West Nile virus forecast model submission form
Email completed form to vbd-predict@cdc.gov

Team name
University of Illinois

Team leader

Name	Institution	Email
Johnny Uelmen	University of Illinois	uelmen@illinois.edu

Other team members

Name	Institution	Email
Rebecca Smith	University of Illinois	rlsdvm@illinois.edu
Bo Li	University of Illinois	libo@illinois.edu
Bill Brown	University of Illinois	wmbrown@illinois.edu
Chris Stone	University of Illinois	cstone@illinois.edu
Liudmila Sergeevna Mainzer	University of Illinois	lmainzer@illinois.edu
Weihao Ge	University of Illinois	weg2@illinois.edu
Shubham Rawlani	University of Illinois	rawlani3@illinois.edu
Wayne Wan	University of Illinois	guangya2@illinois.edu
Joshua Allen	University of Illinois	jallen17@illinois.edu

Model description
Provide a brief summary of the model methods with sufficient detail for another modeler to understand the approach being applied. If multiple models are used, describe each model and how they were combined.

The WNV forecasts are estimated using a zero-inflated Poisson Regression in the JMP statistical software (SAS Institute, Cary, NC, USA). The model first conducts a logistic regression to account for the majority of the counties with no neuroinvasive cases. It then runs a second model (Poisson) on the counties that do have at least n=1 neuroinvasive cases. The model can then output a prediction profile with an estimated zero-inflation parameter based on the two equations for a given county. The probabilities for each case bin (e.g. 1-6, 6-11, etc.) were estimated via difference in cumulative probabilities of the upper bound for each bin. Each county's point prediction and past year's probability distribution provided the mean and standard error, respectively, to calculate the normal distribution.

Our model will be fit for each county. In total, our models used 93 environmental, climatic, and socio-demographic factors specific to each county (see more info below). Climate data was generated from monthly ascii grids for the contiguous United States (PRISM Climate Group) from 1997-2018. Environmental data was provided from the Multi-Resolution Land Characteristics Consortium (MRLC), sourced from the newest available dataset for each respective category. The most recent data available for each socio-demographic characteristic were gathered from the American Fact Finder (United States Census Bureau). The temporal resolution varied from weekly to monthly averages for each county, and remained consistent in its use for each county.

Final models were selected via backward selection process, based on the Akaike Information Criteria (AIC) value.

Variables

List each variable used and its temporal relationship to the forecast. If multiple models are used, specify which enter into each model.

1. There are a total of 93 variables used in our analysis, but the main groupings of the variables are as follows:
2. Land Cover (several here – n=16 variables, including tree canopy, and % impervious); annual average were used per pixel for each county
3. Mean precipitation – weekly averages for each county
4. Mean temperature – weekly averages for each county
5. Max temperature – weekly averages for each county
6. Min temperature – weekly averages for each county
7. Average dewpoint – weekly averages for each county
8. Average age by census block for each county (based on 2010 Census)
9. Race and Ethnicity % by for each county
10. Average household income for each county

Computational resources

Describe the programming languages and software tools that were used to write and execute the forecasts.

Initial data cleaning and processing was conducted in JMP and Excel. Spatial analyses and extraction of data by county were processed in ArcMap 10.5. The statistical analyses were conducted in JMP and R.

Publications

Note whether the model was derived from previously published work and, if so, provide references.

These models were based on the methodology from previously submitted work (currently under review, Uelmen et al. 2020. However, we have changed from a generalized linear mixed model approach to a zero-inflated regression with a Poisson distribution, using Lasso estimation method.

Participation agreement

By submitting these forecasts, the team agrees to abide by the project rules and data use agreements.

Team lead name	Date
University of Illinois – Path	4/29/20