Evaluation of predicted medfly (*Ceratitis capitata*) quarantine length in the United States utilizing degree-day and agent-based models

Travis C. Collier[1,2] and Nicholas C. Manoukis[1,3]

[1]Daniel K. Inouye US Pacific Basin Agricultural Research Center (PBARC), United States Department of Agriculture, Agricultural Research Service, Hilo, Hawaii, 96720, USA
[2]corresponding author; email: Travis.Collier@ARS.USDA.gov
[3]email: Nicholas.Manoukis@ARS.USDA.gov

Abstract
Abstracts should be up to 300 words and provide a succinct summary of the article. Although the abstract should explain why the article might be interesting, care should be taken not to inappropriately over-emphasize the importance of the work described in the article. Citations should not be used in the abstract, and the use of abbreviations should be minimized. If you are writing a Research or Systematic Review article, please structure your abstract into Background, Methods, Results, and Conclusions.

Keywords

Please list up to eight keywords to help readers interested
in your article find it more easily.

Take-homes:

1. There is significant variation in predicted quarantine length at different times and locations.

   (a) Captured by normals

   (b) Climate

2. Variation in prediction within time / location (across years) is important.

   (a) Captured by day-of-year (between-year) variation

   (b) Informs reliability of prediction

   (c) Influenced by rare events (eg. cold snaps)

   (d) Prediction based on normal temps vs normal of predictions based on measured temps

3. DD vs ABS comparison

   (a) ABS is better behaved

      i. Seasonal swings less dramatic; Much less discontinuity at beginning of autumn

      ii. Smaller overall range

      iii. Captures common-sense effects missed by DD: eg. extreme cold kills

   (b) Large disagreement between DD and ABS may indicate DD prediction is unreliable/broken

   (c) Variance in predictions should inform management and planning. ABS variance is easier to interpret (KFAT being a dramatic example).

Introduction

Invasions by insects, pathogens and pests increasingly appear to be a defining challenge of the 21st century, facilitated by global connectivity, climatic shifts, and other factors [1] (find paper on number of invasive species). Invasions by insects that do not become established are by their nature less likely to be detected than those that are "successful" from the point of view of the insect. However, when the invading species is of environmental, human health or economic concern there is a greater chance that cases of invasion followed by extirpation would be detected and studied [2]. Eradication of such insects can be desirable and feasible [?] depending on several factors. One factor might be that the new environment is only marginally or seasonally suitable to the invading insect, facilitating its eradication. Another is that the high cost of allowing establishment leads to extensive efforts for eradication. The invasion of *Anopheles gambiae* into Northeastern Brazil in the 1930's [?] is one example of an invasive insect that was successfully eradicated primarily due to the second of these reasons [? ?].

In the case of *An. gambiae* there have been no reports of re invasion, but there are examples of insects that recurrently invade areas outside their native range and are recurrently extirpated within relatively few genrations. The Gypsy moth *Lymantria dispar* in Canada [?] is one such

species, and arguably, the screwworm *Cochlyomyia hominivorax* around the area at the current northernmost edge of its range in Panama, and more recently in Florida, [?] is another. (add citation on Florida)

One of the most important instances of repeated invasion and extirpation by an economically important pest is that of the *Ceratitis capitata* ("Mediterranean fruit fly", or " Medfly") in California. Though the establishment of these pests is disputed, a pattern of invasion, detection, and response, followed by no further detections for a long period, has been established over the last four decades. Some have argued that Medfly in California is an example of a "metainvasion", consisting of multiple sequential or overlapping introductions [?]. Still other researchers have maintained that Medfly is repeatedly eradicated from the state [?] or for different situations in different regions of the state [? ?]. Medfly is occasionally found in other parts of the mainland US such as Florida (cite), and in other countries or areas that are considered free of the pest such as New Zealand (CHECK), and (some free areas here).

The response plan to Medfly in California, and the other "free" regions mentioned above, is extensive and costly, including a quarantine [?]. A practical and important problem is how long to maintain the countermeasures and quarantine following a last detection. Predicting the likely duration of required quarantines would help with management decision making and planning, including potential cost savings by having sufficient but not excessive «««< HEAD resources available.

We analysed the predicted quarantine length (PQL for short) for 11 sites in the continental United States (figure 1 and table 1) ======= resources available. Currently most programs determine quarantine lengths by calculating the amount of time required for a given number of generations to elapse under a thermal unit accumulation ("Degree day") physiological development model (CITE). Recently another approach to determining effective quarantine durations was introduced via Agent-Based Simulations [?].

Agent-Based Simulations (ABS) or models (also called "Individual-Based" or "Multi-Agent"), which can be minimally defined as those where individuals are described as unique and autonomous, and where they usually interact with each other and their environment on a local level [?]. Such simulations describe a system of interest "from the bottom up": by implementing its constituent particles and then observing the system behaviors and dynamics that result from those particles interacting autonomously with their environment and often each other [?]. These models can include arbitrary amounts of complexity, limited only by computational capacity, and are free of limitations imposed by mathematical tractability [?].

While the freedom to include any degree of complexity in an ABS comes at a cost [? ? ?], the pragmatic and paradigmatic benefits of the Agent-Based mindset can be significant for diverse areas of biology [? ?], including biological invasions. The benefits of the Agent-Based mindset have infrequently been applied to questions of insect invasions [?] (but see ?] for an example). Here,

we analysed the predicted quarantine length (PQL for short) for 11 sites in the continental United States »»»> 7e771b8871ecede7be6fd0ff0a0665a5acfcaf86 based on both the standard thermal accumulation degree day method[3] as well as the MED-FOES[4] ABS under sterile insect technique (SIT)[**?**] eradication.

Seasonal variations dominate the variation in quarantine length predictions, so we aggregated the PQL values for each day of the year (Jan. 1, Jan. 2, ect.) across a large number of years (65 for most of locations) to produce normals.

Figure 1. Location of sites reported on.

Methods

Sites and Temperature Data

Hourly air temperature data for 11 sites was downloaded from NOAA's Integrated Surface Database (ISD) dataset[5, 6]. The airport sites shown in figure 1 were chosen for their biological relevance and availability of high quality hourly data over a long time frame.

Sites are referred to here by the last three letters of the callsign shown in table 1. For 8 sites (SFO, FAT, LAX, RIV, SAN, JAx, TPA, and MIA), temperature data starting on 1950-01-01 was used. The 3 other sites contained large (> 14 days) gaps or other problems in the early years of their data, so data starting on 1970-01-01 for IAH and 1973-01-01 for BUR and MCO was used. For all sites, temperature data from the start date through 2017-05-15 was used for quarantine length predictions for dates ranging from the start date for the site up to 2016-01-01. Data was fetched and parsed using the `Fetching and parsing ISH.ipynb` program. Records for the same station callsign were merged, since identification, format, and precise location of stations has changed over the years. The data was then cleaned using the `Cleaning temperatures.ipynb` program by removing outliers, identifying large gaps (> 3 hours), resampling to every hour on the hour using linear interpolation, and filling the large gaps using day-over-day linear interpolation (interpolating using values for the same hour of day from previous and following days). The processing programs and resulting temperature datasets are provided in the Supplemental Materials.

Degree-Day Calculation

Degree-days were computed by the single-sine method[3] using a base development temperature of 12.39°C (53.3°F) and 345.56 degree-days Celsius (DDc; 622 DDf) per generation following the standard required by California Department of Food and Agriculture regulation 3406(b)[7, 8]. Since we have hourly temperature data, we also calculated degree-days by simple summation for comparison[9]. For each date, the number of days required to pass 3 generations of degree-day based life cycles was computed. These calculations are implemented in `Temperature functions.ipynb` in the Supplemental Materials.

Agent-based Simulations: MED-FOES

MED-FOES[4, 10] is an agent-based simulation explicitly modeling the eradication of a population of Medflies under inundative sterile male releases (aka: sterile insect technique or SIT). A MED-FOES simulation models a single non-spatial population starting from a given age distribution and number of individuals through the time the population experiences extirpation when the last potentially fertile female dies or mates with a sterile male. The simulation is parameterized on the initial population, additional mortality induced by control efforts, the effectiveness of SIT, and a large number of biological parameters for which ranges are known from the literature, including temperature-dependent development and mortality. The simulation is fed the same hourly timeseries of temperature values which was used for degree-day calculations and updated in hourly time steps.

Due to the fact that only ranges are known for many of the parameters, 2500 individual MED-FOES simulations were run for each given start date at each site sampling different regions of parameter-space using a Latin Hypercube Sampling[11] procedure. This set of simulations is referred to as a 'run'.

Varying the start date for different simulations was achieved by simply starting at different points in the input temperature file; for this study a run starting every 7 days over the range of dates available for each site. Each set of runs for a single site over a range of starting dates is referred to as a 'runset'. All runsets were conducted with the same input parameters aside from temperature. Initial population numbers were chosen as a standard outbreak based upon several real outbreaks modeled previously[10].

MED-FOES version 0.6.2 was run under OGS/Grid Engine 2011.11 on a CentOS 6.6 HPC cluster. The MED-FOES code, configuration files, and helper scripts are provided in the Supplemental Materials. Overall, we created 11 runsets (one for each site), each containing runs starting every 7 days over the input temperature data range for that site, where each run contained 2500 individual simulations sampling different regions of biologically plausible parameter space.

The MED-FOES data is summarized here by the number of days from the start date required for 95% of the simulations in a run to be eradicated, referred to as pe95.

Statistical analysis

The main results reported here are 'normals' in a meteorological sense of term, but without the typical running mean smoothing which would complicate interpretation

Table 1. Weather station (NOAA ISD) sites used.

Callsign	Station Name	State	Latitude	Longitude	Elevation	Start year
KSFO	SAN FRANCISCO INTERNATIONAL A	CA	+37.620	-122.365	2.4	1950
KFAT	FRESNO YOSEMITE INTERNATIONAL	CA	+36.780	-119.719	101.5	1950
KBUR	BURBANK-GLENDALE-PASA ARPT	CA	+34.201	-118.358	236.2	1973
KLAX	LOS ANGELES INTERNATIONAL AIR	CA	+33.938	-118.389	29.6	1950
KRIV	MARCH AIR RESERVE BASE	CA	+33.900	-117.250	468.2	1950
KSAN	SAN DIEGO INTERNATIONAL AIRPO	CA	+32.734	-117.183	4.6	1950
KJAX	JACKSONVILLE INTERNATIONAL A	FL	+30.495	-81.694	7.9	1950
KIAH	G BUSH INTERCONTINENTAL AP/HO	TX	+29.980	-95.360	29.0	1970
KMCO	ORLANDO INTERNATIONAL AIRPORT	FL	+28.434	-81.325	27.4	1973
KTPA	TAMPA INTERNATIONAL AIRPORT	FL	+27.962	-82.540	5.8	1950
KMIA	MIAMI INTERNATIONAL AIRPORT	FL	+25.791	-80.316	8.8	1950

of the results. For a variable of interest (eg. temperature or PQL), all values for the same calendar day irrespective of year (eg. 20-July) are aggregated and summary statistics such as mean, minimum, maximum, and standard deviation are computed for each aggregation. Figure 2 shows the minimum and maximum of the normals for temperatures along with the mean of the normal PQL based on 3 generation degree day accumulation and MED-FOES 95% extirpation. Figures 4 and 5 show the standard deviations (σ) of the normals for the degree day and MED-FOES based PQL. `Temperature functions.ipynb` contains the code used to perform normal calculations, and the code generating these figures is `Summary Figures.ipynb`.

The results reported here are the normals of PQL computed using the full temperature time series as opposed to computing PQL from the normal of the temperature time-series. While the latter is fairly common practice, it is not mathematically proper since, as with means, the normal of a function of X is not generally equal to the function applied to the normal of X. Additionally, by computing the normals of the predicted quarantine durations, we can investigate properties of the distribution of values as shown in figures 4 and 5 and the "supernorm" supplemental figures.

Results

There is significant variation in PQL across both time and location. The temporal variation in PQL is dominated by a yearly cycle which is characterized well by the normal values shown in figure 2. Table 2 shows the percentage of variance in quarantine length predictions which is captured by the mean of the normal yearly cycle (aka. R^2) for each site. At all but one site, greater than 75% of the variance in both degree day and MED-FOES based PQL is accounted for by the mean normal, and the majority exceed 90%. SFO is an exception to the overall rule, with the mean normal accounting for only 9.1% of the variation in degree day based PQL and 28.0% of the MED-FOES based PQL. This is also shown in the respective 'supernorm' sup-

Table 2. Percentage of PQL variance captured by the mean of the normal. DD PQL is the 3 generation single sine degree day based prediction, and pe95 is the MED-FOES agent-based simulation predictions.

	R^2	
Site	DD PQL	pe95
SFO	9.12%	28.01%
FAT	93.93%	75.68%
BUR	90.71%	90.88%
LAX	80.17%	83.07%
RIV	92.23%	81.89%
SAN	80.99%	80.91%
JAX	96.45%	94.78%
IAH	95.10%	91.80%
MCO	94.62%	95.77%
TPA	91.91%	94.40%
MIA	88.42%	92.00%

plemental figures S?? and S??.

Seasonal dependence

The seasonal variation, evidenced by the general shape of the curves shown in figure 2, is doubtless familiar to anyone engaged in medfly pest management. Outbreaks starting in the late summer, autumn, or early winter will extend through relatively cold periods where thermal dependent development will be slow and therefore extend the duration of quarantine required for 3 generations of degree days to accumulate (referred to as DD PQL hereafter). Similarly, outbreaks starting in the spring or early summer often lead to short quarantines due to the relatively high temperatures.

This familiar pattern is also predicted by the MED-FOES ABS despite it being quite different in nature from simple degree day accumulation. However, the MED-FOES predictions (pe95) show a smaller seasonal swing. pe95
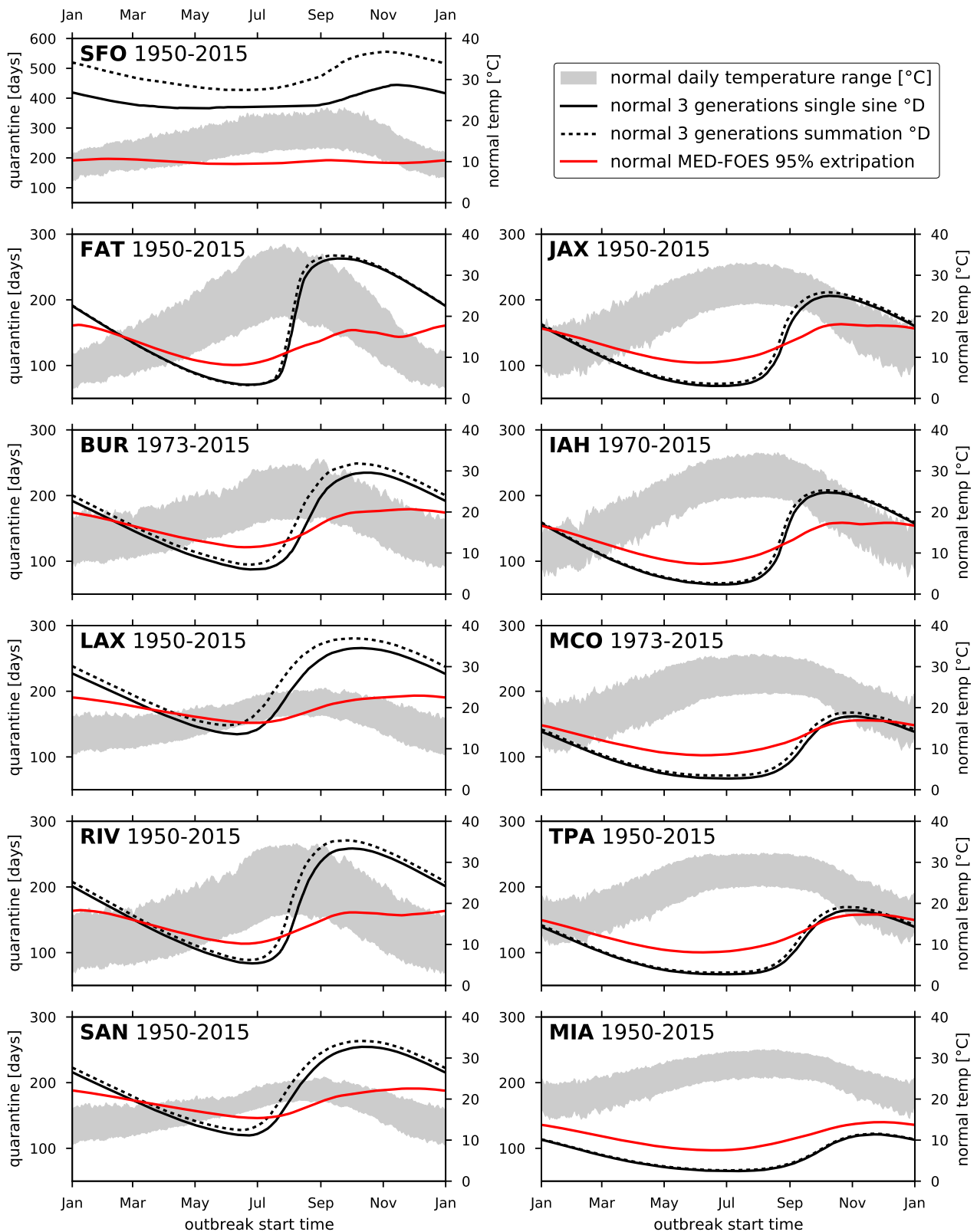
Figure 2. **Summary of normal quarantine length predictions for each site.**

generally predicts as smaller overall range of PQL, with longer quarantines than DD PQL for spring and early summer outbreaks, and shorter quarantines for late summer through early winter.

A particular feature of interest, shown most dramatically

at FAT in figure 2, is that MED-FOES PQL often flattens out or even dips for quarantines starting in the late autumn or early winter. This is due to relatively rare and brief cold-snaps, normally lasting only a few hours, which increase mortality. Since DD PQL does not account for mortality,

it misses the effect of cold-snaps entirely. This cold-snap effect is most clearly seen at more northern and more inland sites where cold-snaps are more likely: particularly FAT and RIV, but also BUR, LAX, JAX, and IAH.

Geographic dependence

PQL generally shows a positive correlation with latitude[**?**]. Sites are ordered by latitude in the figures and tables. As seen in figure 2, higher latitude sites tend to have longer PQL as well as larger seasonal swings for both DD and MED-FOES based predictions.

Figure 3 shows the relationship between PQL and latitude. An ordinary least squares fit to the median PQL at each site shows a significant slope for both DD (F=14.08, p=0.005) and MED-FOES (F=10.55, p=0.010), but the DD based predictions are more sensitive to latitude than MED-FOES (coefficients of 17.39 and 4.78 respectively). Additionally, the predictions for SFO, and to a lesser extent FAT, where the DD model for medfly appears to break down are much better behaved under MED-FOES.

Variance and uncertainty

Figures 4 and 5 report the standard deviation (σ) of the normal for DD PQL and MED-FOES PQL respectively. These indicate the year to year variability of the PQL for outbreaks starting at a given time of the year and can be used to indicate the uncertainty of quarantine length predictions relative to the actual quarantine length which will be required. Similar information is represented by the inter-quartile ranges shown in figure 3 and the 'supernorm' supplemental figures.

With the exception of SFO, the mean normal is a good predictor of DD PQL with the ....

Discussion

The discussion should include the implications of the article results in view of prior work in this field.

Conclusions

Please state what you think are the main conclusions that can be realistically drawn from the findings in the paper, taking care not to make claims that cannot be supported.

Author contributions

In order to give appropriate credit to each author of an article, the individual contributions of each author to the manuscript should be detailed in this section. We recommend using author initials and then stating briefly how they contributed.

Competing interests

All financial, personal, or professional competing interests for any of the authors that could be construed to unduly influence the content of the article must be disclosed and will be displayed alongside the article.

References

[1] Daniel Simberloff, Jean-Louis Martin, Piero Genovesi, Virginie Maris, David A. Wardle, James Aronson, Franck Courchamp, Bella Galil, Emili García-Berthou, Michel Pascal, Petr Pyšek, Ronaldo Sousa, Eric Tabacchi, and Montserrat Vilà. Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*, 28(1):58–66, January 2013. ISSN 0169-5347. doi: 10.1016/j.tree.2012.07.013. URL http://www.sciencedirect.com/science/article/pii/S0169534712001747.

[2] Andrew M. Liebhold and Patrick C. Tobin. Population Ecology of Insect Invasions and Their Management*. *Annual Review of Entomology*, 53(1): 387–408, January 2008. ISSN 0066-4170, 1545-4487. doi: 10.1146/annurev.ento.52.110405.091401. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.ento.52.110405.091401.

[3] G. L. Baskerville and P. Emin. Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology*, 50(3):514–517, 1969. doi: 10.2307/1933912.

[4] Nicholas C. Manoukis, Brian Hall, and Scott M. Geib. A computer model of insect traps in a landscape. *Scientific Reports*, 4:7015, November 2014. doi: 10.1038/srep07015. WOS:000344760700005.

[5] Adam Smith, Neal Lott, and Russ Vose. The Integrated Surface Database: Recent Developments and Partnerships. *Bulletin of the American Meteorological Society*, 92(6):704–708, June 2011. doi: 10.1175/2011BAMS3015.1.

[6] Integrated Surface Database (ISD) | National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC). URL https://www.ncdc.noaa.gov/isd. Last visited 2017-07-05.

[7] Mediterranean fruit fly: Regulation and quarantine boundaries. URL https://www.cdfa.ca.gov/plant/medfly/regulation.html. Last vistited 2017-07-17.
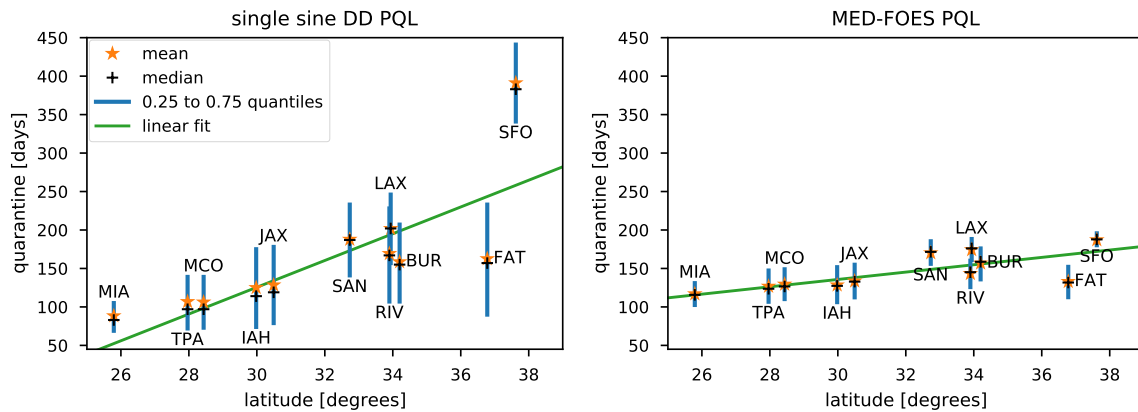
Figure 3. **Predicted quarantine length dependence on latitude. For each site, the mean, median, and inter-quartile range are shown (similar to a boxplot). An ordinary least-squares linear fit to the median values is shown by the green lines. The left pane is for single sine degree day predictions, and MED-FOES based predictions (pe95) in the right pane.**
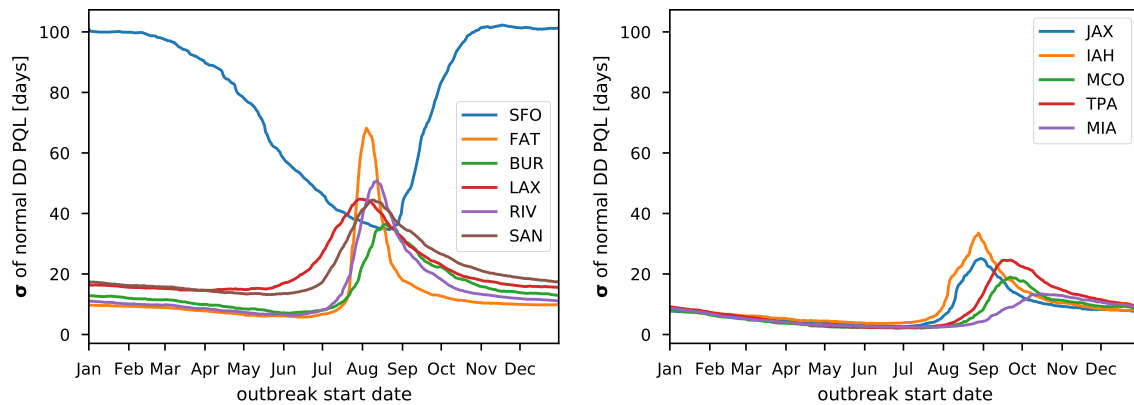

Figure 4. **Variation in quarantine length prediction based on 3 generations of single-sine degree day accumulation.**
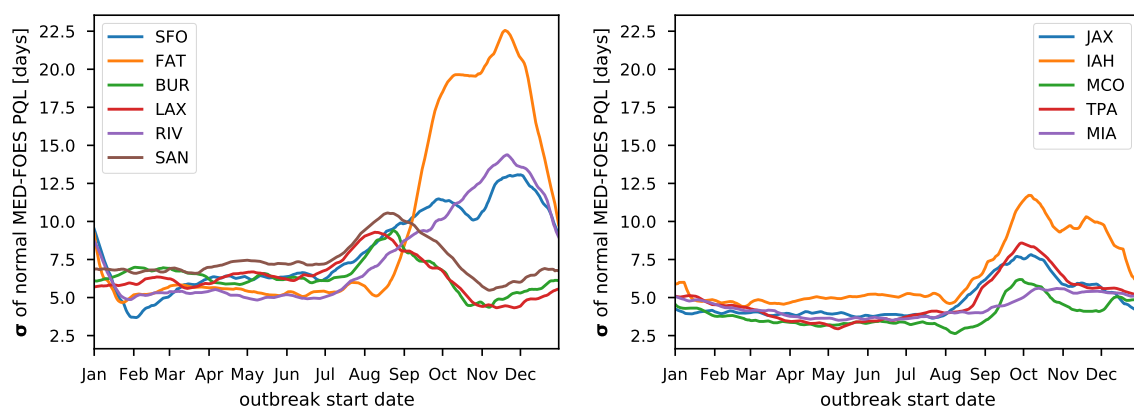

Figure 5. **Variation in quarantine length prediction based on 95% of MED-FOES simulations showing extirpation.**

[8] Califorina code of regulations, title 3, section 3406. URL https://www.cdfa.ca.gov/plant/medfly/docs/regs/3406-TXT-medfly.pdf. Last vistited 2017-07-17.

[9] William J. Roltsch, Frank G. Zalom, Ann J. Strawn, Joyce F. Strand, and Michael J. Pitcairn. Evaluation of several degree-day estimation methods in california climates. *International Journal of Biometeorology*, 42(4):169–176, Mar 1999. doi: 10.1007/s004840050101.

[10] Nicholas C. Manoukis and Kevin Hoffman. An agent-

based simulation of extirpation of Ceratitis capitata applied to invasions in California. *Journal of Pest Science*, 87(1):39–51, March 2014. ISSN 1612-4758, 1612-4766. doi: 10.1007/s10340-013-0513-y. URL https://link.springer.com/article/10.1007/s10340-013-0513-y.

[11] S. M. Blower and H. Dowlatabadi. Sensitivity and uncertainty analysis of complex models of disease transmission: An hiv model, as an example. *International Statistical Review / Revue Internationale de Statistique*, 62(2):229–243, 1994. ISSN 03067734, 17515823. URL http://www.jstor.org/stable/1403510.