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

Medfly (*Ceratitis capitata*) bad... covfefe...
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 resources available.
We analysed the predicted quarantine length (PQL for short) for 11 sites in the continental United States based on both the standard thermal accumulation degree day method[1] as well as the MED-FOES[2] agent based simulation of population of medfiles 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.

Methods

Sites and Temperature Data

Hourly air temperature data for 11 sites was downloaded from NOAA's Integrated Surface Database (ISD) dataset[3, 4]. The airport sites shown in Table 1 were chosen for their biological relevance and availability of high quality hourly data over a long time frame.

Figure 1. Location of sites reported on.

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[1] 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)[5, 6]. Since we have hourly temperature data, we also calculated degree-days by simple summation for comparison[7]. 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[2, 8] 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

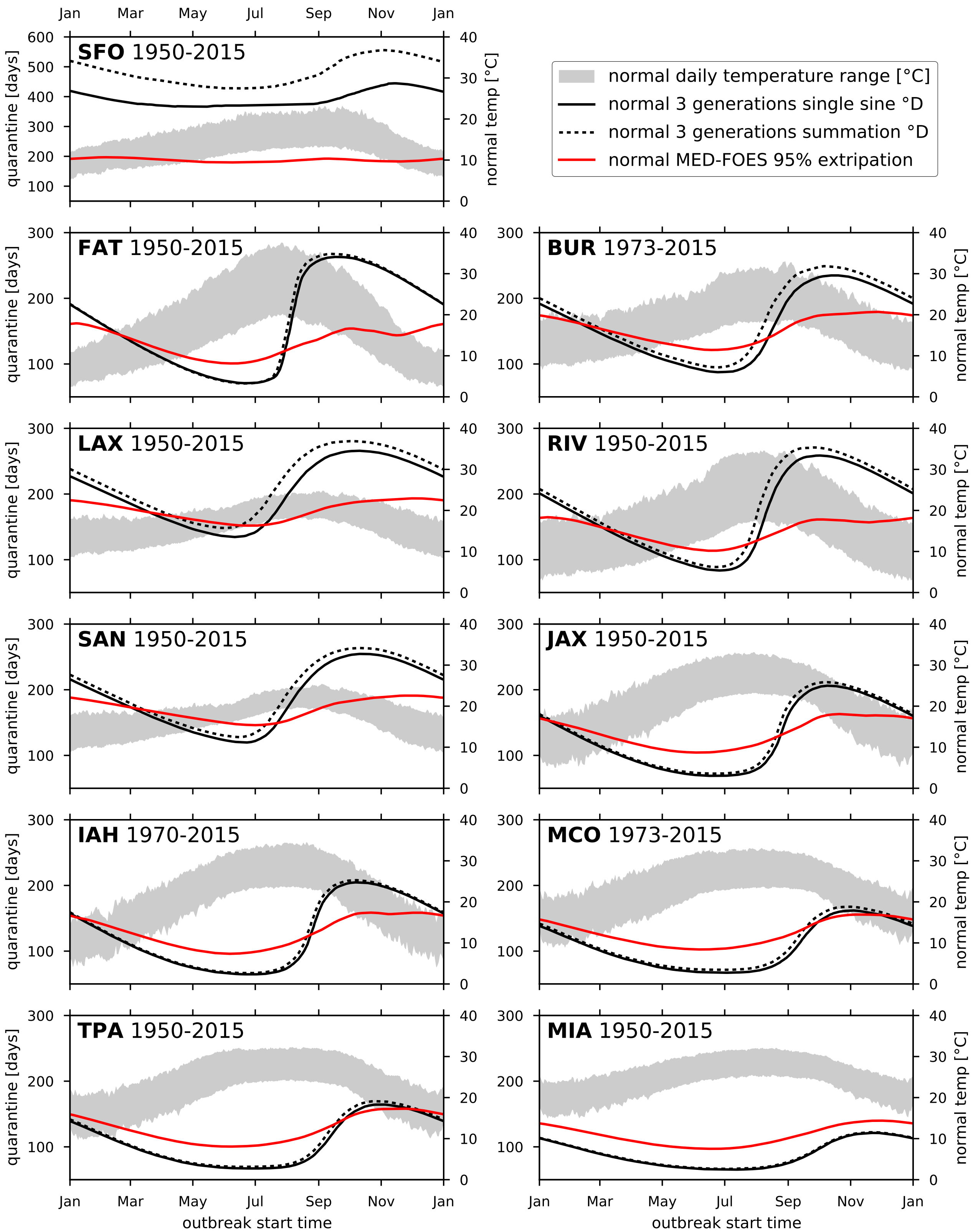

Figure 2. Summary of normal quarantine length predictions for each site.

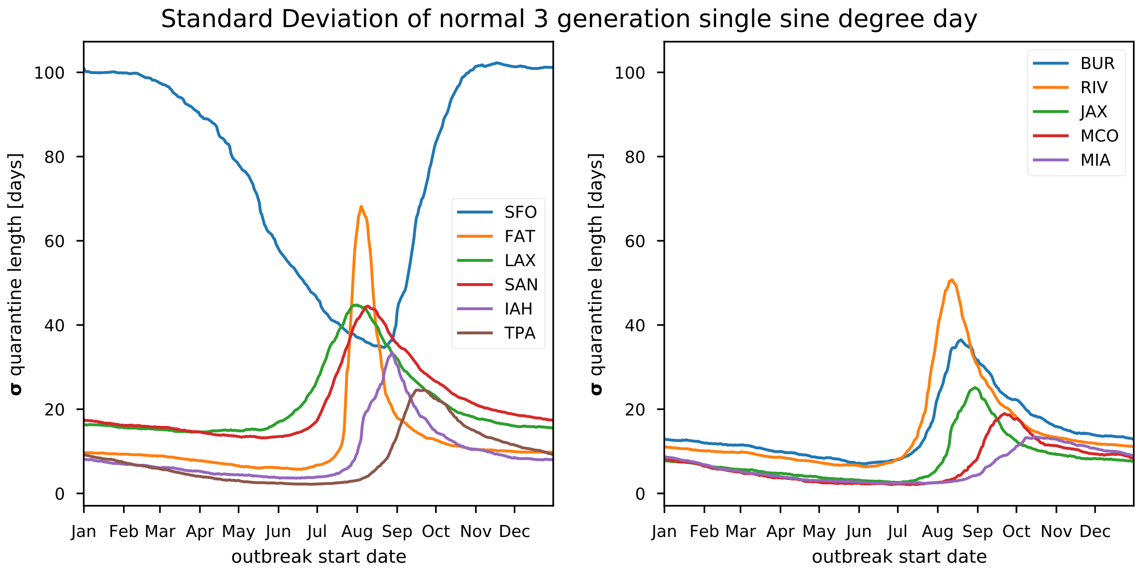

Figure 3. Variation in quarantine length prediction based on 3 generations of single-sine degree day accumulation.

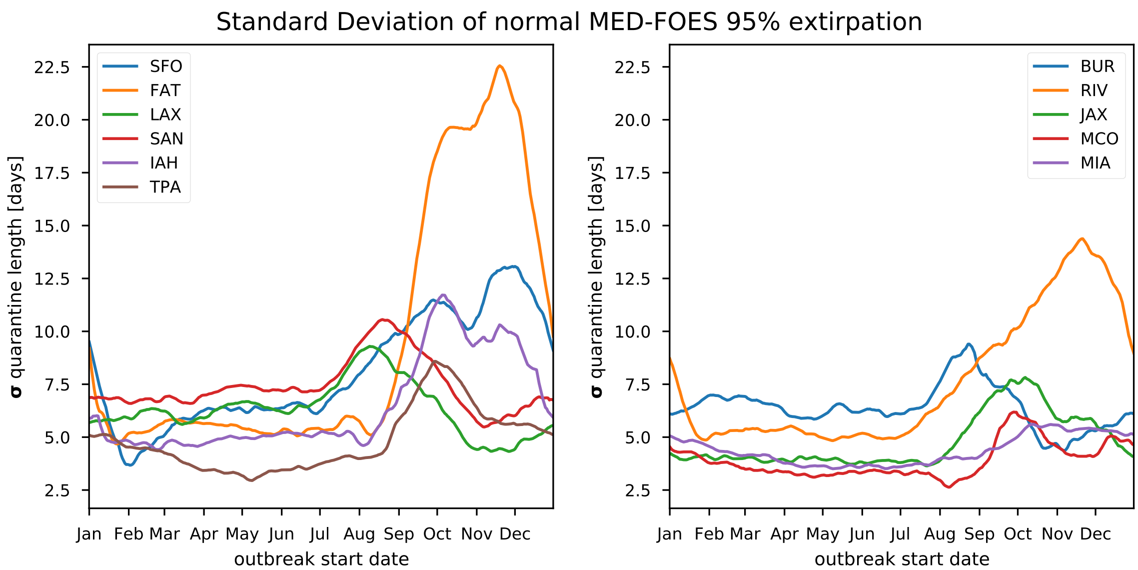

Figure 4. Variation in quarantine length prediction based on 95% of MED-FOES simulations showing extirpation.

Table 1. Sites.

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

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[9] 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[8]. 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 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 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 3 and 4 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 timeseries. 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 3 and 4 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 normal yearly cycle for each site. At all but one site, $>75\%$ of the variance in both degree day and MED-FOES based PQL is accounted for by the normal, and the majority exceed 90%. SFO is an exception to the overall rule, with the normal accounting for only 9.1% of the variation in degree day based PQL and 28.0% or the MED-FOES based PQL. This is more clearly shown in the respective 'supernorm' supplemental figures S?? and S??.

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

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

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%

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 are 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) are show a smaller seasonal swing. pe95 generally predicts longer quarantines than DD PQL for spring and early summer outbreaks, and shorter quarantines than DD PQL for late summer through early winter.

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.

Grant information

Please state who funded the work discussed in this article, whether it is your employer, a grant funder etc. Please do not list funding that you have that is not relevant to this specific piece of research. For each funder, please state the funder's name, the grant number where applicable, and the individual to whom the grant was assigned. If your work was not funded by any grants, please include the line: 'The author(s) declared that no grants were involved in supporting this work.'

Acknowledgements

This section should acknowledge anyone who contributed to the research or the article but who does not qualify as an author based on the criteria provided earlier (e.g. someone or an organisation that provided writing assistance). Please state how they contributed; authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.
Please do not list grant funding in this section.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Mediterranean fruit fly: Regulation and quarantine boundaries. URL `https://www.cdfa.ca.gov/plant/medfly/regulation.html`. Last vistited 2017-07-17.

[6] 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.

[7] 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.

[8] 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`.

[9] 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`.