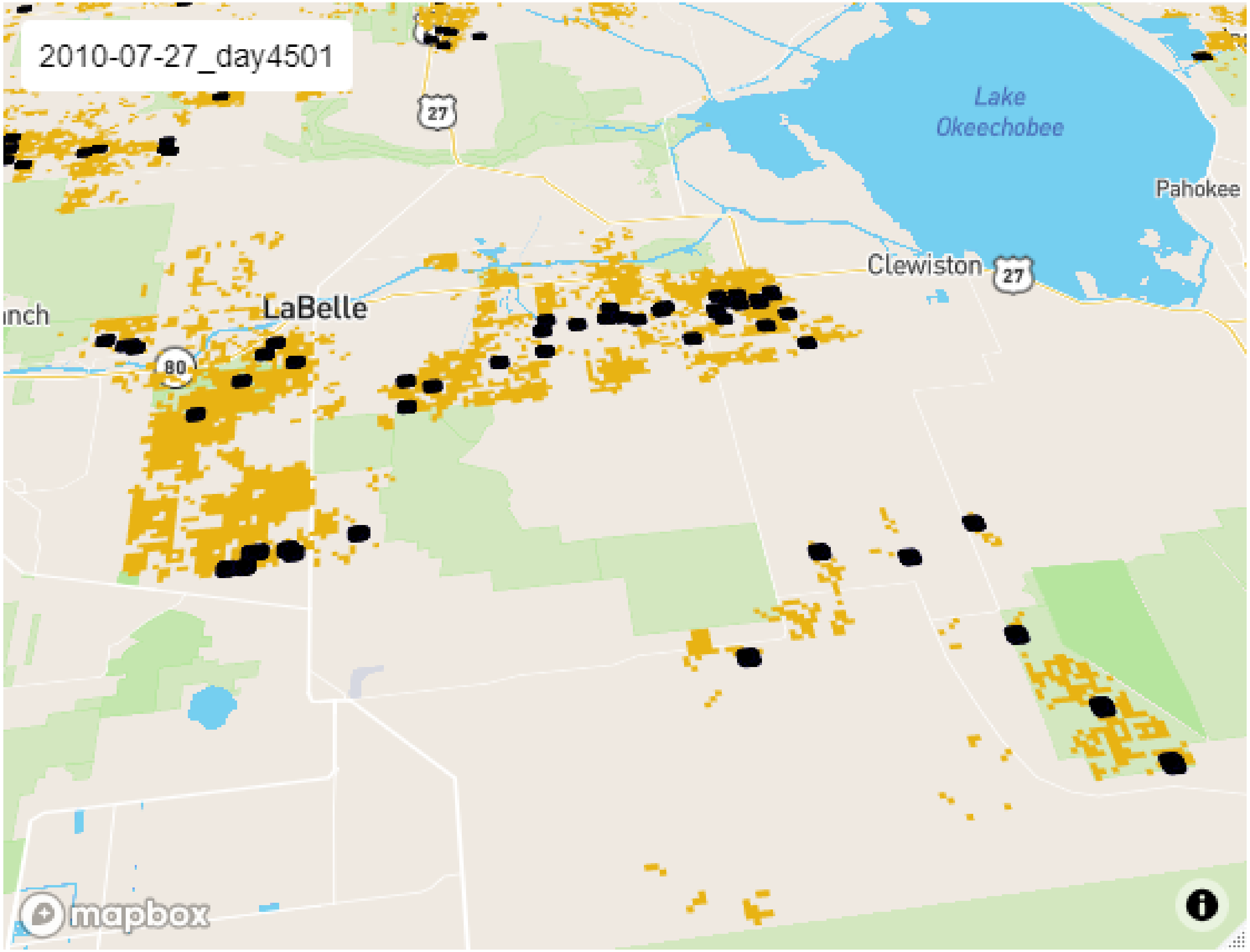
Commerical Groves - Model output

Removal of citrus near Bartow, FL:



Commerical Groves - Model output

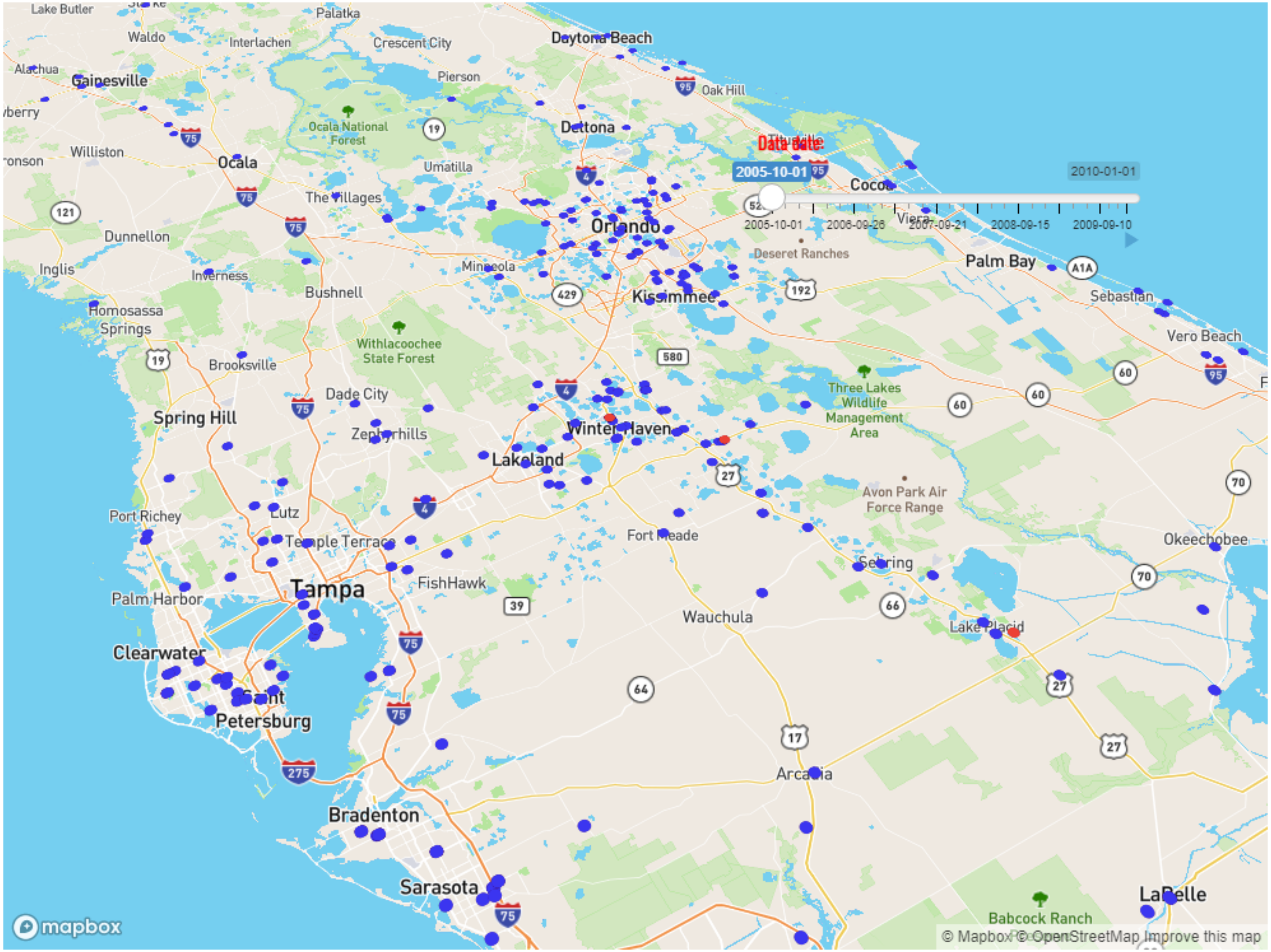
Removal of citrus, Southern Gardens:



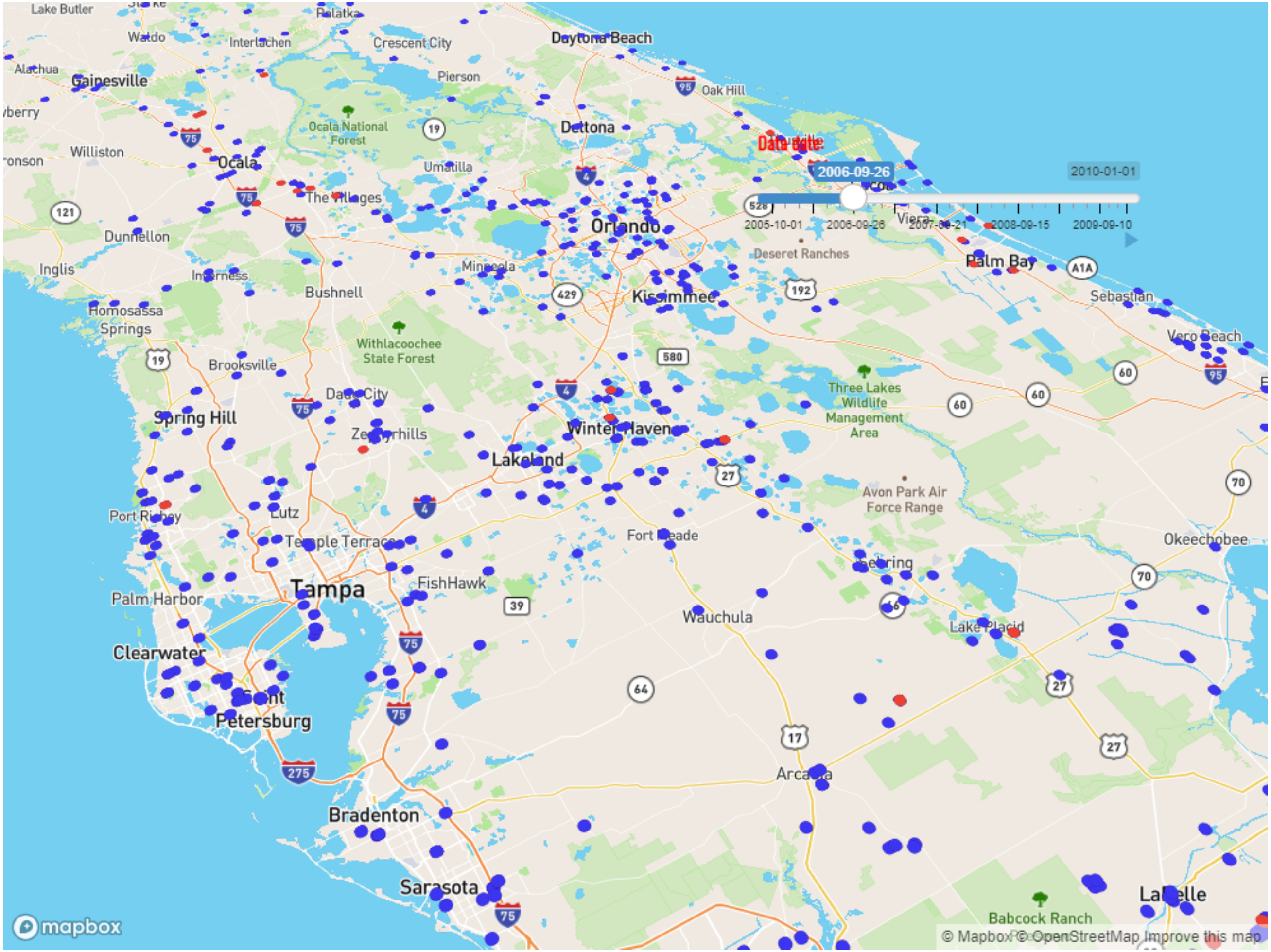
Model vs. Empirical Features

- Model generates more uniform dispersal
  - primarily due to lack of spatial correlation
- Local removal patterns are not an exact match, though still plausible
- Greater (if justified) specificity in modeling assumptions could generate more realistic patterns

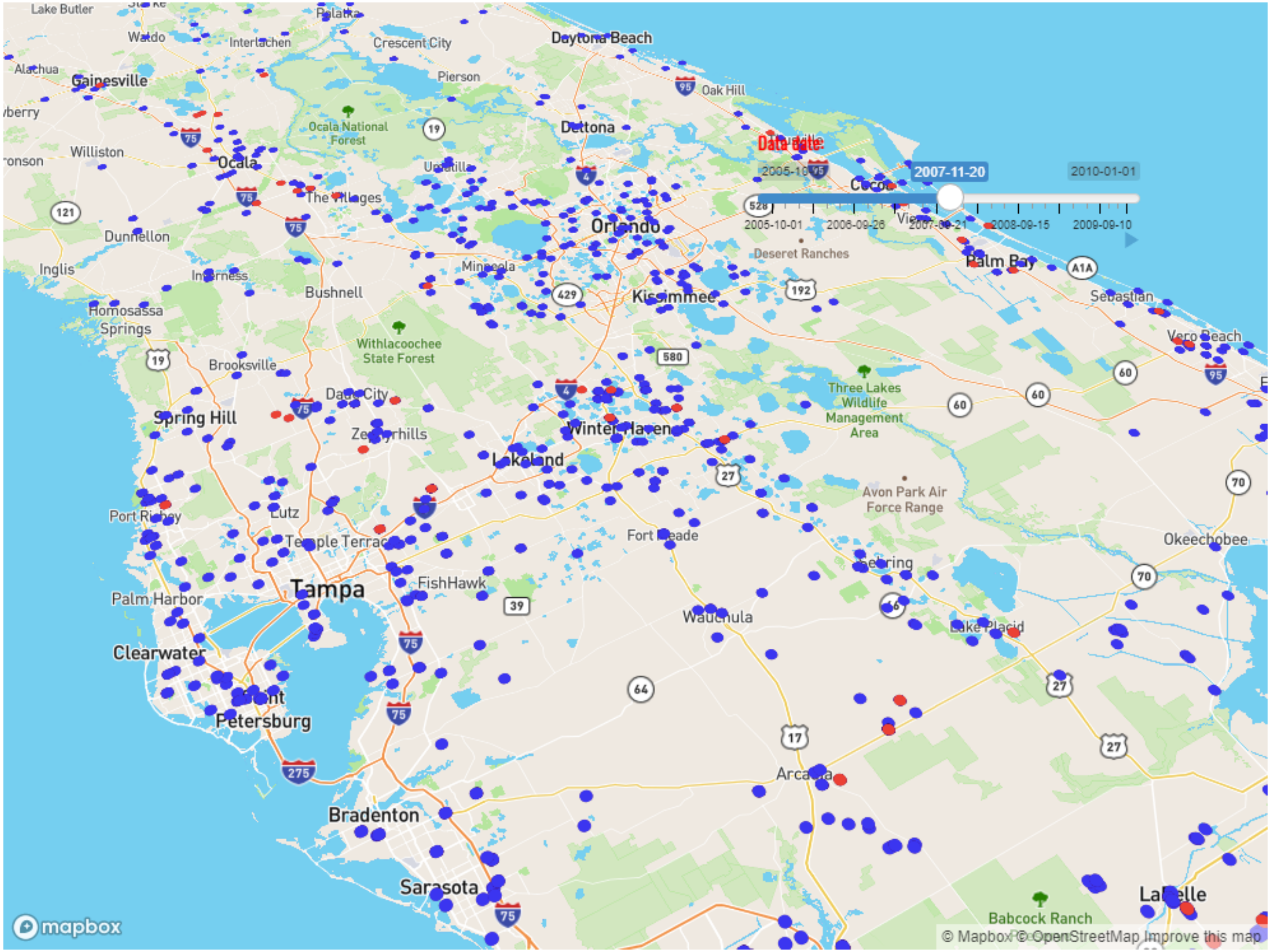
North/Central FL - DPI data set - Oct 1, '05



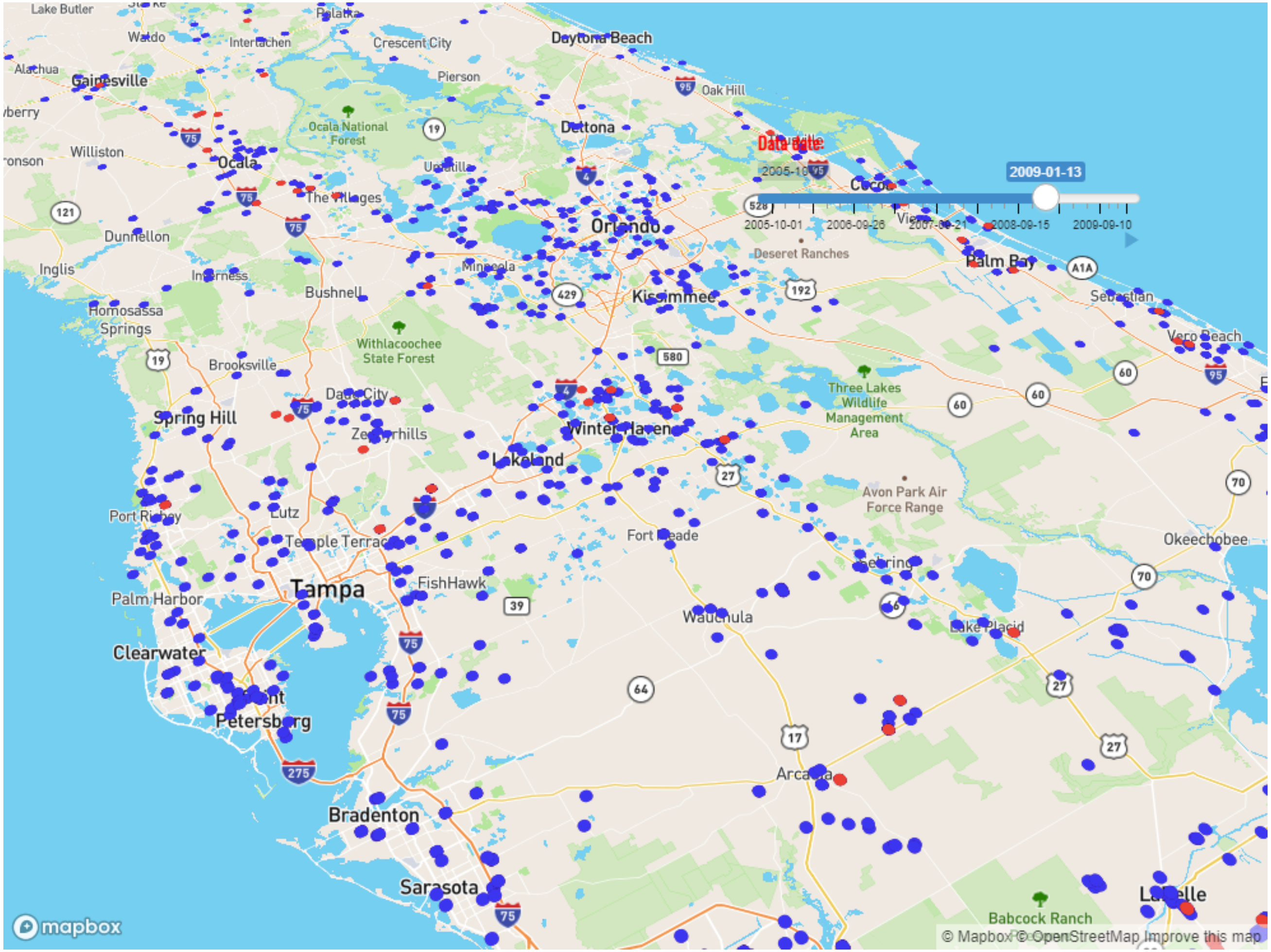
North/Central FL - DPI data set - Sep 26, '06



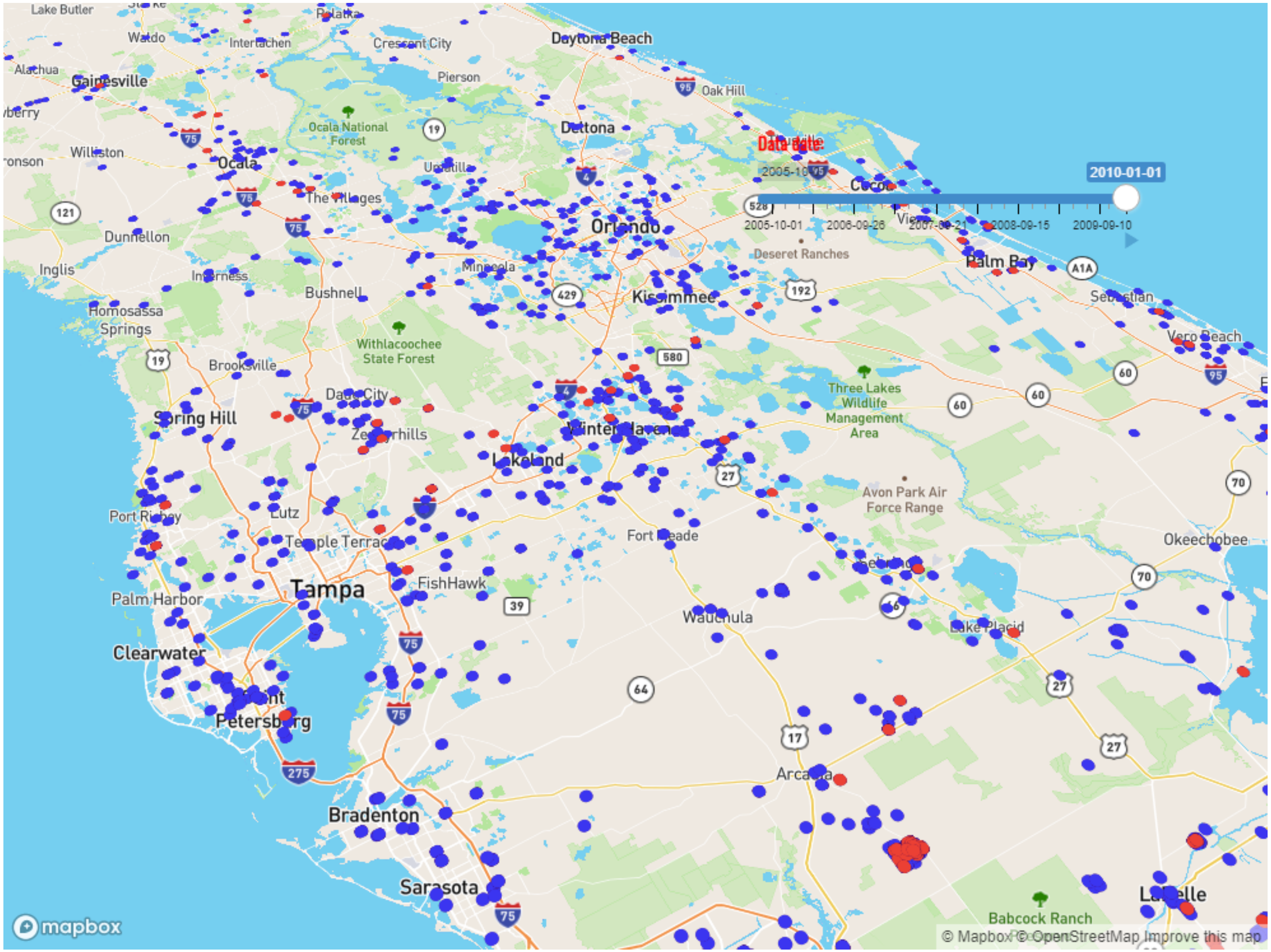
North/Central FL - DPI data set - Nov 20, '07



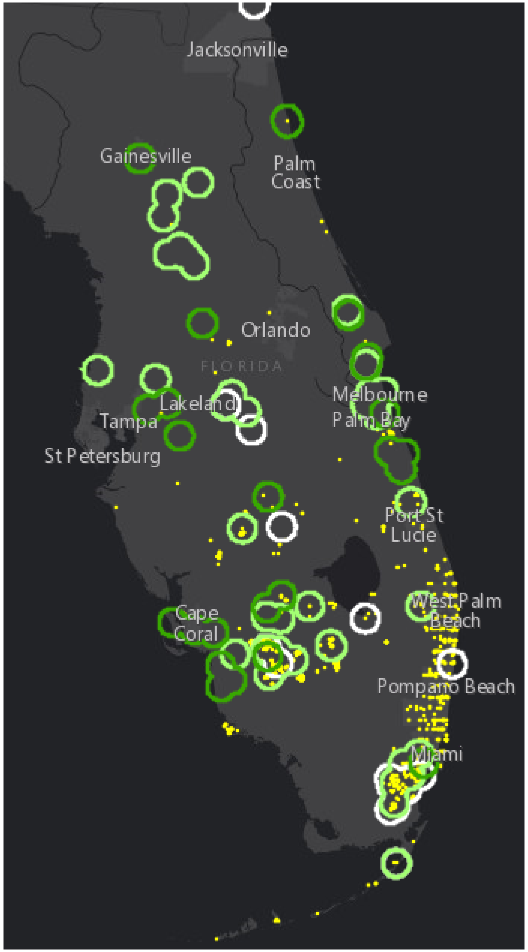
North/Central FL - DPI data set - Jan 13, '09



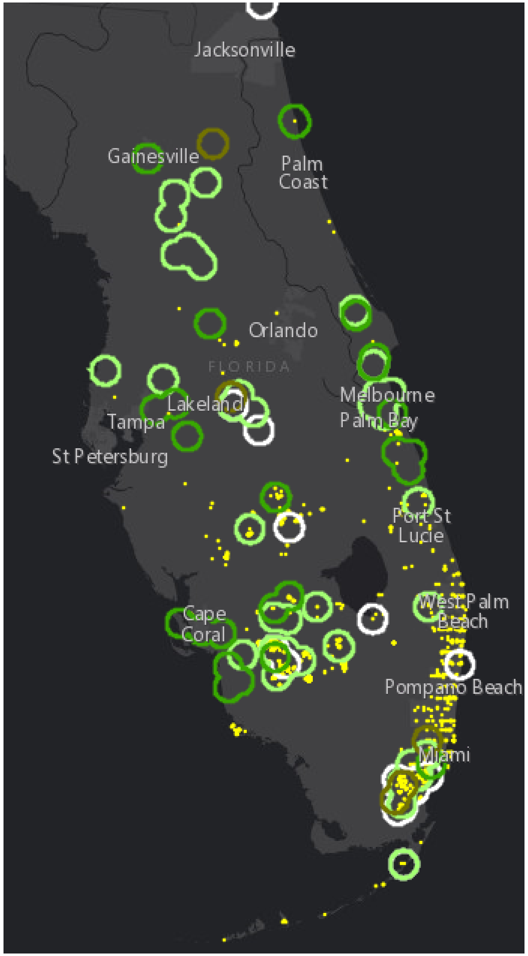
North/Central FL - DPI data set - Jan 1, '10



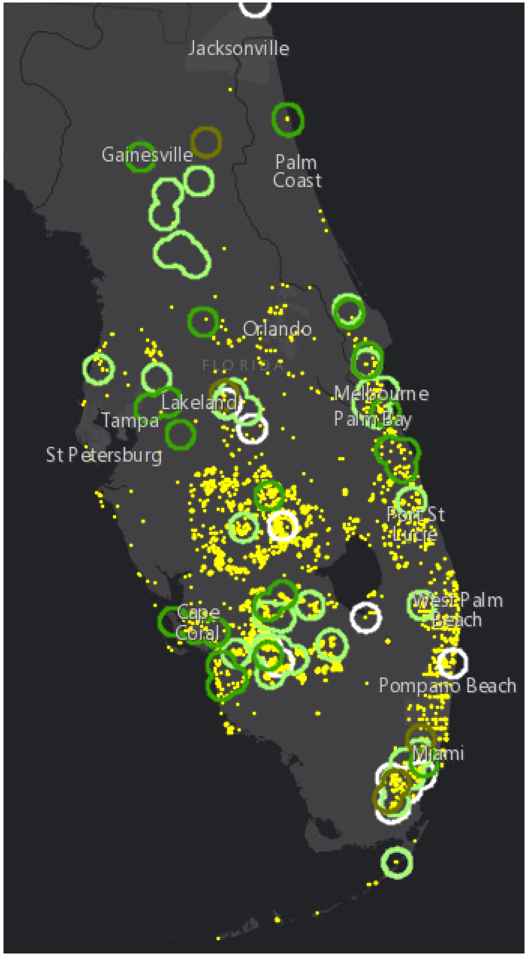
North/Central FL - Halbert Maps - 2007



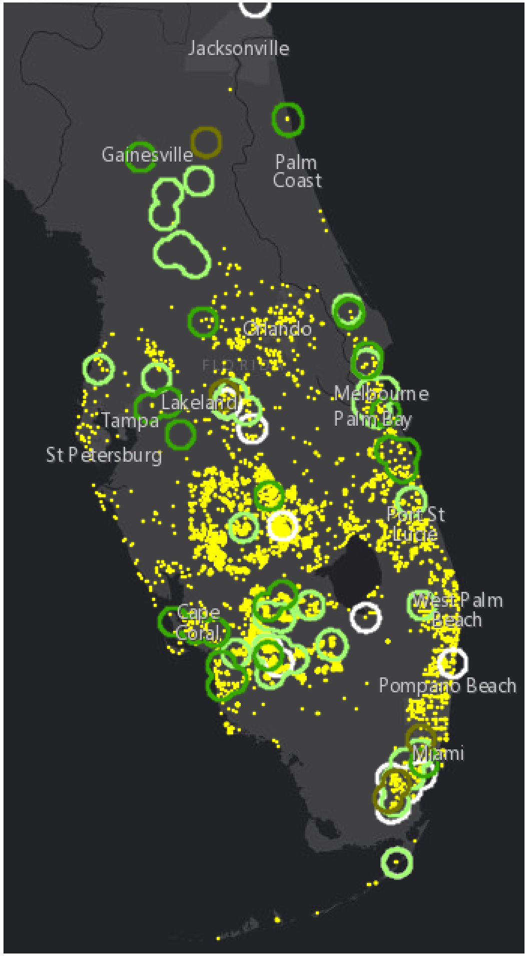
North/Central FL - Halbert Maps - 2008



North/Central FL - Halbert Maps - 2009



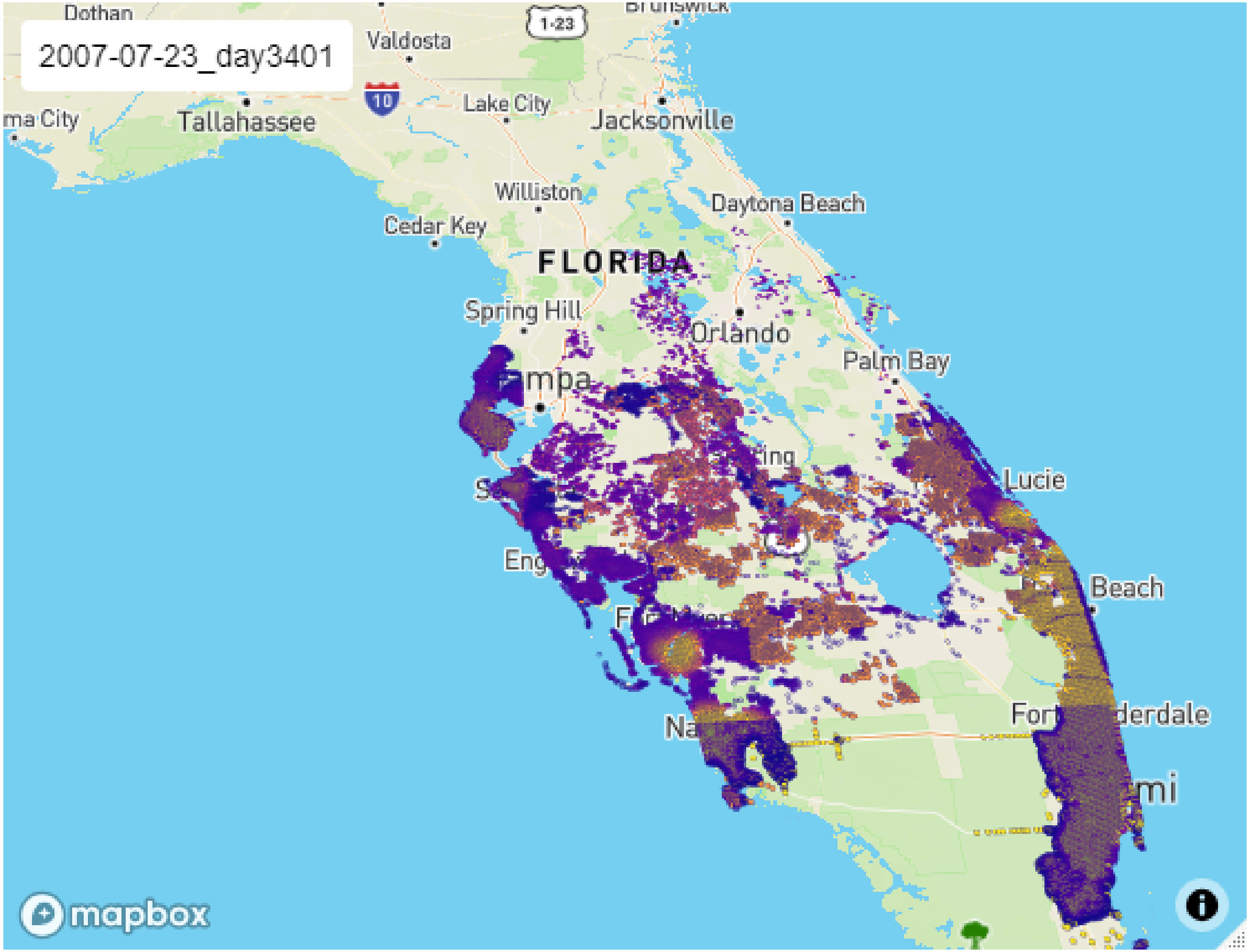
North/Central FL - Halbert Maps - 2010



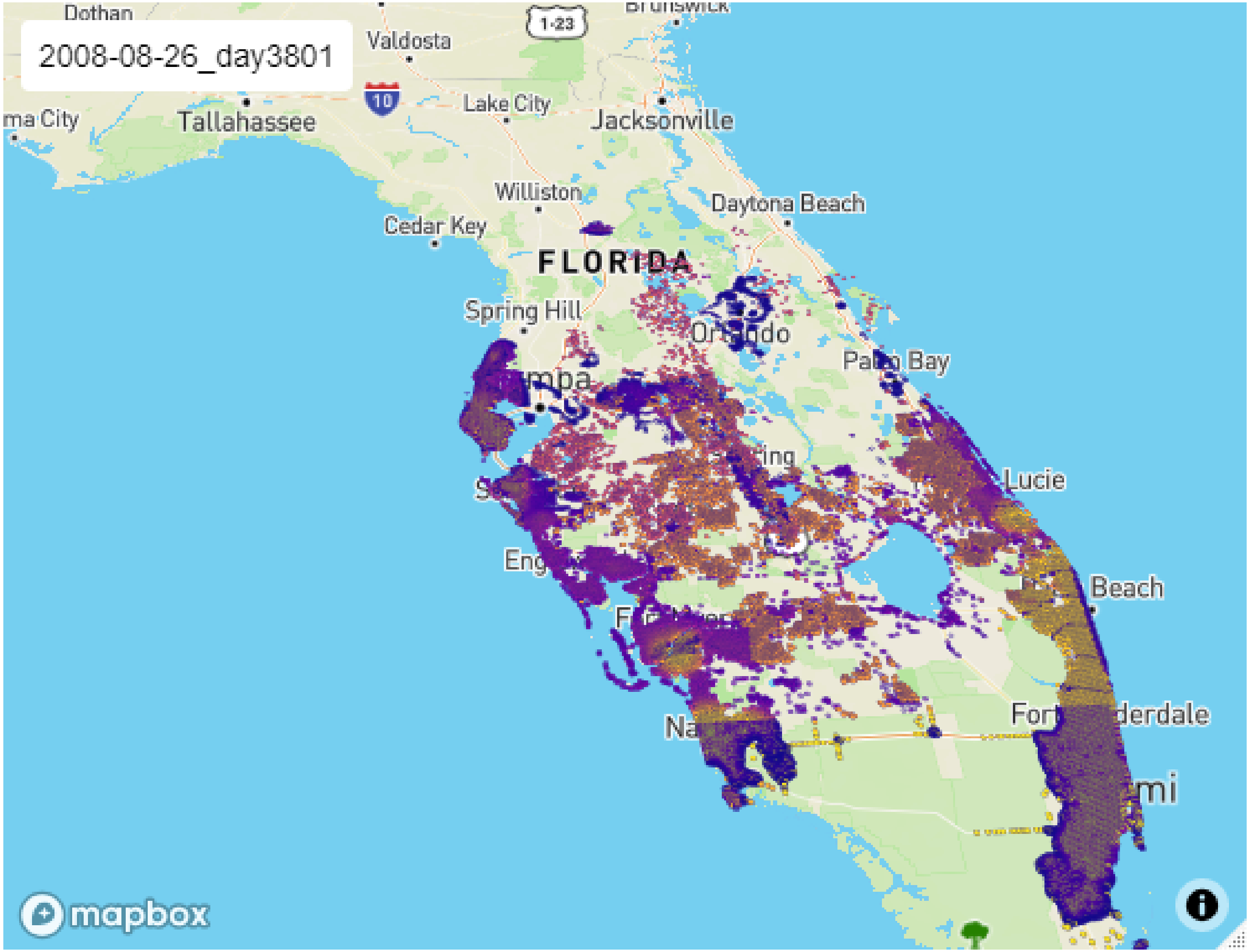
Summary of Features

- Infection density increasing steadily through the period 2007-10
- Increasing range of infection into northern Florida
- Isolated instances of infection in northern locations attributable to plant sales

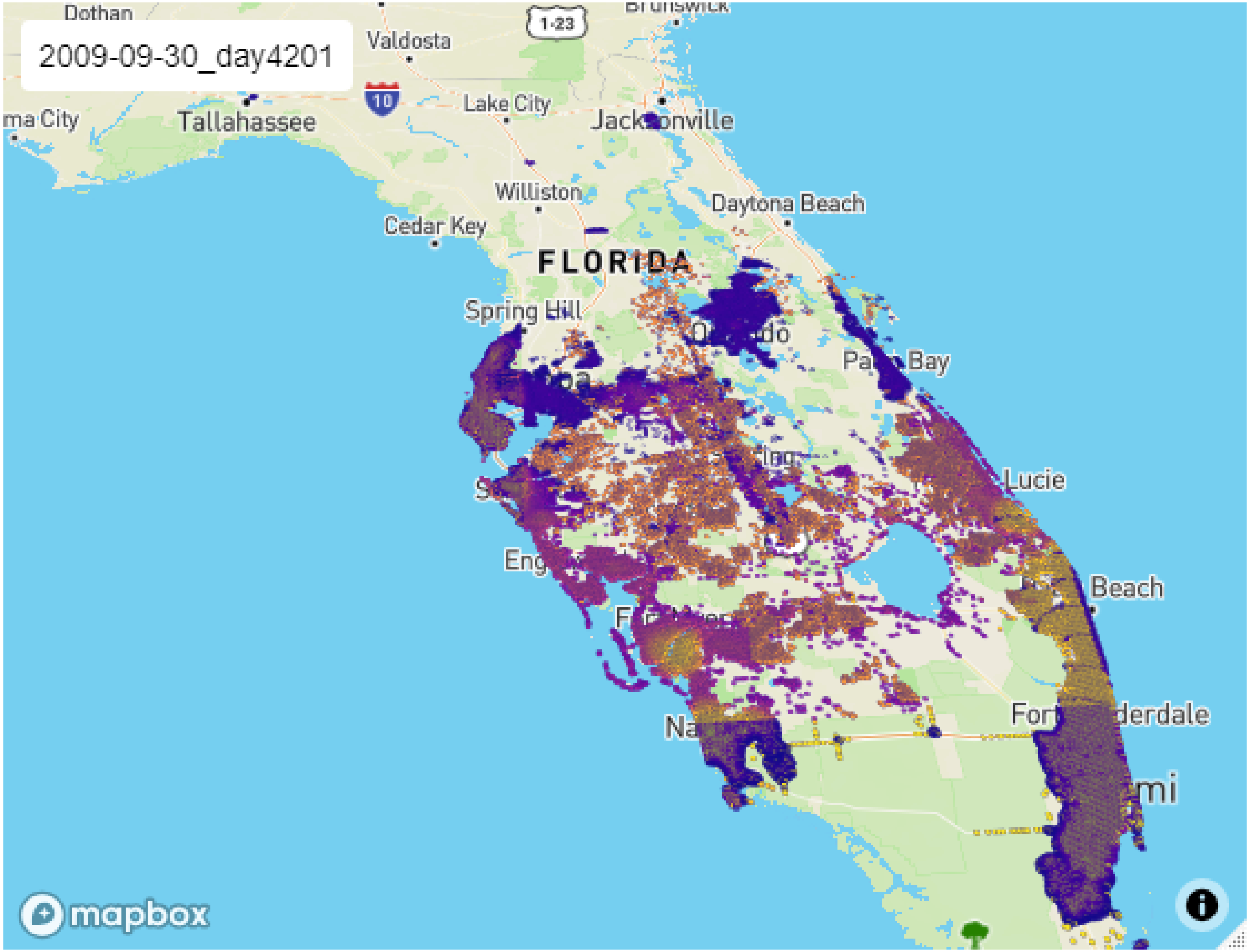
North/Central FL - Model output - 2007



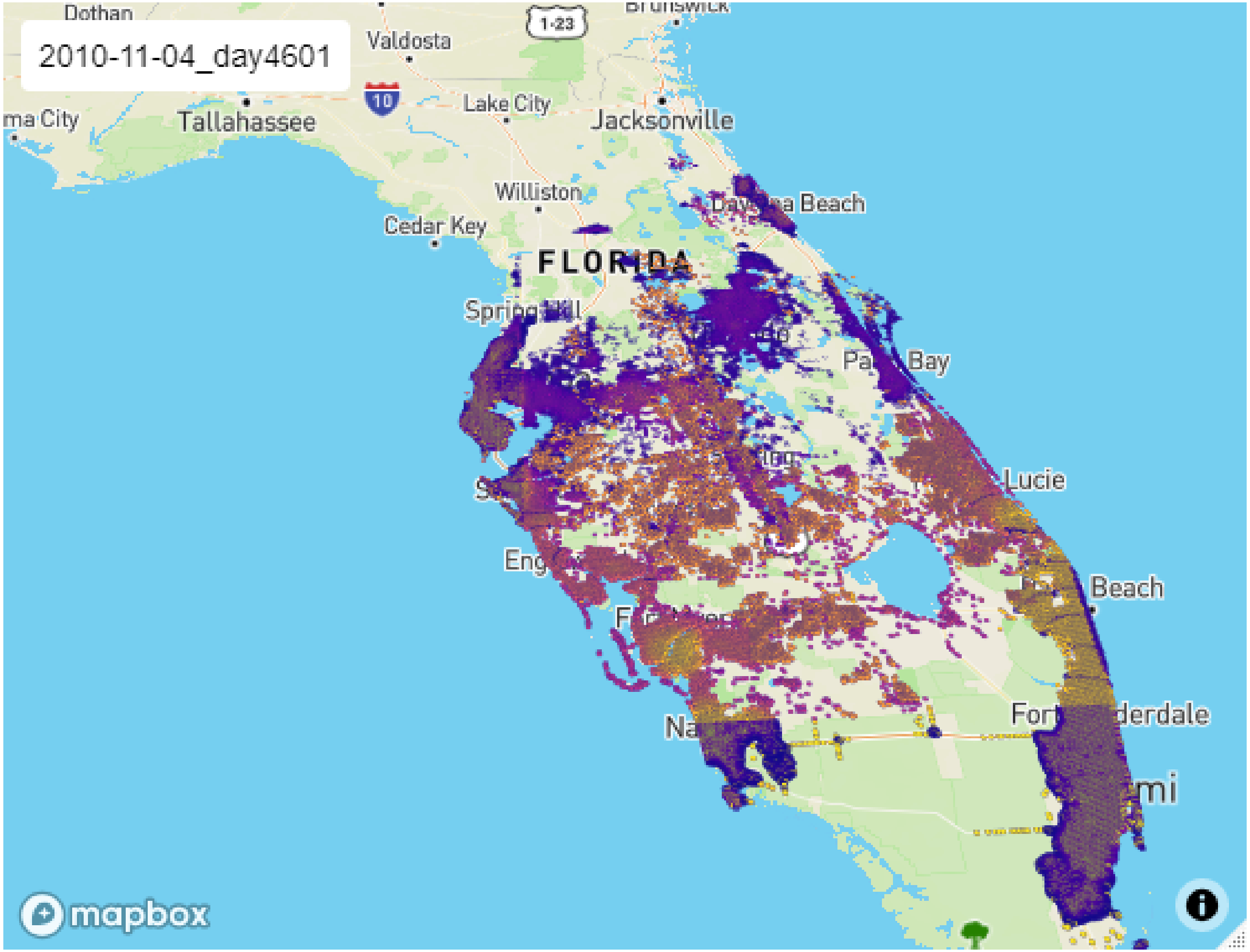
North/Central FL - Model output - 2008



North/Central FL - Model output - 2009



North/Central FL - Model output - 2010



Model vs. Empirical Features

- Model depicts increasing infection density in central Florida, and novel instances of infection in northern Florida (plant sales)
  - model has a bit less infection in north Florida
  - essentially all commercial groves are infected
- Model infection density is higher/earlier than data
  - consistent with lagging detection/undersampling
- Overall pattern of infection generated by the model is quite plausible/realistic

Assessments Summary

- Captures a reasonable, smoothed version of the initial invasion;
- Produces a plausible timeline for the acquisition of *C*Las and subsequent northward migration of infection;
- Generates grove-removal patterns that while not a perfect fit to historical maps, are reasonably close given the necessary lack of modeling specificity;
- Generates realistic infection patterns statewide through the completion of the model (in particular, in later years as infection spreads through north/central Florida).

Quantitative Assessements

Data limitations

Working with the DPI data set, we have:

- Limited number of observations
- No systematic collection
- Exact population levels unknowable
- Presence-only data

Model Expectations

- Should accurately detect presence of uninfected/infected ACP over time
  - for testing, view model as a binary classifier
- Model serves as an interpolation/predication for unmeasured locations
  - no penalty for extra observations

Error Measure

Given:

- loss function L
- realization of random field M under the model at points $(x_1, y_1), \ldots, (x_n, y_n)$ with observed data $z_1, \ldots, z_n$

Natural measure of model performance is the empirical risk:

$$\hat{R}(M) = \frac{1}{n} \sum_{i=1}^{n} L(M(x_i, y_i), z_i)$$

Loss Function

Given data limitations, we take L to be $0/1$ loss:

$$L(a, b) = I(a \neq b)$$

where I is the standard indicator function.

Full model risk

Risk across all years, separated by spatial region and infection status:

Region	i	$\hat{R}(M_i)$	n
Urban	1	0.01408451	420
Citrus	1	0.00000000	132
Urban	2	0.02857143	35
Citrus	2	0.00000000	24

Temporal risk (uninfected)

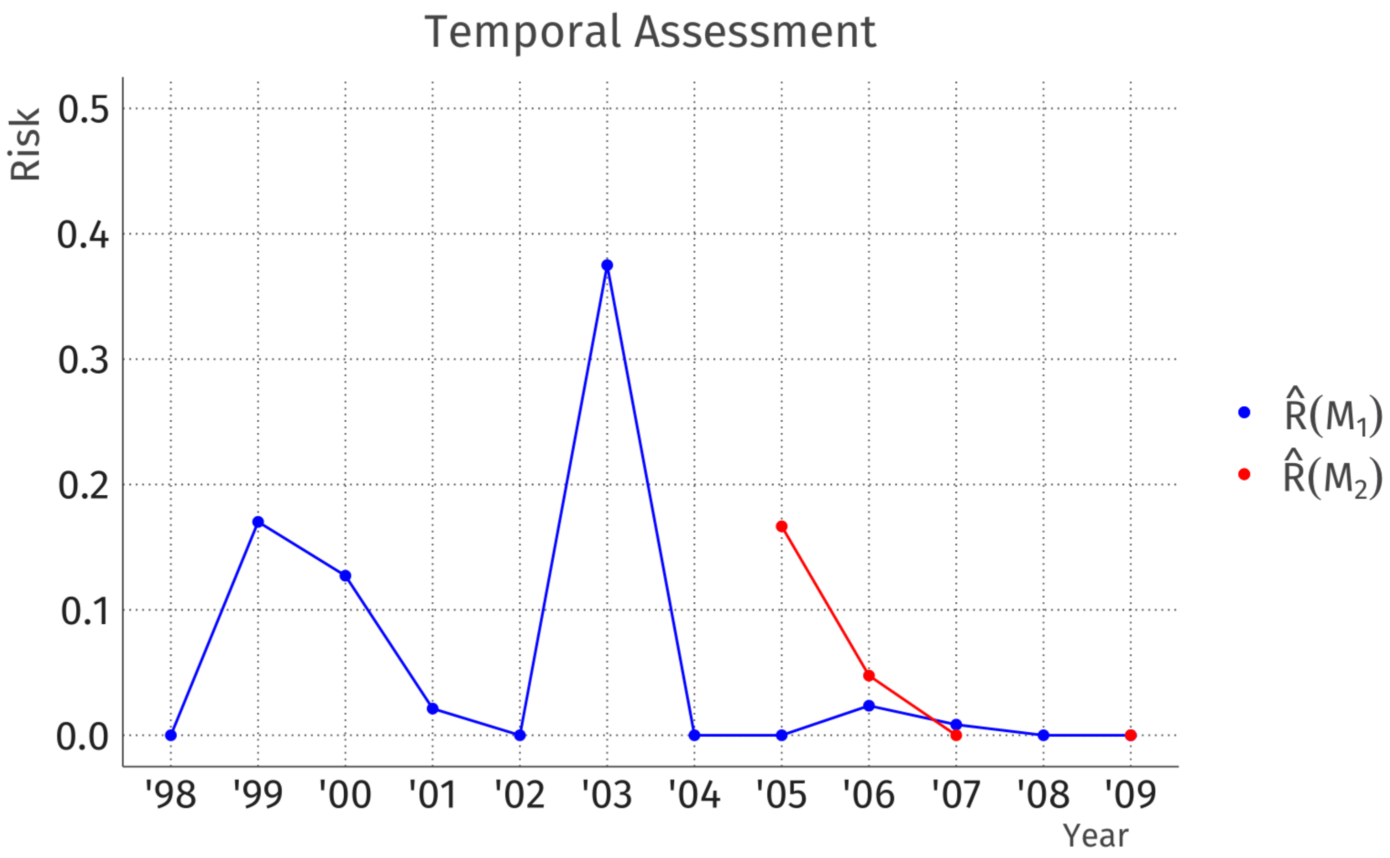
Year	$\hat{R}(M_1)$	n
1998	0.0000000	23
1999	0.1702128	47
2000	0.1272727	55
2001	0.0212766	47
2002	0.0000000	3
2003	0.3750000	8
2004	0.0000000	10
2005	0.0000000	21

Year	$\hat{R}(M_1)$	n
2006	0.0235849	212
2007	0.0084746	118
2008	0.0000000	10
2009	0.0000000	135

Temporal risk (infected)

Year	$\hat{R}(M_2)$	n
2005	0.166667	6
2006	0.047619	21
2007	0	10
2008	n/a	0
2009	0	26

Temporal risk (both)



Risk Summary

- Overall good performance wrt to the data (low misclassification risk)
- Temporal risk worse at the beginning of the model; sharply declines after 2000
- Results are promising, but would like to test against larger data sets

Summary

Summary

- Model is built on a justifiable biological foundation
- Model passes baseline qualitative and quantitative assessments, suggesting it is reasonably effective/accurate
  - imprecision comes from smoothing/uniformity assumptions which vary in their validity
- Can serve as a fundamental tool for the further study of HLB invasion dynamics
  - analyzing historical scenarios
  - interventions to hypothetical future scenarios

Thanks!