Geophysical Research Abstracts,
Vol. 10, EGU2008-A-05647, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-05647
EGU General Assembly 2008
© Author(s) 2008

Automatic Analysis of hydrologic Time Series using Time Series Knowledge Mining to identify System States

O. Gronz (1,2), **M. Casper** (1), P. Gemmar (2)

1) Physische Geographie, Universität Trier, Germany, (2) Institut für Innovative Informatik-Anwendungen, Fachhochschule Trier, Germany (gronzo@fh-trier.de / Phone: +49-651-8103-575)

The response of a catchment considerably depends on the actual state of the system, resulting in enhanced requirements in the modeling of rainfall-runoff processes. An effective approach to reproduce this dynamics is the usage of fuzzy models of Takagi-Sugeno-type to calculate or predict discharge. The design of such a model consists of two essential steps: the definition of the structure and the identification of parameters. The latter one can be effectively solved using data-driven techniques. For the definition of the structure, expert knowledge is mandatory: Usually, numerous time series are available containing measurements of discharge, precipitation, temperature, soil moisture etc. From this large set of time series, the expert has to identify those series that can effectively be used as state variables in a fuzzy system, representing the catchment's state in an optimal way. This knowledge is not always available; for each catchment, this knowledge needs to be gathered anew and moreover, this knowledge is not necessarily complete or correct. To support the identification of state variables, the method Time Series Knowledge Mining (TSKM) has been applied. This iterative and interactive method consists of five steps: in the first step, the preprocessing, common methods are applied to detect and - if possible - remove or damp systematic and random errors like outliers, drift, noise etc. In the second step, the dimensionality of the input space is reduced by combing different time series in new, artificial time series, so-called aspects. Each aspect represents a single time series or the combination of several ones, e. g. by calculating the mean value element by element. In the

third step, the numeric aspects are transformed to a symbolic, qualitative description, the tones, which represent the temporal concept of duration. Each tone represents a specific, persistent state of an aspect and can be combined with labels indicating the meaning of the state descriptively. Tones are defined using characteristic functions, e. g. value-based ones. To calculate suitable intervals for the tones automatically, the Persist Algorithm can be used, which calculates bins resulting in preferably persistent states and not only depending on the distribution of the values. Furthermore, different filters can be applied to fill small gaps in the occurrence of tones or to remove short tones. In the fourth step, the temporal concept of coincidence is represented by chords: the simultaneous occurrence of several tones. Again, several filters can be applied to remove small gaps or short chords. In the last step, phrases represent the temporal concept of partial order: the consecutive occurrence of chords. The knowledge that is discovered by TSKM is presented using Time Series Knowledge Representation, a pattern language that describes temporal knowledge in different, hierarchically structured elements. The patterns are novel, useful, easily interpretable, more abstract and more compact than the underlying time series. TSKM has been applied with data from a $7km^2$ catchment in the northern Black Forest, Germany. From the available time series, four soil moisture time series were selected as aspects, each of which representing the hydrologic behavior of its site. Besides discharge and temperature as further aspects, new aspects were derived from the 1h precipitation accumulation time series, representing increasing time intervals like 2h accumulation, 4h, 8h, 16h, 32h, 64h and 128h to find out, which time interval is the most suitable one for the representation of the system state. Additionally, a new aspect was defined using discharge, precipitation and the catchment's unit hydrograph to calculate the discharge disposition in each time interval. The usage of TSKM enabled the identification of two soil moisture time as potential state variables. The fuzzy systems generated using these two time series were more efficient than all previously generated models, even if those models contained expert knowledge.