Geophysical Research Abstracts
Vol. 13, EGU2011-4623, 2011
EGU General Assembly 2011

How to transfer research efforts on data-driven modeling to technicians? The case of EPR

Luigi Berardi (1), Daniele, B. Laucelli (1), Angelo Doglioni (2), Vincenzo Simeone (2), and Orazio Giustolisi (1)

(1) Department of Civil and Environmental Engineering, Technical University of Bari, Bari, Italy (l.berardi@poliba.it), (2) Department of Environmental Engineering and for the Sustainable Development, Technical University of Bari, Taranto, Italy

In the last two decades the availability of large and detailed databases together with the increased computational capabilities has motivated researchers to propose innovative techniques and methodologies to mine information from data in the field of hydroinformatics. Nonetheless, the transfer process of research efforts to technicians is still far from being accomplished thus making such advancements useless for wider real scale applications. In addition, practitioners are often skeptical from moving from tools/methodologies they are familiar with to newer and more complex ones. Thus, it is necessary to find effective ways to integrate the research achievements within tools already known to technicians in order to facilitate their adoption.

The Evolutionary Polynomial Regression (EPR) [1] has been introduced in the hydroinformatics community as a hybrid data-driven technique, which combines the effectiveness of genetic algorithms with numerical regression for developing simple and easily interpretable mathematical model expressions. Recently, it has been continuously developed in order to match needs of both researchers and practitioners. For example, the multi-objective search paradigm has been introduced [2] for developing multiple models by simultaneously optimizing fitness to training data and parsimony of resulting mathematical expressions. Such improvement allows for a sudden understanding of existing patterns in data (if any) by comparing different optimal models and the selection of the model which best fit the purpose of the analysis.

The EPR integrates some advancement in artificial intelligence and data-driven modeling areas like an efficient multi-objective genetic algorithm and the algorithm for developing nonlinear mathematical structures using an integer coding.

It was used to investigate and model several phenomena/systems including pipe failures in water distribution and wastewater networks, rainfall-groundwater dynamics, scour depths downstream grade-control structures, explicit formulations of Colebrook-White friction factor, and evapotranspiration process. It was also adopted for other applications in the areas of geotechnical and structural engineering.

Over the last four years EPR software has been continuously upgraded and distributed as a stand-alone software application. In order to facilitate its integration with MS-Excel, which is probably the best known data management and analysis tool used by practitioners and researchers, the EPR versions distributed so far allowed the user to arrange input data and get results in separate Excel spreadsheets. Nonetheless, such versions still presented some inconveniences. For example, it required a special arrangement of data in the input spreadsheet and modelling options (i.e. EPR settings) were retrievable after each EPR run, as a figure of the main interface window only.

In order to overcome such difficulties and facilitate a full integration of EPR with tools that technicians are familiar with, an updated version of EPR called EPR-Excel has been recently developed. The whole EPR package is available as an MS-Excel add-in and the user can launch EPR runs as a function in MS-Excel.

Input data can be manually selected from other spreadsheets, without the hindrances of previous versions. A sheet containing all EPR modelling options can be easily modified and retrieved for next analyses. Moreover, the user is guided through proper option setting by some tips, which recall the meaning of each parameter. In order to facilitate multiple analyses, the expression(s) of model(s) obtained, the values model predictions and fitness indicators of each model are stored in a separate Excel file. This allows the user to perform multiple analyses while preserving complete information on both input data and modelling settings.

[1] Giustolisi, O., Savic, D.A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. J. of Hydroinformatics. 8, 207–222.

[2] Giustolisi, O., Savic, D.A., 2009. Advances in Data-Driven Analyses and Modelling Using EPR-MOGA. J. of Hydroinformatics. 11, 225-236