The Apparent Growth Rate of the California HLB Epidemic

March 2019

Neil McRoberts[1] and Holly Deniston-Sheets[2]

[1]*Quantitative Biology & Epidemiology Lab, Plant Pathology Department, UC Davis, CA 95616*
[2] *Citrus Research Board, Visalia, CA 93291*

Executive Summary

- DATOC was requested to analyze the growth rate of huanglongbing (HLB) in Southern California to determine if the increase was following an exponential pattern.

- Observed HLB dynamics follow an exponential-like pattern, but the best-fitting descriptive model to date is a segmented linear model with two distinct patterns of growth.

- The extent to which the observed data might be inherently dependent on the processing rate of the survey and/or detection process has not been fully resolved.

- Data description may not necessarily give the most useful predictions of what will happen in future. We therefore recommend that the program open a discussion about accessing the APHIS/Cambridge model to run periodic predictions of the California situation to help with strategic planning.

The problem posed by the Operations Subcommittee to DATOC

A science task force met in August 2018 to advise the CPDPP Committee on the effectiveness of certain program activities and provide input for future planning. One of the task force's key findings was that the program would need to adapt to emerging trends as they arose. Specifically, the rate of increase in detection of HLB should be carefully monitored, to determine at what point current regulatory activities may become impractical due to an overwhelming infection rate.

In the second half of 2017, the rate of detection of new HLB cases (i.e. trees confirmed to be infected with CLas) increased sharply compared with what had been observed prior. The increased detection rate prompted members of the Operations Subcommittee to ask whether the epidemic had entered a period of exponential growth and requested members of DATOC examine the issue and report back. Herein, we report on those analyses.

Analyses

General considerations about observed HLB dynamics:

In a strict sense, as used in epidemiology and population biology, the dynamics of HLB epidemics are inherently exponential when the disease invaded a new area. This is a simple consequence of two basic facts. First, infected trees and vectors produce more copies of the bacterium and the pathogen population increases at a rate that is proportional to how much is already present; this sets up the possibility of exponential growth. Second, early in the epidemic, the supply of healthy host trees and uninfected vectors is not limited, so there is nothing to check the growth potential of the pathogen. With that important basic concept in mind - that HLB dynamics are inherently exponential in situations such as those in California - the question of interest switches from "is the HLB epidemic going exponential?" to something more like "what is happening with the dynamics of HLB in California?" This report is aimed at answering questions of the second type.

It is important to understand that the results for HLB cases reported by CDFA do not reflect the true HLB epidemic. They measure the *apparent* HLB epidemic - the epidemic that we can "see" given the combined operating efficiency of our sampling and confirmation process. As is now widely appreciated, the sampling process has a rather low efficiency, while the confirmation process (via qPCR) is highly accurate for true positives. The overall procedure combines these qualities and has a performance level that averages across them. The important issue is that because we are detecting a proportion of the infected trees, not the whole population of infected trees, the apparent rate of the epidemic lags behind the actual rate of the epidemic. The data are the outcome from two completely separate processes - the process of infection and disease spread, and the process of disease detection and confirmation.

Statistical analyses of the observed increase in HLB cases:

To develop a comprehensive picture of the rate of the epidemic in California we retrieved the time series data reporting the number of confirmed cases. The data set was compressed by dividing it into 6-month intervals and totaling the cumulative number of cases reported as time progressed over the set of intervals.

We fitted a series of different exponential models to the data, and one "broken stick" or "segmented" linear model. The various exponential models had between one and three parameters, from simplest to most

complex. The segmented linear model had four parameters: two slopes and two "intercept" values. Since the purpose of this report is to provide an interpretation of the observed HLB case count for CPDPC, we omit most of the statistical details from the results; they are available on request.

Results

Although each exponential model has the correct general form to account for the data, in general they make the transition from the early phase of the epidemic to the later phase too smoothly, and overestimate the observed case number from about 2016 to the first half of 2018 (Figure 1). The segmented linear model provides a more accurate approximation compared to the actual observed cases. However, the rate of detection in the second, steeper segment is still approximately 100 times faster than the detection rate in the first segment.
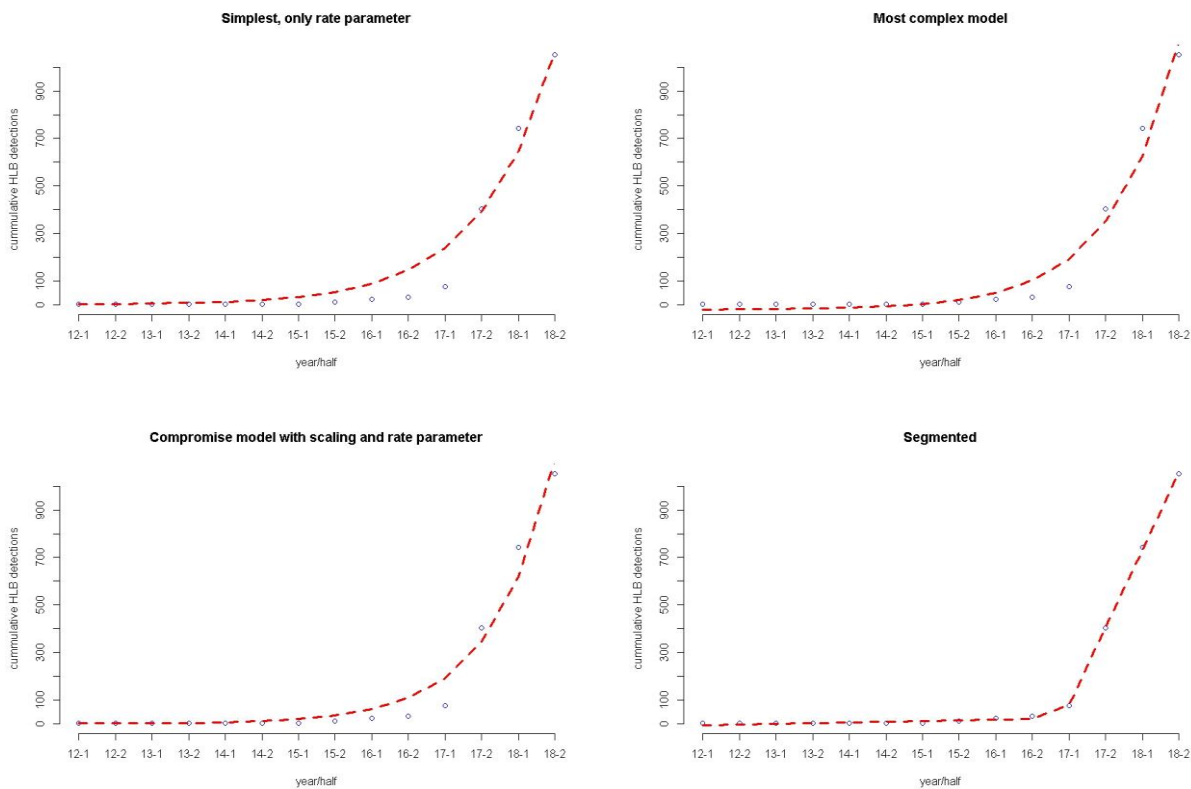

Figure 1. *Observed (blue circles) and fitted values (red lines) for four different statistical models of the six-month cumulative counts of HLB cases in California. The segmented linear model with a break-point in 2017 gives the best overall fit based on information criteria selection.*

Model selection criteria show that the segmented linear model is preferred over the various exponential models (Table 1). The AIC value is a model selection statistic that allows us to compare models that have very different formulations; it is a trade-off between model accuracy and model simplicity, in which very simple but accurate models are favored. It is a penalty score, meaning that lower values are better.

Table 1. *Summary statistics for models fitted to the six-monthly cumulative HLB case data for California*

Model type	AIC	Correlation
Complex exponential	159.15	0.9859
Simple exponential	159.85	0.9845
Compromise exponential	158.24	0.9856
Segmented	102.48	0.9998

Following recommendations from the CPDPC to include sampling intensity in our analyses, and to provide further insight into the observed dynamics, we compiled sampling data for the same periods that were used to estimate the models. Figure 2 shows the sampling intensity data (green data points) on the same chart as the observed HLB cases (black data points), as well as the approximate date that the CPDPC implemented the quadrant sampling protocol (dark red line).
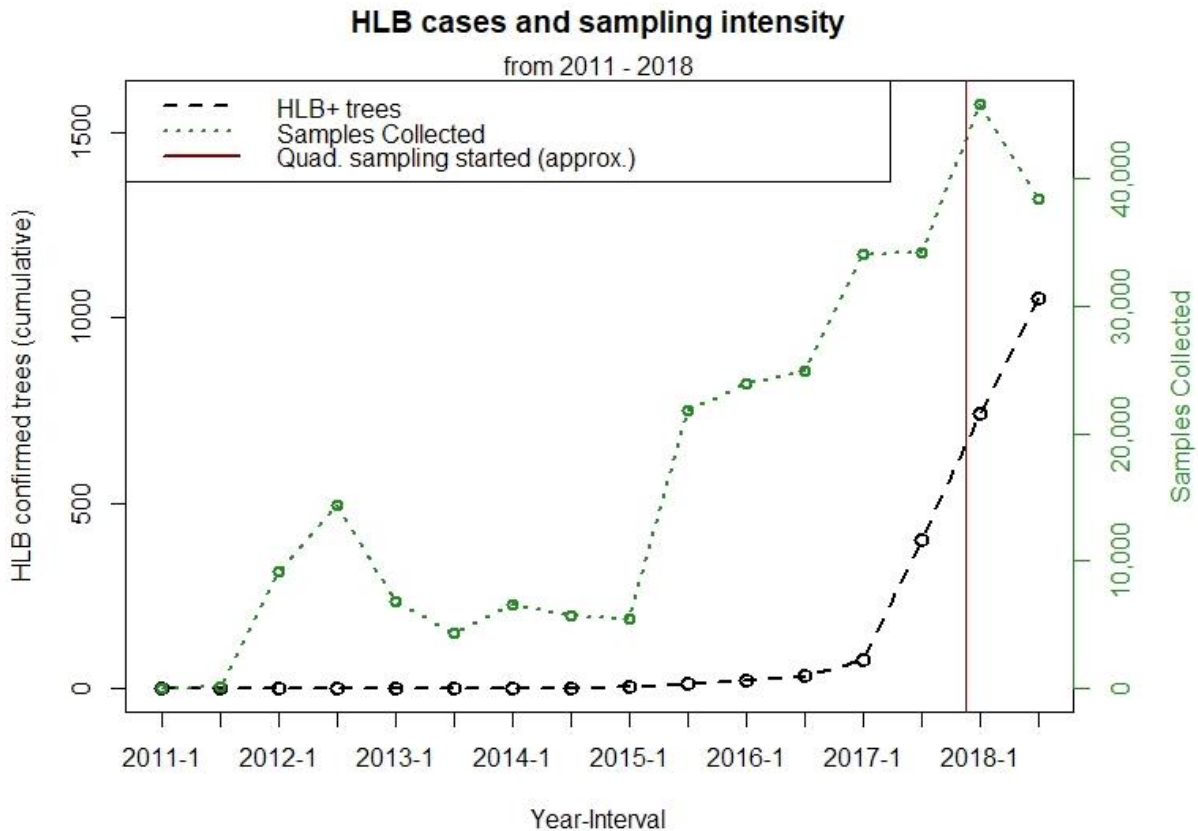

Figure 2. *The observed cumulative HLB case numbers in six-month intervals, and the corresponding number of samples collected from 2011 to 2018.*

Again following direction from the CPDPC committee, in Figure 3 we increased the resolution of the analysis by looking at the change in detection rate month on month for the case data, to see if quadrant sampling had a significant effect on increased detections of CLas. Although changes from mid-2017 are apparent, no pattern directly emerged following the implementation of quadrant sampling.
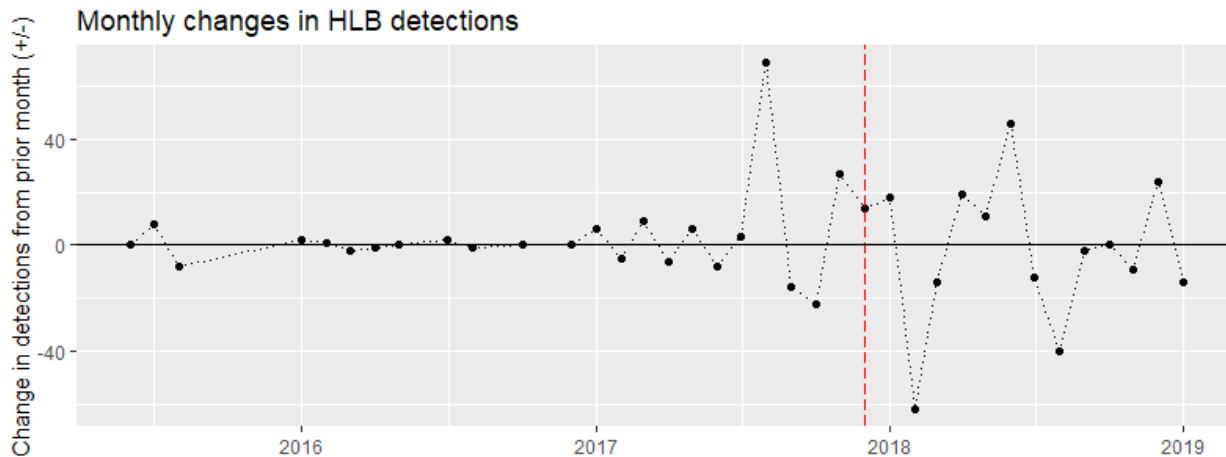

Figure 3. *Month on month changes in the number of confirmed HLB cases in California from 2015 to 2019. The change in the dynamics of detection since mid-2017 is apparent. Earlier dates are excluded from the graph to aid in visualization.*

Conclusions and Recommendations

All of the fitted models confirm that the apparent epidemic is progressing faster now than in the pre-2017 period. The implications for future predictions from the various models are very different, however. All of the exponential models will predict an accelerating number of cases, while the linear model will predict that the case load grows at a constant rate into the future. This difference between the models raises an important question for the CPDPC to consider. Is the linear trend in detection in the second half of the dataset a function of the operational rate of detection and confirmation of the program? In other words, does it represent the maximum reporting rate achievable by the program, or is it a function of the epidemic directly? This distinction matters because if the observed number of cases is a function of processing capacity then the observed, apparent, epidemic will be falling behind the unknown epidemic, and the gap between the two will grow at an unknown rate.

Returning to the concern that prompted this analysis, our main conclusion is that retrospective analysis of the observed data provides only a limited capacity to see into the future; the models that provide the best description of what we have seen so far are at odds with what we know about the dynamics of the disease from first principles. The question of how much of the observed dynamics are driven by the detection process needs to be clarified. Our main recommendation is that the CPDPC, either directly or through DATOC, reaches out to Dr. David Bartels (USDA-APHIS) to explore the possibility of being able to run regular projections of the situation in California using the APHIS/Cambridge simulation model.