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

User Guide Version (v1.0)
Justin K. Davis and Michael C. Wimberly
(justin.k.davis@sdstate.edu, michael.wimberly@sdstate.edu)
Geospatial Sciences Center of Excellence, South Dakota State University

Updated July 30, 2018

What you need to know about WNV and modeling

West Nile virus (WNV) circulates primarily among mosquitoes and birds, but occasionally spills over into human or other hosts. In humans, the majority (~80%) of infections show no symptoms and are not diagnosed except perhaps accidentally during blood donation. Some small proportion of individuals will develop symptoms such as headache, fever, and rash, but in about 1 out of every 100 infected individuals, the disease becomes neuroinvasive and can cause debilitating illness and death.

WNV was first introduced to the US in 1999 on the east coast, after which it spread in a westward-moving wave across the US. The early years produced thousands of cases as WNV invaded new regions. Birds were immunologically naive, so the virus decimated entire avian populations and spilled over into humans in an outbreak that lasted several years in some states.

In South Dakota, there were more than a thousand diagnosed cases in 2002-2003, which implies that many more were actually infected, since most infected individuals do not show symptoms and are never diagnosed. After a number of years of active transmission, case numbers declined and it looked like the disease might vanish from circulation after 2011, in which there were only two cases reported.

However, in 2012 there was another outbreak, with hundreds of cases in South Dakota. Because of these massive fluctuations in annual case counts, there is a need to predict the magnitude and locations of WNV in advance to support proactive mosquito control and disease prevention activities. Thus, a key scientific question is whether there are there any environmental or entomological indicators that could tell us, in advance, how bad a year is likely to be.

In fact, yes. Work by Dr. Michael Wimberly's research group has shown that a relatively simple statistical model, relying on weather and mosquito infection data, can be used to predict the risk of infection on a county-week basis. For more background, you will want to consult relevant journal articles, which are listed at the end of this document. In particular, the paper by Davis et al. (2018), published in Acta Tropica, provides details about the underlying statistical model.

The ArboMAP system was developed to implement this modeling approach via a straightforward workflow. The system uses the R programming language and software environment for statistical computing and graphics along with Google Earth Engine for accessing environmental data. ArboMAP was originally developed to support WNV forecasting in South Dakota, and this guide therefore references specific data formats that are unique to this state. However, it should be straightforward to modify data from other locations into a similar format in order to use them in ArboMAP.

The ArboMAP user guide consists of four main sections. The first section describes how to install the necessary software needed to run the system. The second section describes the steps that need to be taken every year before the beginning of the WNV season to prepare for forecasting. The third section description the steps that need to be taken each time a new forecast is generated. The final section describes the outputs produces by ArboMAP and explains how to interpret the forecast charts and maps.

Setting up for the initial run

Some accounts you'll need

To obtain mosquito infection data for South Dakota, you will need an account at **mosquito.sdstate.edu**. This account will need to be set up by someone at SDSU. If you work for a mosquito control district or another state that collects these data, then you will be able to modify your CSV files to match what we require here, and the ArboMAP system should run in the same way.

Environmental data will be obtained from the gridMET dataset and downloaded using Google Earth Engine (GEE). GEE is a cloud-based platform for processing satellite remote sensing images and other geospatial datasets. Instead of downloading the data and processing it yourself, GEE uses cloud-based computers and algorithms to do in half an hour what might take a standard desktop computer a week to calculate.

To access GEE, you will first need a gmail account at **mail.google.com**, so that you can access Google Drive documents at **drive.google.com**. Next, visit GEE at **earthengine.google.com** and click on "Sign Up" to request an account. You receive confirmation at your gmail address, and any downloads of weather data will be stored in the Google Drive of the account.

If you are associated with the South Dakota Department of Health (SDDOH), you should already have access to up-to-date human case data. If not, check with your group to determine whether there is a data-sharing agreement with the SDDOH. If you are implementing this workflow outside of South Dakota, then it will be up to the you to obtain local mosquito data and put it in the appropriate format.

Some software you'll need

R

R (**www.r-project.org**) is a statistical programming language that runs all of our analyses and produces reports and documentation, including this document. It is free and has a wide variety of packages built by users all around the world to do essentially any statistical analysis you can imagine.

It is easiest to download R from CRAN (**cloud.r-project.org**). Click on the link for "Download R for Windows." Choose "base," then "Download R for Windows." Run this file and install R on your system. Use the default settings for the installer.

RStudio

RStudio (**www.rstudio.com**) is a user-friendly GUI for the statistical programming language R that greatly simplifies a number of tasks for the programmer. Navigate to the site, click on Download, and choose the RStudio Desktop (Open Source License) - that is, the free version. Run the appropriate installer, very likely the Windows Vista/7/8/10 installer.

Run RStudio and close it at least once after you've installed MiKTeX as described in the next section. RStudio may not be able to find MiKTeX unless it's had a chance to search.

MiKTeX

MiKTeX (**www.miktex.org**) is an implementation of LaTeX, which is what allows us to automatically produce PDF reports at the end of all the calculations. Navigate to the site, click on Download -> Windows and then download and run the installer. Default options will suffice, but make one change: "Install missing packages on-the-fly" should be changed to "yes." You will initially install an incomplete copy of MiKTeX, and it will be updated automatically whenever you run the code. This means that the first time you run the code, it might take quite a while.

An alternative is to choose the Net Installer from the "All Downloads" page, and to choose "Complete Installation." This will take a great deal of time, and you should only do it on a fast connection, but it means you will only ever have to do this once. Expect that this will take a few hours - if you can afford the time to do this, it's worth getting all the downloads done. You will probably be asked to select a source/server/mirror from which to download your files - you can choose the 0-Cloud, which will automatically find a fast download site for you, or you can manually pick somewhere close to you physically.

Setting up the directory structure

ArboMAP requires a variety of files to run, and the following directory structure is the easiest way to organize everything that's necessary. Here we have a D: drive with a directory called ArboMAP - this directory contains everything necessary to run the model. Advanced users can change this directory structure by modifying ArboMAP's code.
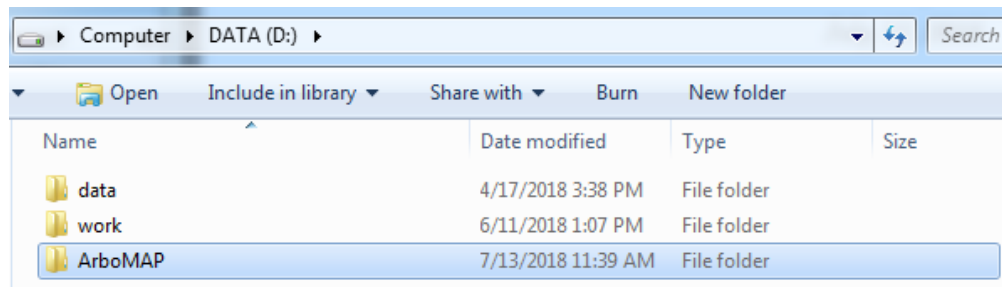
This ArboMAP folder should contain one Rmd file and seven subfolders, each containing either data used to run the model or outputs from the model.
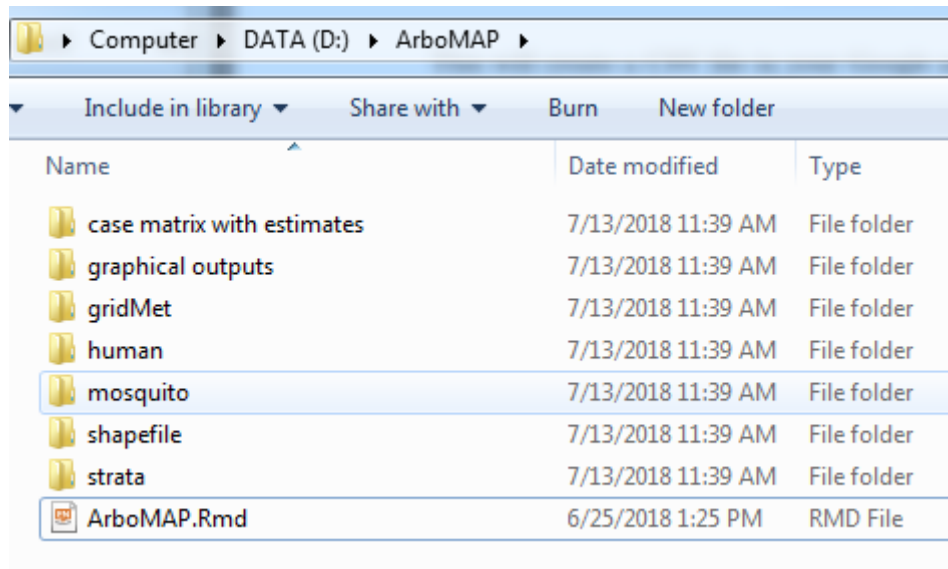
The folders are as follows:

- `case matrix with estimates` will hold a single file at the end of processing, which contains all the numerical data used for modeling and the estimates for each county and week. This can be used for additional statistical analyses if desired.
- `graphical outputs` will contain a number of PNG files at the end of processing. These images are the same images in the PDF output, but are provided so that the images can be used in other media if desired.

- **gridMet** contains a number of CSV files holding the weather data. This folder should only contain weather data, as any additional files placed here will cause errors in processing.
- **human** contains the historical human case data file.
- **mosquito** contains the mosquito pool infection data files.
- **shapefile** contains the files used to map risk. This is by default South Dakota, but these files may be replaced with any similar SHP file.
- **strata** contains a file that tells the model how to split the state into different geographic regions.

Finally, the `ArboMAP.Rmd` file contains all the modeling code. This file is opened in RStudio and pulls data from the directories above, produces estimates, and outputs a PDF file and a variety of other files in the output directories.

Files and settings

The Rmd file, which you will open in RStudio, contains a preface (shown below) with settings that tells the program where to find the various pieces of data used to make predictions. Errors in these filenames are a common reason that the code will refuse to run.

- **graphicoutputdir** tells the program where to store its graphical outputs. You may want to use some of the graphics in the PDF for your own purposes, so they will be stored in PNG format in this directory.
- **humandatafile** contains the human WNV cases
- **maxobservedhumandate** tells the program which is the last human case that should be used in modeling. If you happen to know that there are some cases in 2018 and you're predicting 2018, then you should set this to the end of 2017. We do not use the current year's data in making predictions, and trying to do so will mess up the predictions. This is an extra safeguard in case you do accidentally update the case file to include data from this year.
- **weekinquestion** tells the program which week you'd like to obtain predictions for. Weeks in this program begin on Sunday and run until the next Saturday, and the weekinquestion date will automatically be rounded to the previous Sunday. For example, if 2018-07-16 happens to be a Monday, then the week examined is 2018-07-15 (Sunday) through 2018-07-21 (Saturday).
- **gridmetpathstr** indicates the directory where all the gridmet CSV files are stored
- **gridmetsummaryfile** is the name of the file that the program will create to summarize all the CSV files. Since it's possible you've downloaded a year many times (especially the current year), the program has to create a summary file and stores it with the rest of the gridMet data. Other than this extra file and all the CSV files, there should be nothing else in this directory.
- **acttestfile** and **arctestfile** tell us where to find the active and archive mosquito testing data
- **stratafile** tells us where to find the file containing the data to split up the state into four regions to model the mosquito infection rate (see below)
- **countyshapefile** points to the SHP file for mapping SD counties
- **compyear1** and **compyear2** indicate which two years in the past we'd like to compare the current year to in one of the graphics.

`Maxobservedhumandate` should be updated at the beginning of every season to point to the end of last year, so that only human cases in previous years are used to condition the model's predictions. `Weekinquestion` should be updated weekly.

```
# where do we want the outputs?
graphicoutputdir <- "D:\\ArboMAP\\graphical outputs\\"
fullcasematoutputdir <- "D:\\ArboMAP\\case matrix with estimates\\"

# where are the human data located?
humandatafile <- "D:\\ArboMAP\\human\\17-11-16 - reconciled human WNV.csv"

# what is the date of the last human case we're willing to believe?
# probably, cut this off at the end of last year
# DO NOT use any human cases from the year you're modeling
maxobservedhumandate <- as.Date("2017-12-31", "%Y-%m-%d")

# which week are we producing the graphs for?
weekinquestion <- as.Date("2018-07-15", "%Y-%m-%d")

# where are the gridmet csv files stored?
gridmetpathstr <- "D:\\ArboMAP\\gridMet\\"
# what is the name of the summary file to be created?
gridmetsummaryfile <- "gridmet internal summary.csv"

# where are the mosquito test files located?
acttestfile <- "D:\\ArboMAP\\mosquito\\active_testing.csv"
arctestfile <- "D:\\ArboMAP\\mosquito\\archive_testing.csv"

# which county stratification scheme are we using?
stratafile <- "D:\\ArboMAP\\strata\\17-04-20 - classified strata - classic.csv"

# where is the countyshapefile
countyshapefile <- "D:\\ArboMAP\\shapefile\\cb_2014_us_county_5m - in EPSG 5070 - only SD.shp"

# to which two other years do we want to compare the current year's predictions?
compyear1 <- 2012
compyear2 <- 2017
```

Obtaining data at the beginning of the season

Obtaining the gridMET data

ArboMAP requires weather data to make its predictions. You will use Google Earth Engine (GEE) to download historical weather data beginning a full year before your human case data (e.g. if you begin modeling in 2001, you will need to download all data beginning in 2000). If you want to download all data, it will not hurt the calculations; simply begin in 1950). You will only need to download historical data once; every other time you run the program, you will only need to download the most recent year's weather data from GEE.

You will be given a link to code you can use to download the weather data. Once you click on this link and sign in to the Google account you used to sign up for Google Earth Engine, you should see a page like the following:
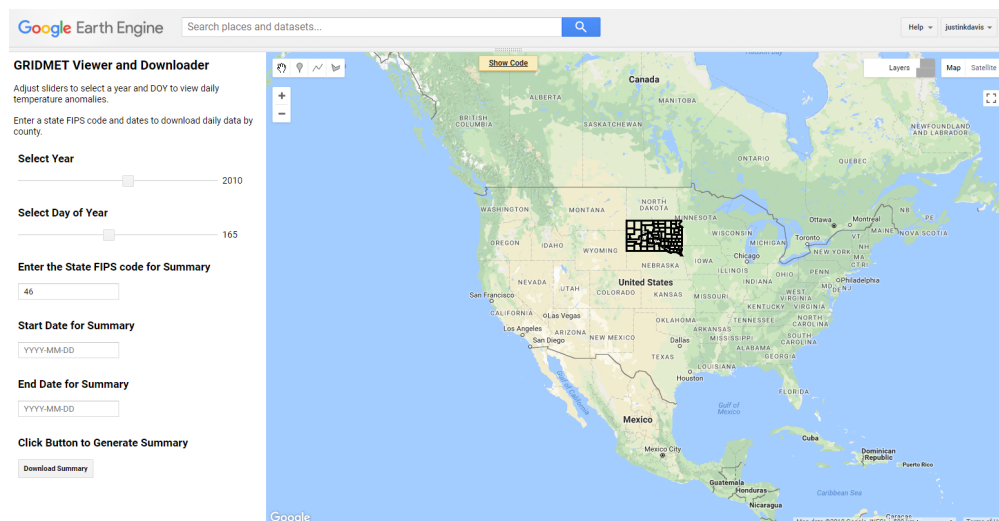
Click on the textured bar above "Show Code" and drag down to view the code. You don't need to modify

any of the code, but this will allow you to see the Tasks tab, which will allow you to download the data. Set the "Start Date for Summary" to YYYY-01-01, where YYYY is the year of data you wish to begin in, and set the "End Date for Summary" to some point in the future. You can set the End Date to 2050-01-01, and you will be given the most up-to-date weather data.

**Enter the State FIPS code for Summary**

46

Start Date for Summary

2000-01-01

End Date for Summary

2050-01-01

Click Button to Generate Summary

Download Summary

You can change some of the other options, if you'd like to modify your download. If you're not using the data for South Dakota, for example, which has FIPS code 46, you can choose another state from the list of codes (**click here for codes**).

Click on "Download Summary." This will create a task called myExportTable Task under the Tasks tab. Click on Run. If you do see this or any of the code, click on the textured bar directly above the map and drag down. This should reveal the code windows.
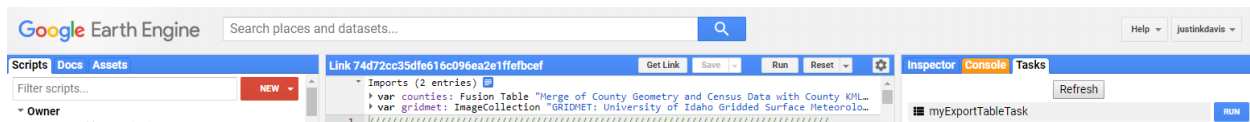
Name the task "historicaldownload" and name the file "historicaldownload" - note that there are no spaces. Click on "Run" in this window in the Tasks tab. If the Tasks tab is not visible, remember to pull down the bar directly above the map.
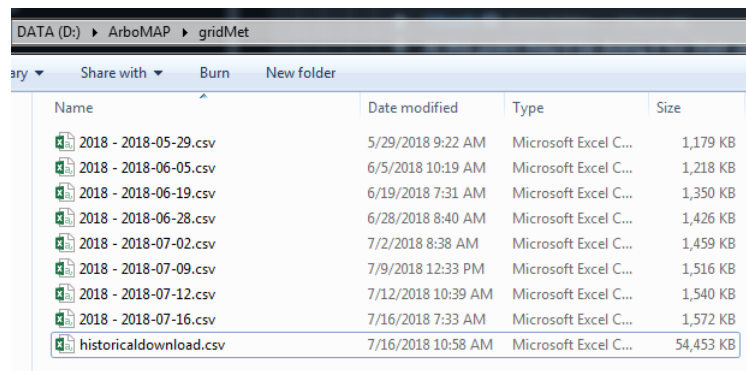
Processing can take half an hour for this initial data pull, depending on who else is using the system, and very rarely your job can be cancelled - you might receive an error about "too many objects," in which case you should try again later or split the download into individual years. Once it's done, there will be an extra

file in your Google account's drive at **drive.google.com** called "historicaldownload.csv".

Download this file (right-click -> download) and save it in a directory with all the other gridMet data; by default this is "D:/ArboMAP/gridMet/". You will eventually add updates for weather data in the current year (2018 at the time of writing), as discussed below, and these are the files beginning with "2018 - ". These updates are stored as extra CSV files and, along with the historical download, give you a complete picture of the daily weather in every county.

Note that it doesn't matter how many times you download a date's weather data, as long as it appears at least once in the data. Every single one of these 2018 files contains weather data for 2018-01-01, for example. Additional measurements will not affect the predictions, and any data that are represented more than once are simplified before inclusion into the model.

Obtaining the human case data

You will only need to update the human case data file once at the beginning of each season. For example, to make WNV forecasts for 2018, the model is calibrated using human case data through the end of 2017, and 2018 case data are not incorporated until data are finalized after the end of the 2018 WNV season. Only two pieces of information are needed for every human case: a date and a county. The date is the symptom onset date (in the case of clinical cases) or the date of blood donation (in the case of viremic blood donors). The county is the patient's county of residence. Each row represents a single case.

The CSV file you use should that contains the human case data should have a unique column named "creationdate" that is formatted MM/DD/YYYY, and a column named "county" that contains standard

names for SD counties. Here, any of the following are acceptable: BROOKINGS, Brookings, brookings, Brookings County, BROOKINGS COUNTY, and they will all be reduced to "brookings" in the output.

It does not matter which other columns are present, as long as each row contains at least these two values. Cases without either of these values will not be modeled. Below is an example of an acceptable (fake) data file. Note that other columns are present, but they are not used by ArboMAP.

caseid	strippedcaseid	creationdate	county	zip	age	disease status	infectiontype
125662233	125662233	5/25/2017	Davison	90210	31	CONFIRMED	FEVER
1290875908	1290875908	6/21/2012	Brookings	12345	27	CONFIRMED	FEVER

Save this file in its own directory. By default this is "D:/ArboMAP/human/human case data.csv".

Do not include any human case data from the year you are modeling. If you're making predictions for 2018, your model should be based on all human case data up to December 2017. If you include any human case data for 2018 - for example, if you know that there were a handful of cases already in mid-July 2018 and update the human case data file during the season - then the model will assume that human case data for 2018 are complete. This is almost certainly incorrect, since some cases are only reported weeks or months after diagnosis, and human case counts during the season will therefore underestimate the actual disease burden.

Obtaining the mosquito data

South Dakota users will be given a login for **mosquito.sdstate.edu** by collaborators at SDSU. Log in and navigate to Mosquito Data. There are two databases to download - the active data view, which contains tested pools from the current year, and the archive data view, which contains tested pools from prior years.

To download data from the current year, click on "Testing Records" under Active Data Views and scroll to the bottom of the page. There will be a small orange button labeled "CSV." Click on this to download the mosquito infection data for this year. Save this file in the directory dedicated to the mosquito infection data with a name that indicates it is the active view; for example "D:/ArboMAP/mosquito/active_testing.csv".

	A	B	C	D	E	F	G	H	I	J
1	Title	VR	Collection Date	Species	Number in Pool	WNV Result	Collect Ref.	Edit	County	Trap
2	mark_hoven_2649	2649	6/4/2018	Culex tarsalis	21	Negative	1306-04-1812	Edit	Brown	Cemetery BG
3	mark_hoven_2648	2648	5/31/2018	Culex tarsalis	24	Negative	1305-31-1812	Edit	Brown	Cemetery BG
4	mark_hoven_2646	2646	5/30/2018	Culex tarsalis	45	Negative	1305-30-1812	Edit	Brown	Cemetery BG
5	mark_hoven_2647	2647	5/30/2018	Culex tarsalis	14	Negative	1305-30-184	Edit	Brown	Wylie Park
6	mark_hoven_2650	2650	6/5/2018	Culex tarsalis	32	Negative	1306-05-1812	Edit	Brown	Cemetery BG
7	chris_carlson_4001	4001	5/30/2018	Culex tarsalis	6	Negative	11705-30-180	Edit	Stanley	Larson's House - BG

Note that this file must include a column named "Collection Date" in MM/DD/YYYY format, a "Species" column, a "WNV Result" column coded as Positive or Negative, a "Trap" location, and a "County" column. "Trap" is needed for backwards compatibility with earlier versions of the database.

To download data from previous years, click on "Testing Records" under the Archive Data Views and scroll to the bottom of the page to find the CSV button. This can take a little longer to compile since it includes all past mosquito infection data. Save this file as "D:/ArboMAP/mosquito/archive_testing.csv".

	A	B	C	D	E	F	G	H	I	J	K
1	Title	VR	WNV Result	Number in Pool	Collection Record	Trap	County	Collection Date	Nid	Species Old	
2	archive_6545	8609	1	24		Zoo - # 27	Minnehaha	7/28/2006	1007111	Culex restuans	not found
3	archive_6623	8654	1	28		Zoo - # 27	Minnehaha	8/4/2006	1007116	Culex pipiens	not found
4	archive_6651	8656	1	7		Zoo - # 27	Minnehaha	8/8/2006	1007121	Culex pipiens	not found
5	archive_11567	10483	0	1		YMCA	Minnehaha	8/27/2009	1007126	Culex tarsalis	not found
6	archive_11581	10485	0	1		YMCA	Minnehaha	8/28/2009	1007131	Culex tarsalis	not found
7	archive_13814	17869	0	1		Yankton		7/27/2011	1007136	Culex tarsalis	not found

This file must include a column named "WNV Result," which is either 0, which means a Negative pool, or 1, which means a Positive pool. There must be "County," "Trap," and "Collection Date" columns in MM/DD/YYYY format. There must be a "Species.Old" column. We only model mosquitoes labeled "Culex tarsalis".

If you are downloading these files from the mosquito.sdstate.edu website rather than creating them on your own, these should be simple downloads. However, it is possible that something has changed in the format of the download, and we recommend that you check the formats of the files.

A note on standardized place names

ArboMAP was designed to model WNV on a county-week basis, but can work with any geographical unit as long as each area is assigned a unique, stable identifier. Ideally, you would like to work with something like the FIPS code, but here we've chosen to work with county names as that is how the data are usually delivered. This presents some difficulties, since these names are probably not completely standardized in your sources; e.g. some sources will list "Brookings County" while others will simply use "Brookings".

ArboMAP reduces all names by deleting "County" and "Parish", removing all spaces from the county name, and putting the county name in lowercase. Therefore, "BROOKINGS COUNTY" becomes "brookings" and "Charles Mix County" becomes "charlesmix". This gives the greatest chance that all places will be recognized in all data sources. However, incorrect spellings and invalid date formats (e.g. DD/MM/YYYY instead of MM/DD/YYYY) cannot be fixed by the program and will result in errors.

Note that in 2015, Shannon County, SD became Oglala Lakota County. There are similar changes in other states. You are welcome to use either name in the human case data file, but the output will use "shannon" for backwards compatibility. If the output is to be distributed to the public, this should be changed.

The typical run during the season

Updating the data during the season

Before you run the model during the WNV season, you will want to be sure you have the most up-to-date weather and mosquito data. These updates are simpler than the updates that need to be run once per year at the beginning of the WNV season, but will need to be performed every time you want to make an updated prediction.

Updating the gridMET weather data

Follow exactly the process described above, but only download data for the current year. For example, if you are modeling in 2018, set the start and end dates to 2018-01-01 and 2019-01-01. This will include any gridMET data available for 2018. Name the task "2018" and name the file "2018 - 2018-07-16" for example, if you downloaded the data on 2018-07-16. Save this file in the gridMet directory with the historicaldownload.csv file and the other updates to 2018. Eventually, you will have a large number of these update files along with the original "historicaldownload.csv" file that contained all available past data.

Updating the human case data

Do not update the human case data during the season. If you learn that there was a case last week, for example, this should not yet be included as a new line in the "human case data.csv" file. Next year, when the historical human case data are updated, this new human case will be included in the file, but during a year the "human case data.csv" file should remain untouched.

Updating the mosquito data

Use the process described above to log in to the SDMIS website. This time, however, only download the active view, and replace "active_testing.csv" with the new file. During the season, the "archive_testing.csv" file will not change, since the historical data should not be revised.

Setting up the Rmd file

You will be given a file called "ArboMAP.Rmd", which should be placed in "D:/ArboMAP/". This is an Rmarkdown file, which does all of the calculations of modeling and produces a PDF in the end to summarize the outputs. Open RStudio and open this file.

Running the code

Open the Rmd file in RStudio, change the weekinquestion, and press ctrl+shift+K. This will tell RStudio to compile the PDF. It should not take very long - on a slow laptop, modeling takes around 10 minutes at most. If it takes substantially longer than this, it is likely that there is an error somewhere in the data. At the end, you will see a PDF in the same directory as the Rmd file, and a number of graphics will be stored in the directory you requested.

Interpreting the output

Weather data

Each forecasting run will generate a report containing a number of graphics to interpret; an example from 2018 is below. First, statewide mean daily temperature in the current year (red) is compared to the median (black) and max/min (grey) temperatures for all historical gridMet data that you are currently using. Higher temperatures are associated with more WNV, so red above black typically means more human cases. Higher temperature, more WNV.

Temperature in 2018
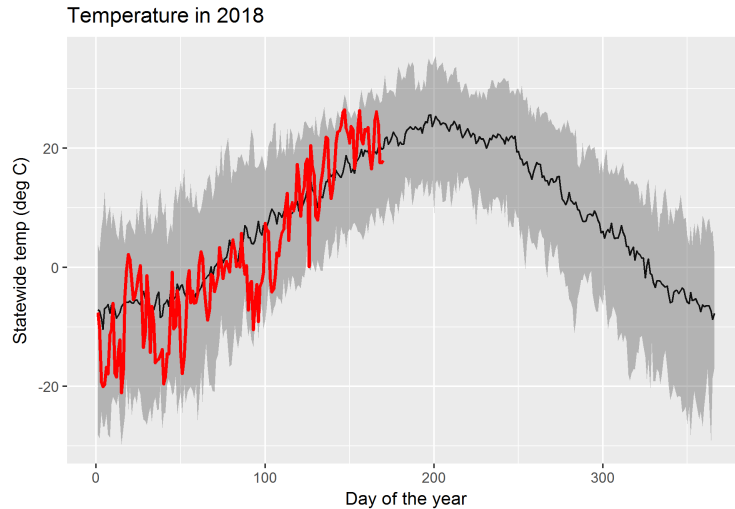
The model of WNV for South Dakota also uses the vapor pressure deficit; an example from 2018 is below. This is the force exerted by the atmosphere to extract moisture from something - the higher the deficit, the drier the air. Usually, if the atmosphere is more humid (lower deficit), there is more WNV. Lower deficit, more WNV.

Vapor pressure deficit in 2018 (higher means drier)
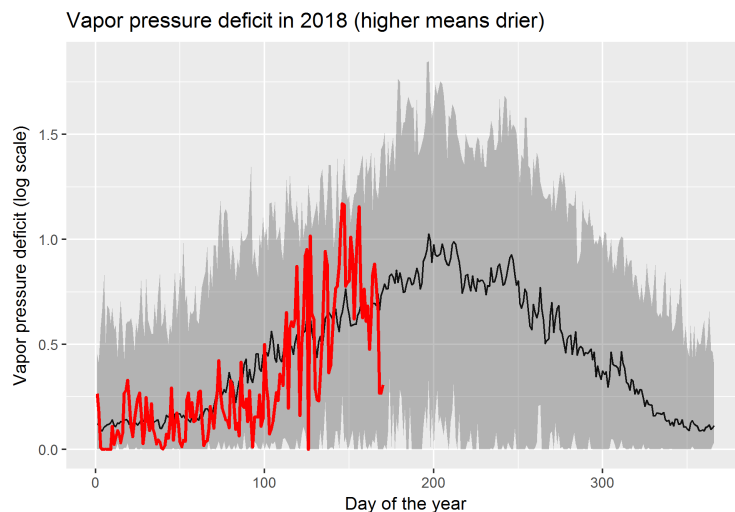
Mosquito data

The model of WNV also summarizes the mosquito infection growth rate (MIGR) for multiple strata within the state. Every winter, the virus goes into hiding. In the early season, WNV begins replicating and spreading among birds and mosquitoes. The MIGR is a measure of how quickly that's occuring. We cannot estimate the MIGR for every county, but instead split the state into four strata, shown below, and estimate the MIGR within the strata. If the map does not resemble this, especially if the map is entirely grey, it's likely that the program cannot locate the strata file.

State stratification map
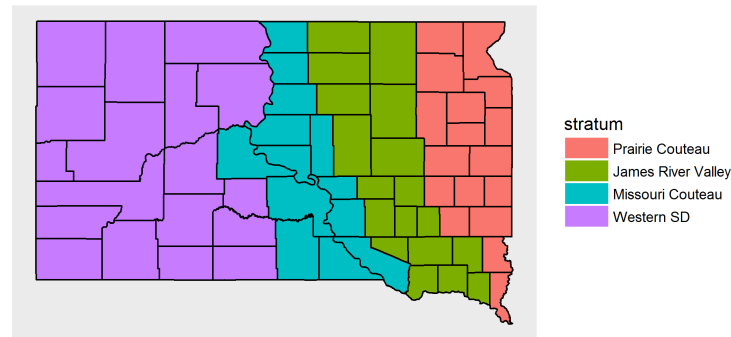
Below we show the estimates of the MIGR per year, per stratum. If the line is above 0, then mosquito infections are growing in pools more quickly than in the average year, and more human WNV should be expected. If lower, then risk to humans is lower. Note that the state does fluctuate up and down as a whole (e.g. in 2011 when barely any positive mosquito pools were found, and only two human cases were diagnosed), but sometimes the strata do differ (e.g. the MIGR in western SD was relatively low in 2014, but was relatively high in 2015).


Note that the MIGR is a single number, derived from a model that fits all of a year's mosquito infection data. Below, we show observations from 2018 (red dots), along with the curve that best fits these data (red), and samples from two other years, 2012 (blue) and 2017 (blue, dashed). The faster this line rises, the higher the MIGR, and the higher the risk.
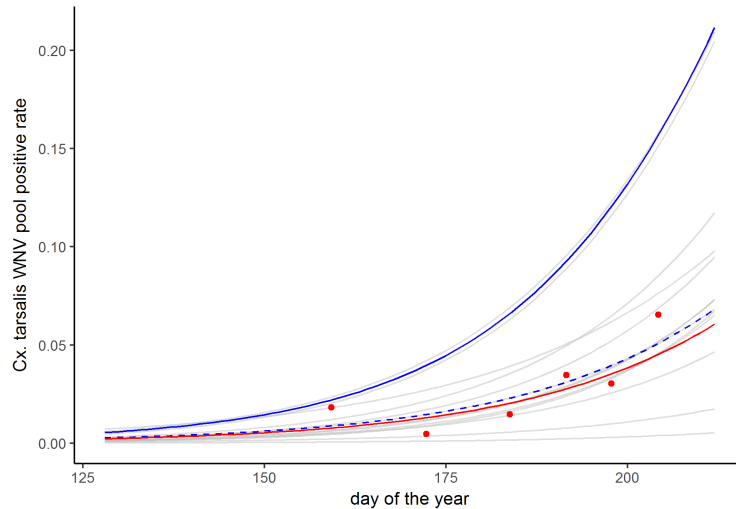
Multi-year WNV forecasting chart

Next are the model outputs. For each week of a year, the model estimates what proportion of counties will report at least one human case based on fit to historical data. This is displayed first on a statewide basis - red are estimates and black are observations. For the year in which you're making predictions, observed (black) should be constantly zero. Red should fit historical observations reasonably well - if there are major deviations, then it is likely that some data, probably the human or mosquito data, are in the wrong format.

Statewide model predictions
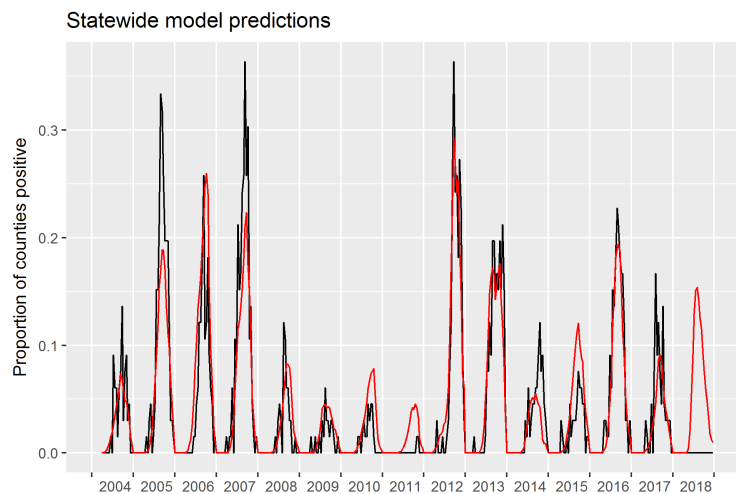
Current-year WNV forecasting chart

Next are model outputs for the year you're estimating and the two years you've selected for comparison. We typically choose 2012, an outbreak year, and the year one before you're estimating (e.g. if you're estimating 2018, then we will also display 2017). Predictions for the forecasting week are in red, and the week in question begins at the red dot.
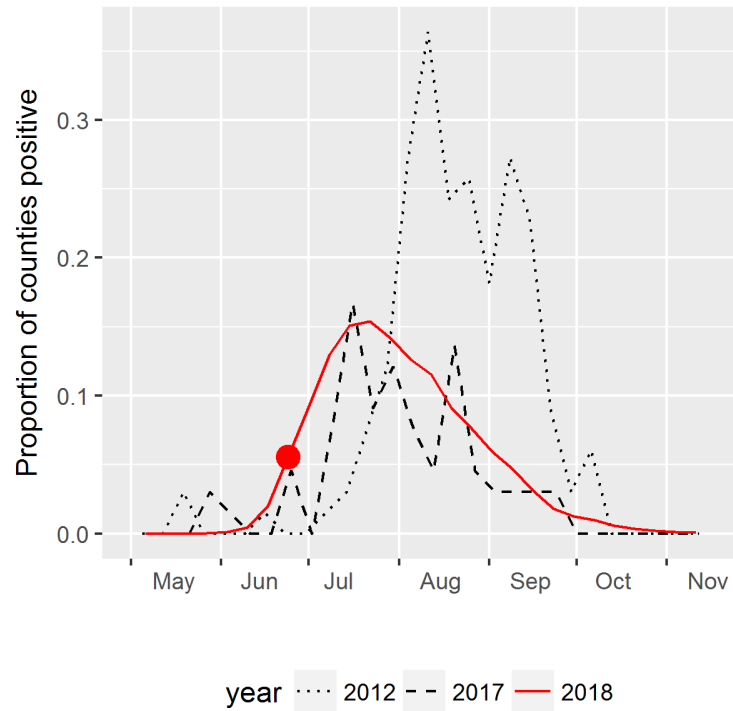
Current-week WNV absolute risk map

Next are the absolute probabilities of each county reporting at least one human case during the week in question. If the county is darkest blue, then the model says there is no probability of reporting a case during the week you've selected. This is often the case in the early or late seasons, especially in the lower-population counties. If a county is brightest red, then the model says the county will definitely report at least one human case this week.

Estimate for week beginning 2018-06-24

14
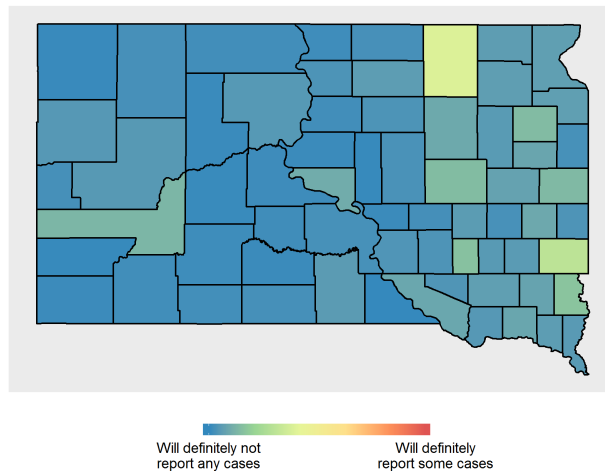
Current-week WNV relative risk map

Next are the relative probabilities. A county in the early season may have low risk because the virus is not yet circulating, but is that lower or higher than usual, for that county, during that week of the year? Brookings County on May 15th 2018 will have low risk, but that risk may be higher than usual when compared to Brookings on May 15th in 2004-2017? If a county is red, this means that risk is predicted to be substantially higher than average in that county, during the week in question. If blue, then lower. If yellow, then average. It is important to remember that red and blue do not mean high or low absolute risk. They mean high*er* or low*er than usual*. A county that is lower than usual can still have many cases in a week; e.g. Brown County in early August can be lower than usual, but will still likely report cases.
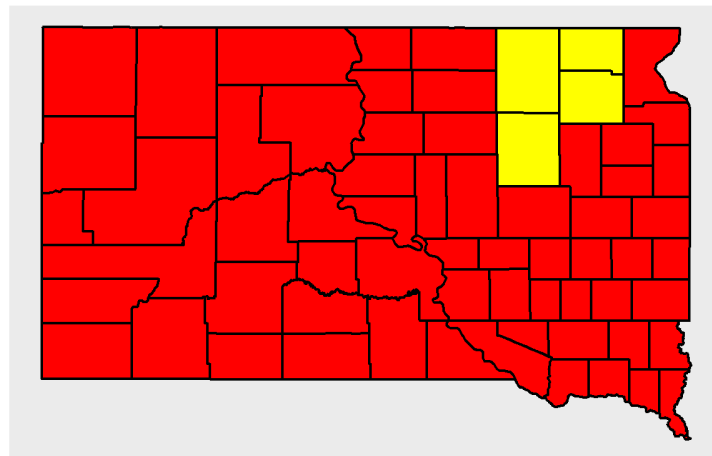
Risk for 2018-06-24 to 06-30

About average
Higher than usual
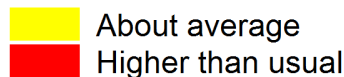
Combining the two maps above, we conclude that it's still early in the season and absolute risk is low, but there are indications that risk is higher than usual in most counties. You will want to interpret the temperature, humidity, and mosquito infection charts to understand exactly why this is the case.

Troubleshooting

Filenames and unrecognized escapes

The most common error will be an error in filenames. If you are told that a file is not found, it is likely that you have not correctly told the Rmd file where to look. Another common error with filenames concerns the slashes: if you see something like the following, then it is because you have used a single \, which is not acceptable, instead of \\ or / in your filename. Note that we have written "d:\work" rather than "d:\\work" or "d:/work".

```
D:/work/git/WNV/WNV/WNV_modeling_-_main.Rmd
 Line 38 Error: '\w' is an unrecognized escape in character string starting ""D:\w" Execution halted
```

```
# where are the human data located?
humandatafile <- "D:\work\\git\\WNV\\packaged WNV\\human\\17-11-16 - reconciled human WNV.csv"
```

No estimates, blank maps

Unfortunately, if the Rmd file refuses to produce estimates (e.g. the red line is missing in the graphs or the maps are completely blank) then it's likely that a file is missing, is in an improper format, or there are extra files where they do not belong; e.g. only gridMet files should go in the gridMet data directory.

Bad estimates

This is usually because of bad mosquito infection data. You will want to carefully check the mosquito infection data to ensure they are in the correct format. You will often be able to see this on the graph of the mosquito infection growth rate - a year in the past will dip down to very low levels of infection, but this does not seem historically accurate. Another problem can occur if you enter county names incorrectly. If the data tell the program that a case occurs in a county that does not exist, then this case will not be counted, and a county's risk will be artificially low. If this happens with enough counties (e.g. if every county is followed by a space in the data file) then a whole year's risk will be 0.

Relevant scientific papers

- Davis, J. K., G. P. Vincent, M. B. Hildreth, L. Kightlinger, and M. C. Wimberly. 2018. Improving the prediction of arbovirus outbreaks: a comparison of climate-driven models for West Nile virus in an endemic region of the United States. Acta Tropica 185: 242-250.
- Davis J. K., Vincent G. P., Hildreth M. B., Kightlinger L., Carlson C., and M. C. Wimberly. 2017. Integrating Environmental Monitoring and Mosquito Surveillance to Predict Vector-borne Disease: Prospective Forecasts of a West Nile Virus Outbreak. PLoS Currents Outbreaks. 2017 May 23. Edition 1. doi: 10.1371/currents.outbreaks.90e80717c4e67e1a830f17feeaaf85de.
- Wimberly, M. C., A. Lamsal, P. Giacomo, and T. Chuang. 2014. Regional variation of climatic influences on West Nile virus outbreaks in the United States. American Journal of Tropical Medicine and Hygiene 91: 677-684.
- Wimberly, M. C., P. Giacomo, L. Kightlinger, and M. B. Hildreth. 2013. Spatio-temporal epidemiology of human West Nile virus disease in South Dakota. International Journal of Environmental Research and Public Health 10: 5584-5602.
- Chuang, T., C. W. Hockett, L. Kightlinger, and M. C. Wimberly. 2012. Landscape-level spatial patterns of West Nile virus risk in the northern Great Plains. American Journal of Tropical Medicine and Hygiene 86: 724-731.
- Chuang, T., G. M Henebry, J. S. Kimball, D.L. VanRoekel-Patton, M. B. Hildreth, and M. C. Wimberly. 2012. Satellite microwave remote sensing for environment modeling of mosquito population dynamics. Remote Sensing of Environment 125: 147-156.
- Chuang T., and M. C. Wimberly. 2012. Remote Sensing of Climatic Anomalies and West Nile Virus
- Chuang, T. M. B. Hildreth, M. B., D. L. VanRoekel, and M. C. Wimberly. 2011. Weather and land cover influences on mosquito populations in Sioux Falls, South Dakota. Journal of Medical Entomology 48: 669-679.
- Wey, C. L., J. Griesse, L. Kightlinger, and M. C. Wimberly. 2009. Geographic variability in geocoding success for West Nile virus cases in South Dakota, USA. Health & Place 15: 1108-1114.
- Wimberly, M. C., M. B. Hildreth, S. P. Boyte, E. Lindquist, and L. Kightlinger. 2008. Ecological niche of the 2003 West Nile virus epidemic in the northern Great Plains of the United States. PLoS One 3: e3744.