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

Team name

Keyel_etal_RandomForest

Team leader

Name	Institution	Email
Alexander Keyel	NYS Wadsworth Center	akeyel@albany.edu

Other team members

Name	Institution	Email
Oliver Elison Timm	University at Albany	oelisontimm@albany.edu
Laura Kramer	NYS Wadsworth Center	laura.kramer@health.ny.gov

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.

We used a Random-Forest-based approach to predicting WNV, modified from [1]. A Random Forest [2] was run using all input predictor variables. A mean importance score [2] was then calculated. All variables below the mean importance were dropped, and the model was refit. The variance partitioning approach in [1] was excluded due to processing constraints. Only data from the months leading up to the desired forecast date were used. Weather data from January to March were aggregated, and used for the initial forecast. Weather data and soil moisture data were then added as monthly variables for remaining forecasts adding data for April, May, and June, for the May, June and July forecasts. Soil moisture was added beginning in April as a previous study [1] found spring and summer soil moisture to be relevant. A second modification is that a quantile Random Forest approach [3,4] was used instead of the original Random Forest method [2,5] in order to produce probabilistic forecasts.

Variables

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

1. Mosquito Ranges for Cx. pipiens, Cx. tarsalis, Cx. quinquefasciatus, Cx. salinarius, Cx. nigripalpus; no temporal variation, digitized from: [6]
2. Air Conditioning: total number of homes, total homes with AC, percent of homes with AC by broad US region; no temporal variation included, [7]
3. Census Data: estimated population by year; temporal variation by year, provided by A. Tyre, pers. comm

4. GRIDMET weather data: Minimum, mean, and maximum temperature, log precipitation, relative humidity, and vapor pressure deficit; compiled to 3-month composite for Jan - Mar, monthly thereafter. Anomalies from the 1999 – 2020 baseline were also included [8–10]
5. NLDAS NOAH soil moisture; monthly, beginning in April [11] (not included for April 30 forecast, as the monthly aggregate is not available until early May)
6. County case information: Total county cases, an indicator if a county has had greater than 100 cases, an indicator if a county has had greater than 10 cases, and an individual indicator for each of the four counties with over 500 cases.
7.
8.
9.
10.

Computational resources

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

Model is in R and is based on the open-source rf1 package as called by the dfmip package. The actual implementation differed slightly, as some of the diagnostic processes in the rf1 package did not have sufficient memory to process the US-wide data. Both packages available on www.github.com/akeyel, the script used for analysis available upon request. Models were run on a 64-bit Windows PC with a 2.9 GHz processor and 16 GB RAM.

Publications

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

The model is a derivation of the model used in Keyel et al. 2019 [1]. Major changes include a switch from deterministic forecasts to probabilistic forecasts, and reformatting of the code to be compatible with the dfmip model comparison framework. The analysis also switched to include monthly weather data, due to the rolling nature of the forecasts. The input data sets were also adjusted to include data streams that could be updated on a continuous basis.

Keyel et al. 2019. Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut. PLOS ONE 14: e0217854. https://doi.org/10.1371/journal.pone.0217854.

Participation agreement

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

Team lead name	Date
Alexander C Keyel	04/30/2020

Literature Cited

1. Keyel AC, Elison Timm O, Backenson PB, Prussing C, Quinones S, McDonough KA, et al. Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut. PLOS ONE. 2019;14: e0217854. doi:10.1371/journal.pone.0217854

2. Breiman L. Random forests. Machine learning. 2001;45: 5–32.

3. Meinshausen N. Quantile regression forests. Journal of Machine Learning Research. 2006;7: 983–999.

4. Meinshausen N. quantregForest: Quantile Regression Forests. 2017. Available: https://CRAN.R-project.org/package=quantregForest

5. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2: 18–22.

6. Bernard KA, Kramer LD. West Nile virus activity in the United States, 2001. Viral Immunology. 2001;14: 319–338.

7. EIA USEIA. 2015 RECS Survey Data. 2018 May. Available: https://www.eia.gov/consumption/residential/data/2015/

8. Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. International Journal of Climatology. 2013;33: 121–131.

9. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 2017. doi:10.1016/j.rse.2017.06.031

10. Wimberly MC, Davis JK. GRIDMET_downloader.js. University of Oklahoma; 2019. Available: https://github.com/ecograph/arbomap

11. Xia Y, Mitchell K, Ek M, Cosgrove B, Sheffield J, Luo L, et al. Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. Journal of Geophysical Research: Atmospheres. 2012;117.