Abstract

Author(s)	Azadeh Golshan

Title

Modern machine learning tools to predict cyanobacteria in Chichester Dam

Description

We used two machine learning methods (random forest, RF; and recurrent neural network, RNN) for predicting *Microcystis spp.* counts in Chichester Dam, which is an on-stream reservoir in the Lower Hunter region of NSW. Cell counts were predicted using historical weekly observations of ~100 parameters over a 20-year period. The parameters included physical observations (temperature, water column height, etc.), chemical observations (pH, dissolved nutrients, etc.), and others (weather, inflow, soil moisture, etc.). A goal was to use the trained prediction algorithms to explore the likely effects of future changes in climate and land use on the prediction of *Microcystis* populations in this reservoir.

Predictive performance was evaluated by cross validation. For the RF algorithm, performance was examined using a range of different lags (i.e. using observations from a set range of different previous time steps to predict current cell counts). The RNN algorithm has the key advantage of handling this lagged-prediction problem internally.

Initial results demonstrate that both methods provide useful predictions for cell counts. The algorithms capture systematic seasonal changes and accurately predict many of the high-cell-count periods. Individual lagged inputs in the RNN method proved to be far superior in providing a fit to the actual data. Results of analysis have indicated that dissolved oxygen, pH, and nutrient loads were the predominant variables in determining the onset and duration of cyanobacterial growth.

The trained algorithms also provide easy-to-use platforms for exploring the sensitivity of cell counts to changes in any (or all) of the predictor observations. These explorations can be used to assess risks posed by climate change and other future scenarios.