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.) will be based on the prediction from our generated equation using the upper bound value.

Our model will be fit for each county. Although not yet entirely finished in processing, explanatory variables comprised of >20 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.

Covariate inclusion criteria was conducted via forward selection process (initial inclusion criteria, $p \leq 0.2$). Final models were selected by the Bayesian Information Criteria (BIC) value. Lastly, locations that

have never had a WNV neuroinvasive case in the past were denoted values of 0.001 for the annual total.

The total number of included covariates and data sources used may change before the final models are submitted by July 31. Due to other commitments, and awaiting preliminary 2019 data, I am holding off on conducting these analyses for at least another month.

Variables

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

1. Land Cover (several here – n=16 variables, including tree canopy, and % impervious); annual average were used per pixel for each county
2. Mean precipitation – weekly averages for each county
3. Mean temperature – weekly averages for each county
4. Max temperature – weekly averages for each county
5. Min temperature – weekly averages for each county
6. Average dewpoint – weekly averages for each county
7. Average age by census block for each county (based on 2010 Census)
8. Race and Ethnicity % by for each county
9. Average household income for each county
10.

Computational resources

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

A large component of these models were processed in ArcMap 10.5. The statistical analyses were conducted in JMP, SAS, and R (each program has different analysis that we prefer to use on that respective platform). Initial data organization and storing was in Excel.

Publications

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

This model is a product of ongoing research in our lab on West Nile Virus in the Chicago, IL area. Our primary work focuses on a 2-county area and for the purpose of the WNV predict challenge, our existing model has been modified to incorporate predictors that accommodate for nationwide predictions.

Participation agreement

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

Date

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