AOGS 2006

Extracting Knowledge from Parsimonious Genetic Programming Evolved Rainfall-Runoff Models

NITIN MUTTIL[1], AMITHIRIGALA W. JAYAWARDENA[1]

[1] *Department of Civil Engineering, The University of Hong Kong*

The problem of accurately determining river flows from rainfall, evaporation and other factors, occupies an important place in hydrology. Practitioners in water resources have embraced data-driven modeling approaches enthusiastically, as they are perceived to overcome some of the difficulties associated with physics-based approaches. In the recent past, an evolutionary based data-driven modeling approach, genetic programming (GP) has been used for rainfall-runoff modeling. As compared to traditional data-driven techniques, the key advantage of GP for the modeling process is its ability to produce models that build an understandable structure, i.e., a formula or an equation. Interpretation of the GP models can shed physical insight into the hydrological processes involved. In this study, GP models are evolved using datasets from various catchments located in Southeast Asia and an attempt is made to extract knowledge from those models. To evolve physically more interpretable GP models, various GP parameters (including the equation size, the function set used, etc.) are intelligently set so that the evolved models are simple and easily understandable (i.e., are parsimonious). The GP evolved models are examined closely and it is observed that the equations give a clear indication of the time of concentration of the catchments and the antecedent moisture content of the soil. For river flow data from a catchment exhibiting chaotic behavior, the GP approach can guide in the search for the appropriate embedding dimension. The GP models also have the automatic ability to select significant input variables that contribute to the modeling and to disregard those that do not, which result in the model's interpretability. These significant variables can be further used as the only inputs for prediction into a data-driven model, thus reducing the dimensionality of the inputs.