Transforming data into information

M.B. Beck and Z. Lin
Warnell School of Forest Resources, University of Georgia, Athens, Georgia 30602-2152, USA
(E-mail: mbbeck@arches.uga.edu)

Abstract In spite of a long history of automated instruments being deployed in the water industry, only recently has the difficulty of extracting timely insights from high-grade, high-volume data sets become an important problem. Put simply, it is now relatively easy to be “data-rich”, much less easy to become “information-rich”. Whether the availability of so many data arises from “technological push” or the “demand pull” of practical problem solving is not the subject of discussion. The paper focuses instead on two issues: first, an outline of a methodological framework, based largely on the algorithms of (on-line) recursive estimation and involving a sequence of transformations to which the data can be subjected; and second, presentation and discussion of the results of applying these transformations in a case study of a biological system of wastewater treatment. The principal conclusion is that the difficulty of transforming data into information may lie not so much in coping with the high sampling intensity enabled by automated monitoring networks, but in coming to terms with the complexity of the higher-order, multi-variable character of the data sets, i.e., in interpreting the interactions among many contemporaneously measured quantities.

Keywords Biological wastewater treatment; filtering-smoothing algorithms; models; nitrification; recursive estimation; signal extraction

Introduction
To imagine one might have too many data – to be data-rich, yet information-poor – would have been a startling thought just a few years ago. Whatever the reasons for this change in our fortune, in both problem-solving opportunities and the different challenges following from these new opportunities, our concern in this paper is that of making the most sense of the great volumes of data to which we can now have access. We acknowledge the substantial interest in further developing the technologies for generating ever more intense and varied streams of water quality data; and we will presume that creating systems and procedures of analysis for acquiring insights into the practical implications of all these data, in real time, in particular, will be especially challenging. However, our focus herein is on addressing just two issues: an outline of one candidate methodological framework for such procedures of analysis, essentially founded upon the principles of recursive estimation (Young, 1984, 1998; see also Beck, 1987); and an assessment of its success for a single case study dealing primarily with aspects of biological nitrification in an activated sludge process of wastewater treatment.

Methodological framework
Suppose the behavior of the system being monitored can be defined according to the following (lumped-parameter) representation of the state variable dynamics,

\[
dx(t)/dt = f\{x,u,\alpha; t\} + \xi(t)\]  \(1a\)

with observed outputs being defined as follows,

\[y(t) = h\{x,\alpha; t\} + \eta(t)\]  \(1b\)