ArboMAP: Arbovirus Modeling and Prediction to Forecast Mosquito-Borne Disease Outbreaks

Summary of Model Outputs (v2.2)

Justin K. Davis and Michael C. Wimberly
(justinkdavis@ou.edu, mcwimberly@ou.edu)
Geography and Environmental Sustainability, University of Oklahoma

Updated March 04, 2020

Data used for predictions

Believable data

ArboMAP is set to believe weather data from the years 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018. If weather are data are necessary outside of this range (e.g. predicting into the far future) then the program will use daily averages from observed years to fill in times where data are not available. A year appearing in this list does not mean that those data are actually available to ArboMAP - solely that the program will permit their use if they are present. These data need to be actually present in the weather directory, but observations will be discarded if ArboMAP has not been instructed to believe those years.

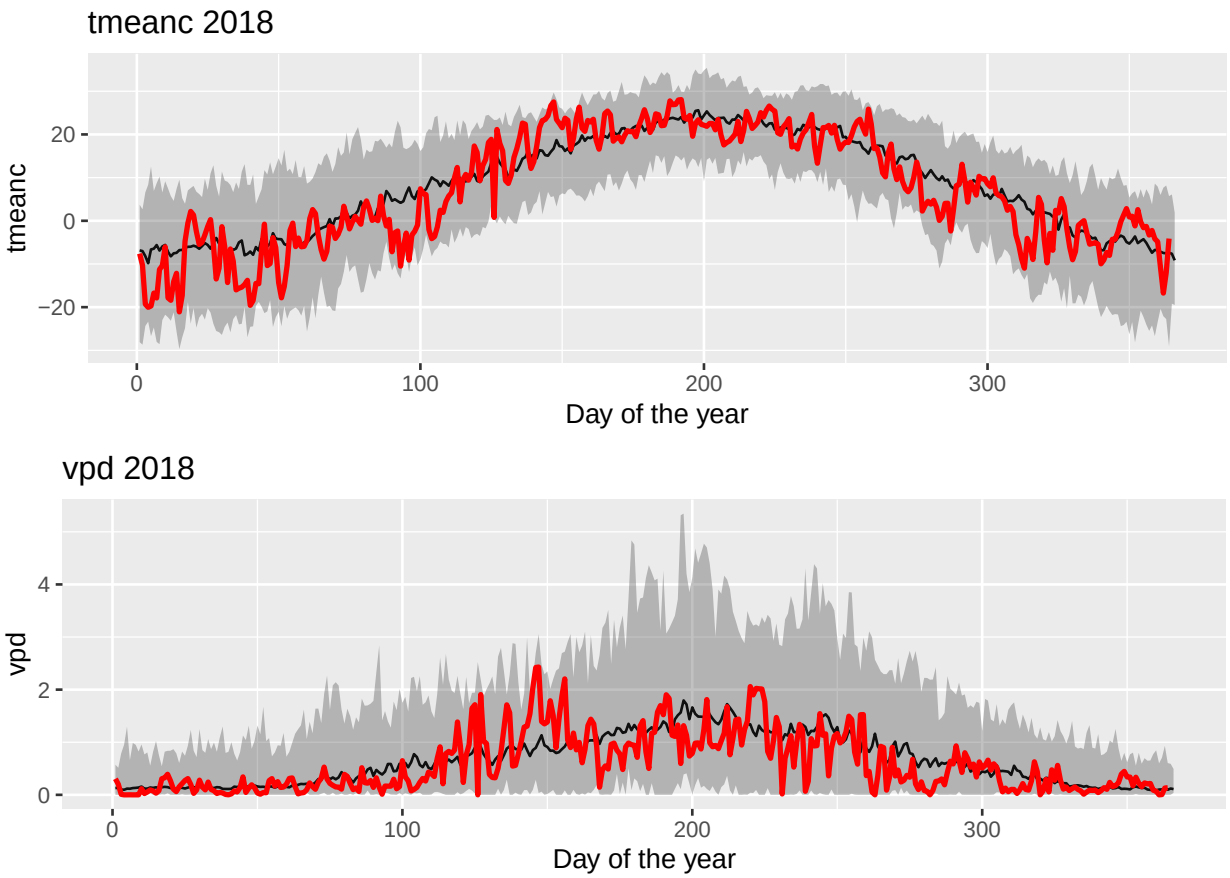
ArboMAP is set to believe mosquito infection data from the years 2004, 2005, 2006, 2007, 2008, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018. Years without mosquito data are assumed to have average mosquito infection rates. Just because a year is present in this list does not imply that those data have actually been made available to ArboMAP - just that ArboMAP will use those data if they are available. Even if there were no positive pools in a given year, if there were any pools tested then the data will be useful; zero infection rates do predict low-risk years and should be believed.

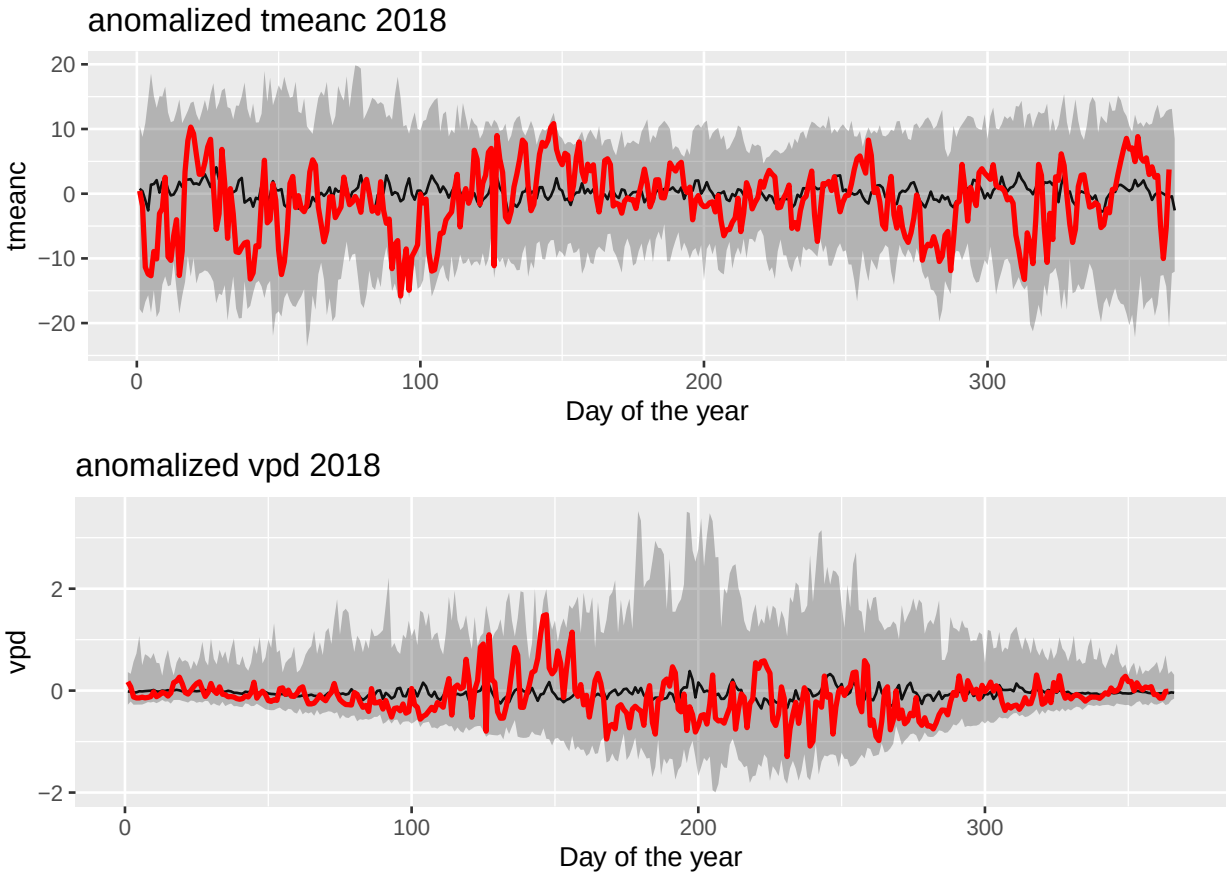
ArboMAP is set to believe human infection data from the years 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018. This list **should not** include the year in which human cases are being predicted; e.g. if you are predicting for 2020, then 2020 should not be in this list. If, however, you are running historical estimates (as opposed to future predictions), then it is reasonable to let ArboMAP trust these data. Even if there were no cases in a year, if zero was the actual reported number of cases, then the year should be believed.

If you believe human data in a year and either 1) do not believe the mosquito data from that year or 2) are not able to obtain an estimated mosquito infection growth rate for that year (e.g. too few mosquito pools reported), then ArboMAP will substitute the mosquito data with a single constant for risk that year. This allows us to estimate relationships with environmental data even when mosquito data are not available, but it means the fit in that year will be more accurate than would have been permitted by the use of mosquito infection data.

Weather data

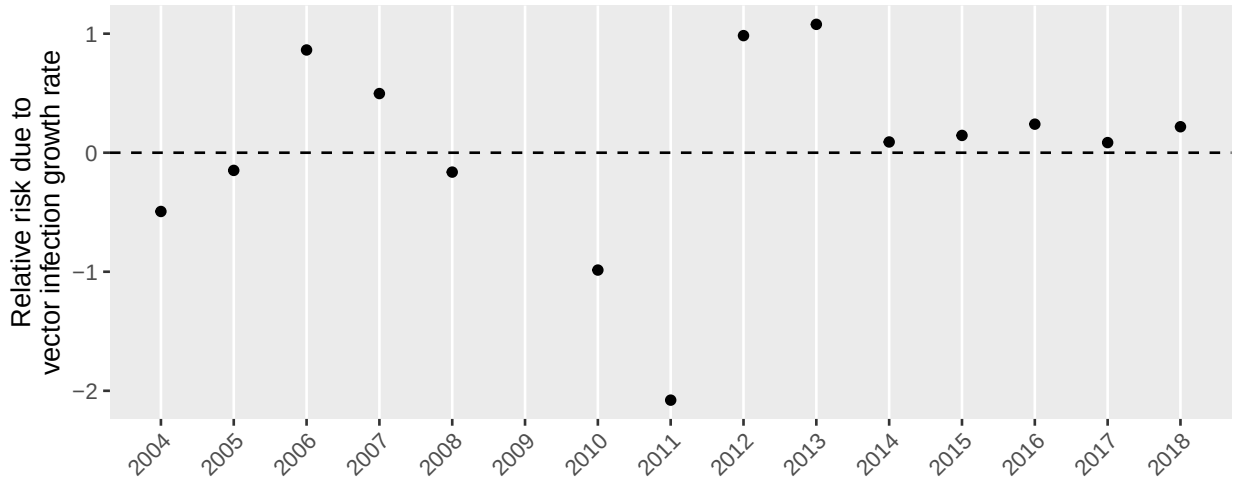
Weather data from the gridMET data set range from 2000-01-01 to 2018-12-30. Below are graphs of statewide daily averages of tmeanc and vpd. Observations for the current year are in red. Black is the medium from all other years, and the grey band indicates the max/min ever observed. Below this are the anomalized weather indices, from which the weekly averages have been subtracted to show deviations above/below the mean.



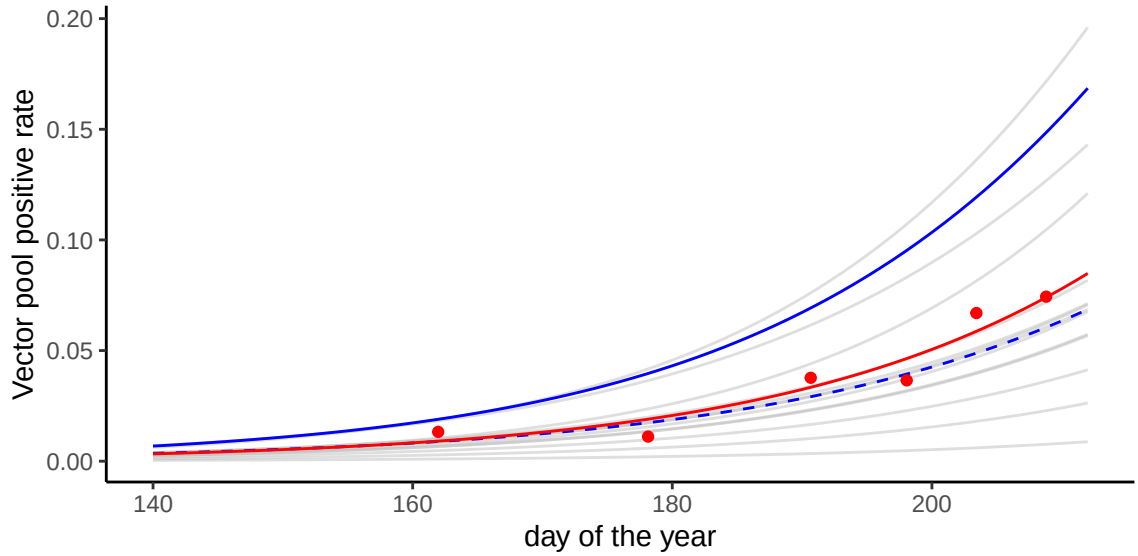


Vector infection data

There are 15914 samples in the vector testing database. For 2018, there are 1697 tested samples, with 66 (3.9%) positive. The estimated risk of human infection due to the early-season vector infection growth rate is shown below. Higher means that the pathogen is spreading more rapidly among vectors, and more human cases should be expected. If an estimate does not appear in a year, it is likely because the `mosquitobelievableyears` variable in ArboMAP was not set to include it; i.e. ArboMAP was told to disregard mosquito infection data for this year.

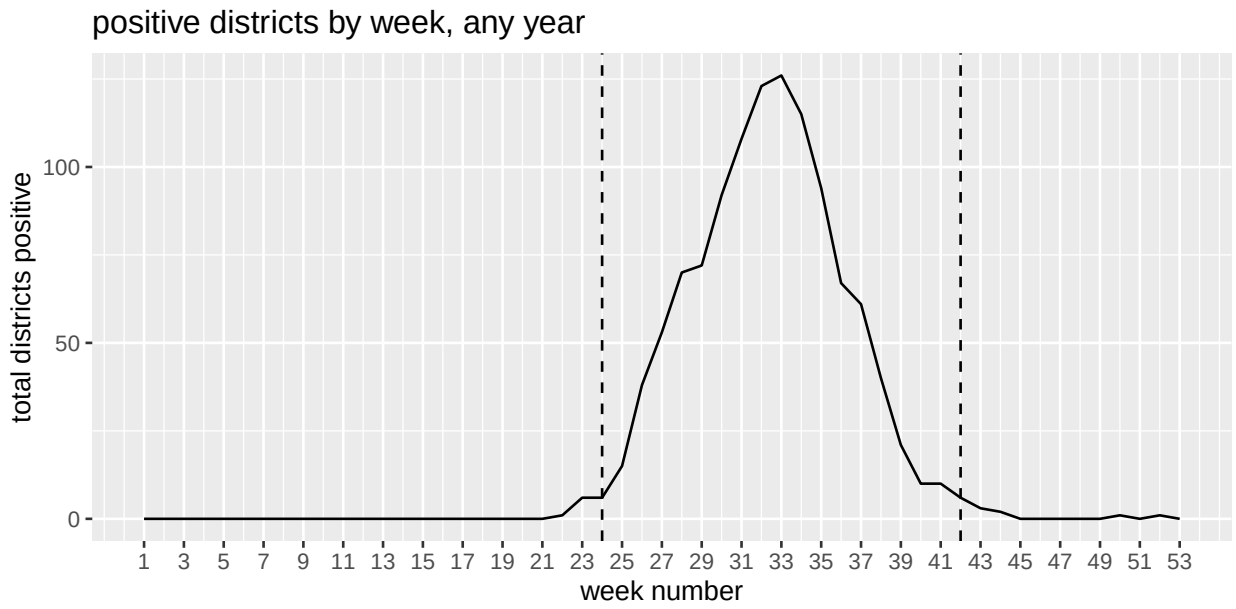


The following graph shows the estimated growth of positive samples for every year (grey), with 2012 (blue) and 2017 (blue, dashed) selected for comparison, and estimates and observations for 2018 (red). The lines are modeled sample positive rates; the actual statewide positive sample rate for 2018 is shown here by grouping observations nearby in time.



Human data

In the graph below, positive districts by week are shown over all years. Only weeks within the two dashed lines (excluding 2% total of the earliest and latest cases) are used in modeling for numerical stability.



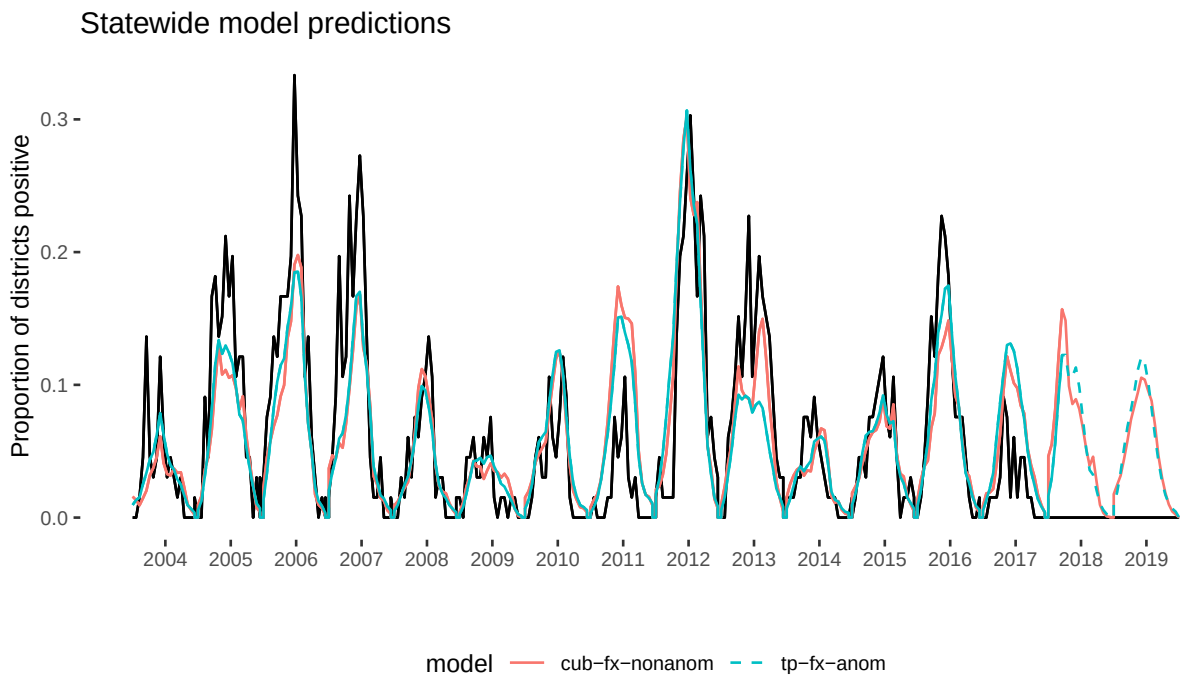
The predictive model of human cases was calibrated using 1300 historical cases, not including any cases from 2019. No cases from 2019 are used to make predictions; the estimates for this year are based solely on weather and vector data. Typically, 23% of a year's cases occur before the end of this week in any given year.

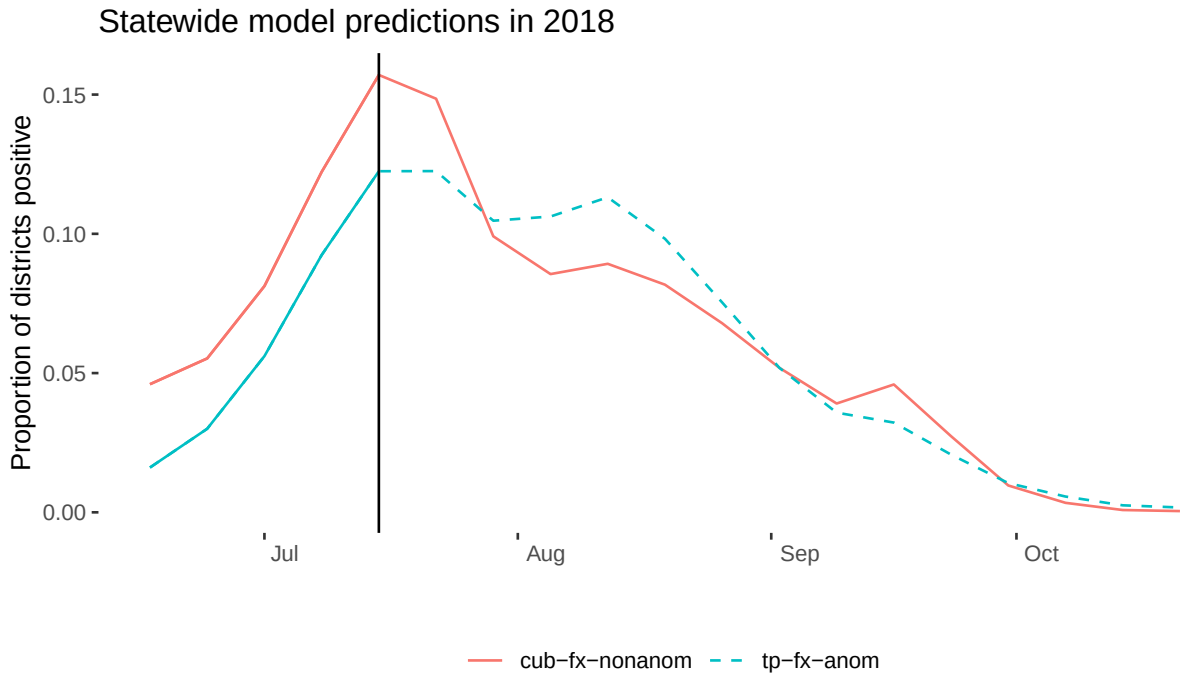
Model results

Statewide trends

The graphs below show observed statewide risk (black) and estimated risk (red) up to Sunday 2019-12-29. Observed risk should be completely zero during the last year - these data are not used in the model, and will only be updated once final human case data are received at the end of the year.

ArboMAP is performing a exact fit for data in years: 2009. An exact fit occurs whenever there are human data in a year (the year is in `humanbelievableyears`) but no corresponding mosquito data (the year is not in `mosquitobelievableyears`) or for some reason the risk factor due to mosquitoes could not be estimated. We do not want to ignore human data if no mosquito data are available, since we can still examine the relationship with environmental covariates.



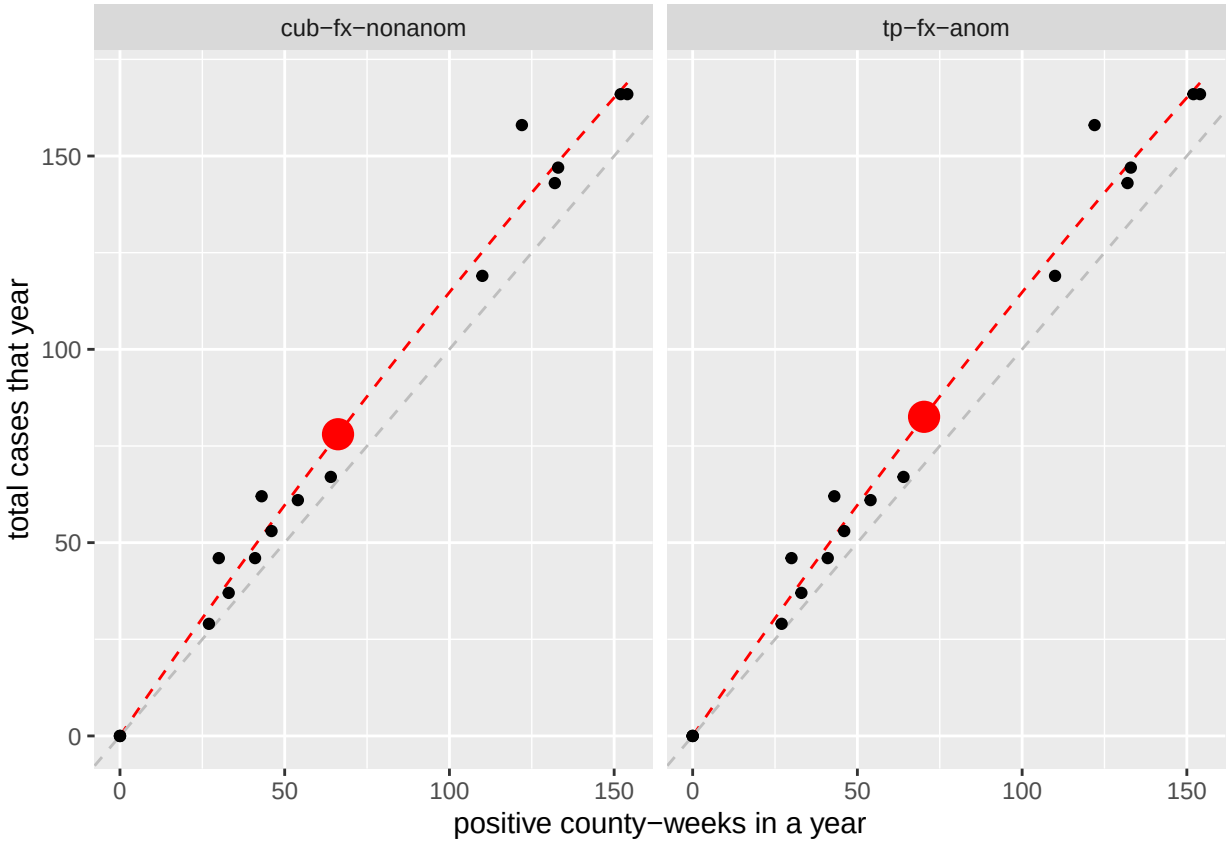


Model fit statistics are shown below.

	model	aic	auc
2	tp-fx-anom	6938.60	0.84
1	cub-fx-nonanom	6942.97	0.84

Estimates for the week in question are shown below.

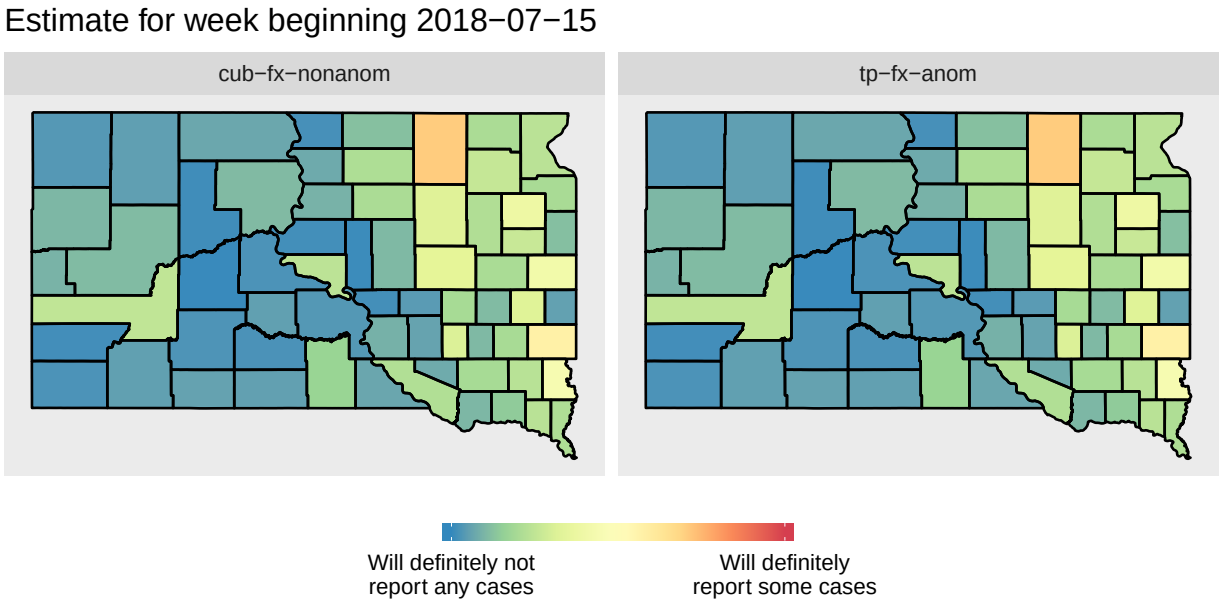
model	prop. positive	districts positive
cub-fx-nonanom	1.8%	1.2
tp-fx-anom	1%	0.7



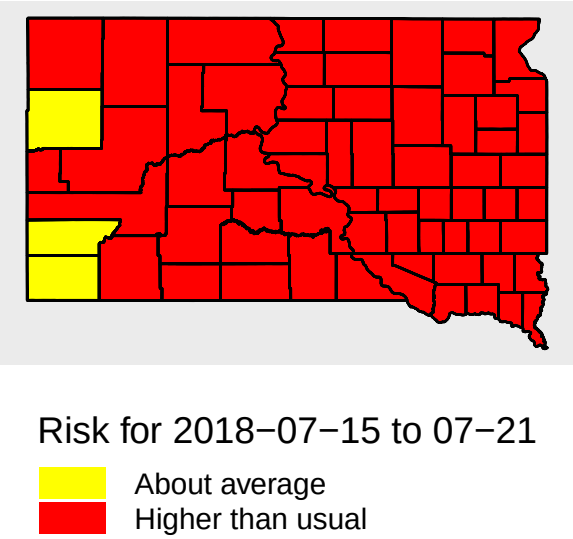
model	est. positives	total est. cases
cub-fx-nonanom	66.2	78.0
tp-fx-anom	70.2	82.6

Results for 2018-07-15 to 07-21

We visualize the raw estimated risk for 2018-07-15 below. If a district is darkest blue, then we estimate that there should be no human cases reported for this district, during this week. If a district is brightest red, we are certain that there will be at least one human case reported for this district, during this week.

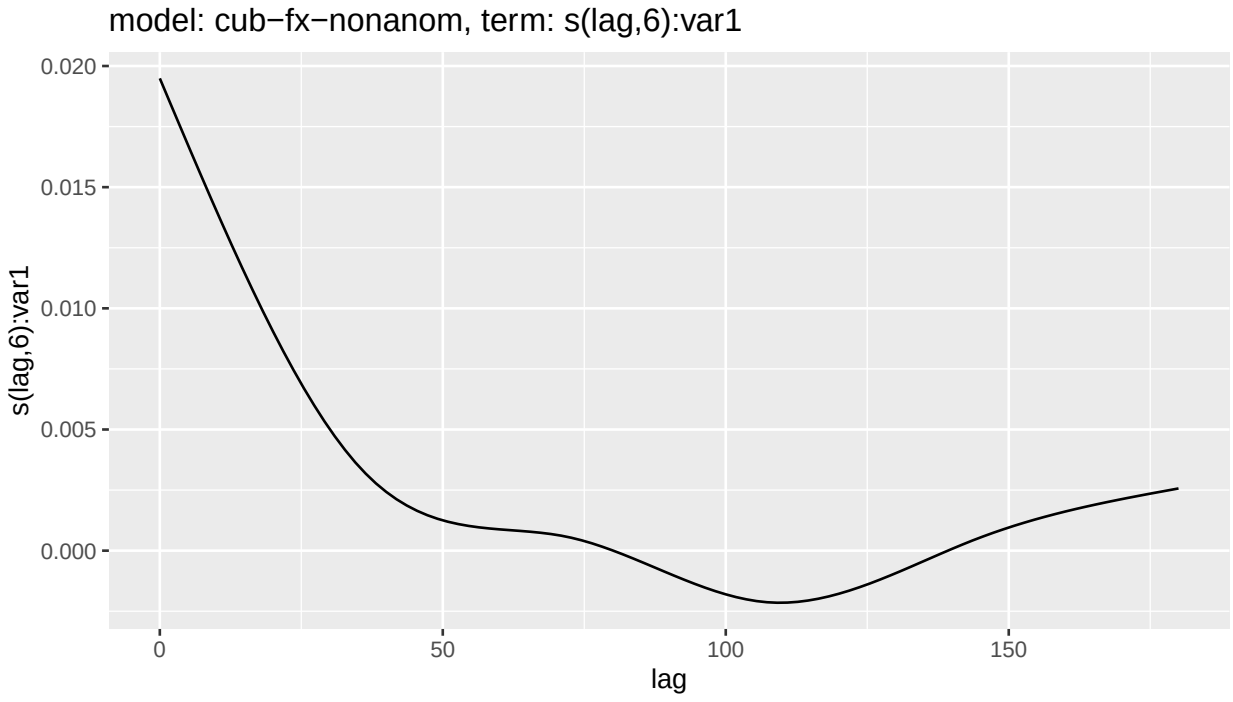


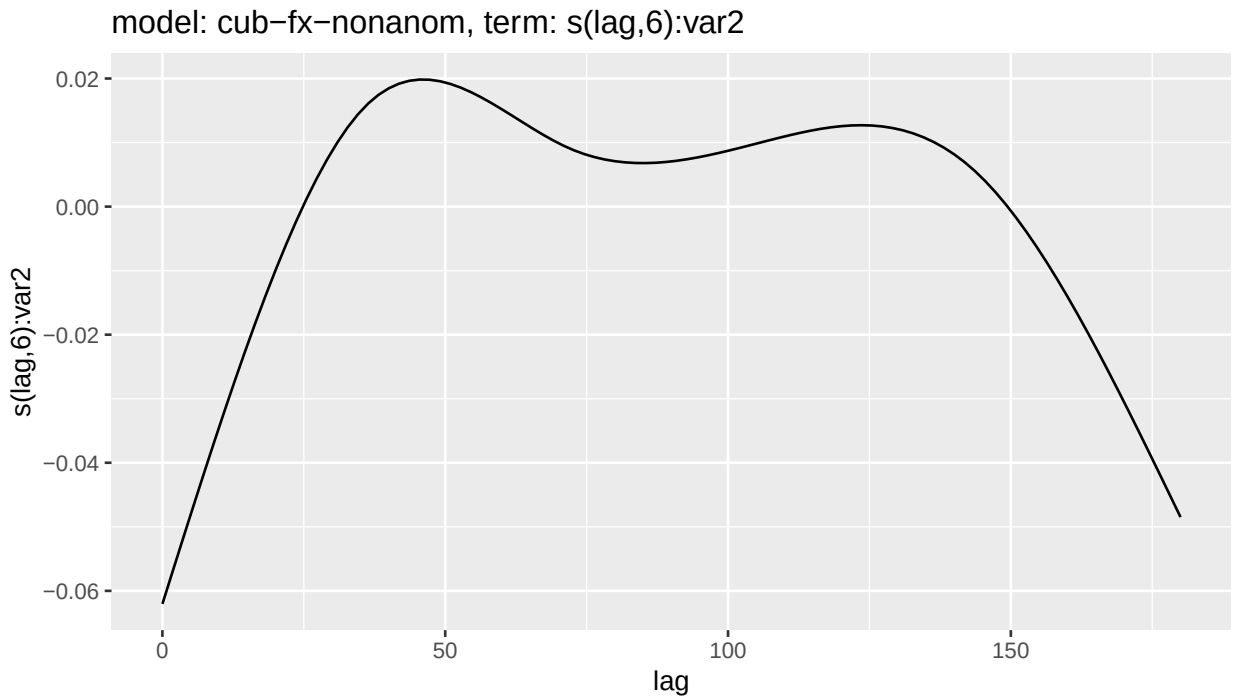
This map indicates whether probabilities reported in the previous map are higher (red) than average, lower (blue) than average, or right about normal (yellow) compared to the same week in previous years.

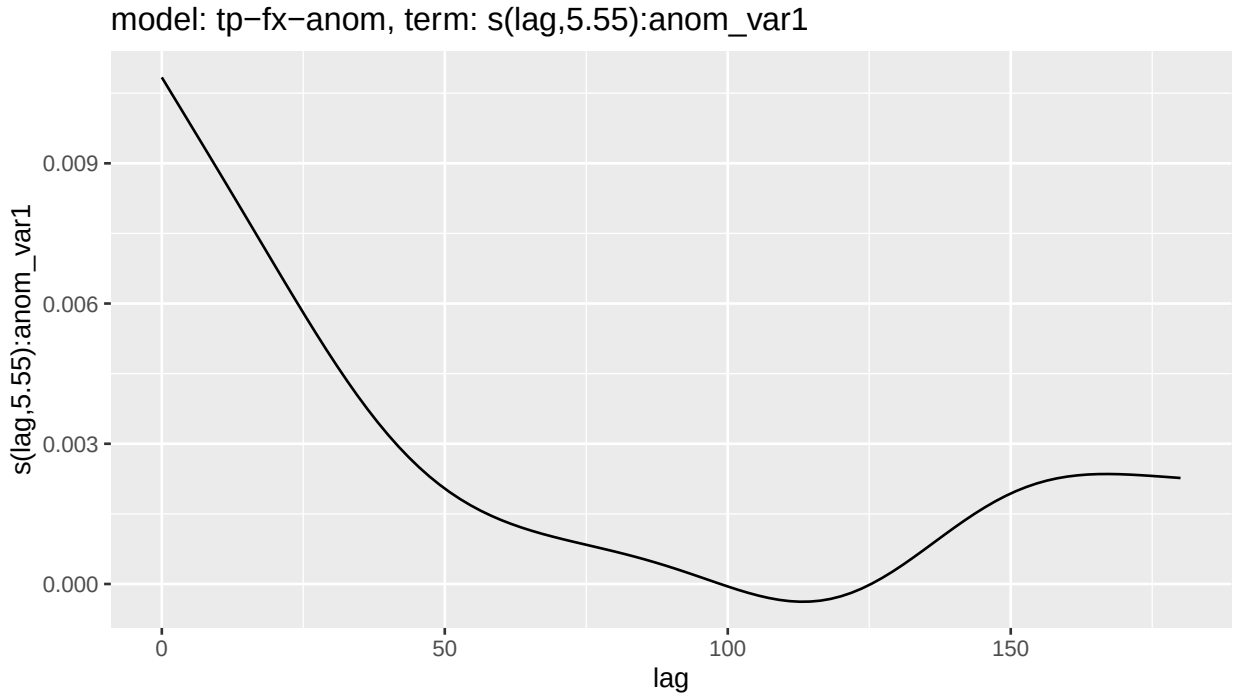


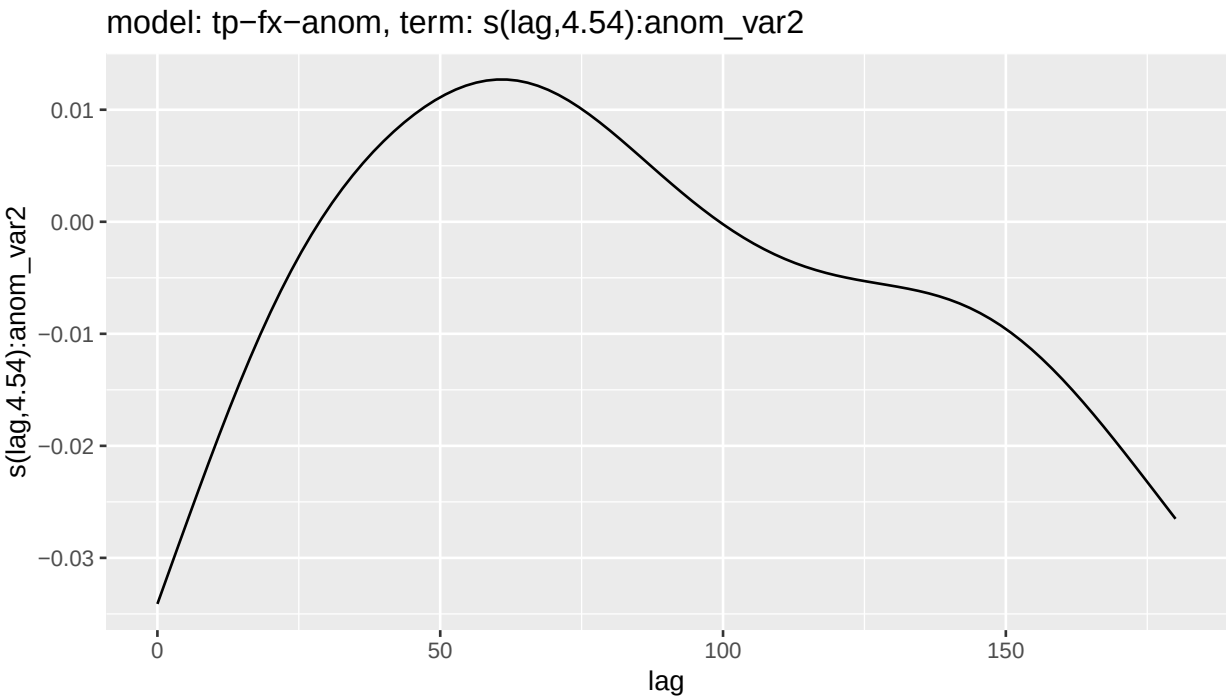
Estimated dependence functions

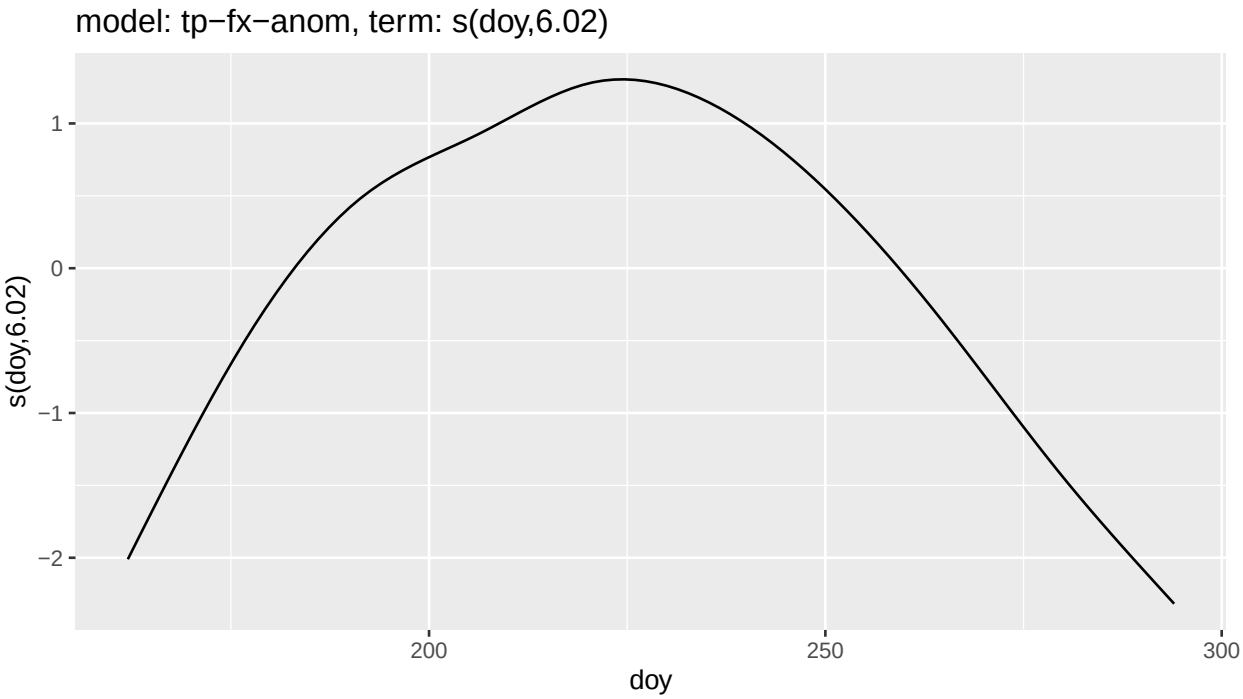
Because the user can write model formulas from scratch, we are not able to assume anything about the regression model. Therefore, plotting the model is difficult when you don't know what components the model has. In what follows, every 1-/2-D plot from the chosen models are shown. These will have names assigned by the gam function, which should be comprehensible if the user built the model.




model: cub−fx−nonanom, term: s(lag,6):var2
0.02
0.00
−0.02
−0.04
−0.06
s(lag,6):var2
0
50
100
150
lag


model: tp−fx−anom, term: s(lag,5.55):anom_var1
0.009
0.006
0.003
0.000
s(lag,5.55):anom_var1
0
50
100
150
lag


model: tp−fx−anom, term: s(lag,4.54):anom_var2
s(lag,4.54):anom_var2
0.01
0.00
−0.01
−0.02
−0.03
0
50
100
150
lag


model: tp−fx−anom, term: s(doy,6.02)
s(doy,6.02)
1
0
−1
−2
200
250
300
doy

District names

It is possible that district names in the human, mosquito, or weather data might disagree with the names found in the district shapefile. We have chosen the shapefile to unify all of the various data sources, so the human, mosquito, and weather data files should be updated to match these. Typically, check the TIGER shapefiles or census for standardized names.

There are 66 districts in the shapefile. This is the maximum number of districts ArboMAP will recognize in the human, mosquito, or weather data.

There are 67 districts in the human WNV data. If some districts never reported cases, this number may be less than the number of districts in the shapefile. If there are more, there are likely misspellings in the human data file. Human districts not found in the district shapefile: greg0ry.

There are 29 districts in the mosquito WNV data. If some districts never reported mosquito data, this number may be less than the number of districts in the shapefile. Mosquito districts not found in the district shapefile: none.

There are 66 districts in the mosquito WNV data. If these data were downloaded with the GEE app, they should match the shapefile districts exactly. Weather districts not found in the district shapefile: none.