Jun 26th, 10:40 AM - 12:00 PM

Using the LASSO to understand groundwater residence times

Jeff Jeffrey Starn
USGS NAWQA Program, East Hartford, CT, jjstarn@usgs.gov

iEMSs International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software
Fort Collins, Colorado, USA, Mazdak Arabi, Olaf David, Jack Carlson, Daniel P. Ames (Eds.)
https://scholarsarchive.byu.edu/iemssconference/2018/

Using the LASSO to Understand Groundwater Residence Times

Jeff Starn[1]
[1]*U.S. Geological Survey, National Water-Quality Assessment; jjstarn@usgs.gov*

Abstract: Groundwater residence-time distributions (RTDs) are critical for understanding lag times between recharge at the water table and base flow in streams. However, RTDs cannot be measured directly—they must be inferred from an analysis of data using models. Glacial aquifers present challenges to modeling approaches because they are spatially discontinuous and have highly variable properties. An innovative approach by the USGS uses machine learning in conjunction with numerical models that results in a rapid and robust way of generating RTDs. To demonstrate the method, computer programs were used to automatically create generalized finite-difference groundwater flow models in 30 watersheds across the northeastern glaciated U.S. RTDs were calculated from these models using flux-weighted particle tracking. Targets for machine learning were created from the simulated RTDs by fitting 3-parameter Weibull distributions. A form of penalized linear regression called Multitask LASSO (Least Absolute Shrinkage and Selection Operator) regression was trained on the Weibull parameters using hydrogeographic variables of the modeled domains as explanatory features. Because LASSO features are standardized, coefficient magnitudes can be compared to determine the relative importance of the features. Multitask LASSO was used to estimate the three Weibull parameters simultaneously, thus ensuring that the same features were used to estimate all of the parameters. The results show that aquifer heterogeneity and exchange of water between glacial deposits and bedrock and surface water are important for estimating RTDs. The quantitative understanding gained from the LASSO permits RTDs to be estimated across the glaciated region.

***Keywords**:* groundwater; residence time; machine learning; metamodel; statistical learning