Environmental Modelling & Software 97 (2017) 229–242

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management

Tassadit Bouadi [a, *], Marie-Odile Cordier [a], Pierre Moreau [b], René Quiniou [c], Jordy Salmon-Monviola [b], Chantal Gascuel-Odoux [b]

[a] University of Rennes 1, IRISA, F-35042 Rennes, France
[b] UMR 1069, Soil Agro and HydroSystem, INRA, Agrocampus Ouest, F-35000 Rennes, France
[c] INRIA Bretagne Atlantique, IRISA, F-35042 Rennes, France

ARTICLE INFO

Article history:
Received 14 September 2016
Received in revised form
20 July 2017
Accepted 28 July 2017

Keywords:
Multidimensional modeling
Simulation data
Data warehouse
OLAP
Water quality
Nitrogen
Catchment
Distributed agro-hydrological model

ABSTRACT

Spatially distributed agro-hydrological models allow researchers and stakeholders to represent, understand and formulate hypotheses about the functioning of agro-environmental systems and to predict their evolution. These models have guided agricultural management by simulating effects of landscape structure, farming system changes and their spatial arrangement on stream water quality. Such models generate many intermediate results that should be managed, analyzed and transformed into usable information. We describe a data warehouse (N-Catch) built to store and analyze simulation data from the spatially distributed agro-hydrological model TNT2. We present scientific challenges to and tools for building data warehouses and describe the three dimensions of N-Catch: space, time and an original hierarchical description of cropping systems. We show how to use OLAP to explore and extract all kinds of useful high-level information by aggregating the data along these three dimensions and how to facilitate exploration of the spatial dimension by coupling N-Catch with GIS. Such tool constitutes an efficient interface between science and society, simulation remaining a research activity, exploration of the results becoming an easy task accessible for a large audience.

Software availability

Name of software: N-Catch (Nitrogen in Catchment data warehouse)
Developers: Tassadit Bouadi and Sylvain Dousset
Contact: tbouadi@irisa.fr
Hardware required: Experiments were performed on an Intel Core i7 CPU at 2.8 GHz and 16 GB of RAM on a Mac OSX platform
Software required: The relational database management system MySQL, Quantum GIS (QGIS) and Microsoft Windows, Mac OSX or Linux operating system
Program language: Perl, Python
Software availability: source code can be provided through collaborative arrangements

1. Introduction

Agro-hydrological models have been used extensively by researchers and stakeholders as the scientific basis for environmental management by estimating nonpoint-source pollution, identifying source areas, predicting effects of climate and land-use changes and testing the efficiency of mitigation plans to improve water quality at the catchment level (Ferrant et al., 2014; Rode et al., 2010; Wellen et al., 2015). Extensive research has focused on improving the ability of these models to consider the heterogeneity of structures and processes within agricultural landscapes by representing their spatial distributions.

These models generate a large volume of spatiotemporal results of various formats and semantics. Generally, only daily flux and concentrations at the catchment outlet are analyzed, even though finer-grained variable, temporal or spatial resolutions are potentially available. The main reason for this is a technological barrier that prevents efficient data processing. To address this issue, efficient tools are needed to store, display and analyze this

* Corresponding author.
E-mail address: tbouadi@irisa.fr (T. Bouadi).

http://dx.doi.org/10.1016/j.envsoft.2017.07.019
1364-8152/

spatiotemporal information and turn it into useful knowledge that enables better understanding of agro-environmental systems and adaptations to meet environmental targets.

Tools that analyze and visualize simulated data distributed in space and time could help researchers and stakeholders explore effects of different scenarios at multiple temporal and spatial resolutions, such as plot or sub-catchment levels, or for specific periods of the year. Furthermore, they should provide the means to analyze and cooperatively identify effective and locally-adapted solutions to improve water quality in agricultural catchments. For illustration, scenarios that can be tested with agro-hydrological models include agricultural practices, such as the introduction of catch crops (Moreau et al., 2012a) or hedgerow spatial arrangement (Benhamou et al., 2013) within a catchment. The water and solute recharge into the groundwater calculated at the plot level results from interactions between environmental conditions and agricultural practices, which can vary greatly across a catchment due to soil or hydrological conditions. Exploring simulation data is a useful means to analyze these local interactions, if considering them is in the modeling structure, and to propose specific changes according to the location within the catchment.

To this end, recent studies have shown how a data warehouse (DW) and On Line Analytical Processing (OLAP) technologies are used to analyze environmental simulations (Boulil et al., 2013; Mahboubi et al., 2010). A DW is a subject-oriented, integrated, time-variant, non-volatile collection of data that supports the management decision-making processes (Chaudhuri and Dayal, 1997; Inmon, 2005). DWs are emerging as a key technology for organizations seeking to use their data to keep track of activities and improve data analysis. DWs are used (i) to provide access to massive data accumulated over time from many sources and in various formats (computer files, traditional databases, text documents, etc.) and (ii) to support multi-dimensional data analysis to make strategic and tactical decisions. DW users can extract trends and variability from the data according to various criteria to better support decision-making or to discover hidden information.

Few DWs have been developed in the agro-environmental sciences (Abdullah, 2009; Boulil et al., 2013, 2014; Nilakanta et al., 2008; Pinet et al., 2010). They support analyses of agricultural data along different dimensions. Nilakanta et al. (2008) developed the National Agricultural Resources Information System DW for the Indian agricultural sector. This DW provides strategic and periodic information to researchers and planners to facilitate decision making. Abdullah (2009) developed an OLAP tool, ADSS-OLAP, to analyze mealybug incidence on cotton crops. The dimensional model of ADSS-OLAP includes different dimensions for analysing the effect of climate, pesticide and geography. These dimensions are aggregated as a logical OLAP cube, with a classical multidimensional model. Pinet et al. (2010) considered ways to use the Unified Modeling Language to design agricultural DWs. They presented a method for designing a DW and applied it to analyze spatial impacts of pesticide use in agriculture. Boulil et al. (2013) developed an OLAP system to store and analyze pesticide transfer data generated at the soil-column scale by a model called MACRO, to validate the model and compare results of different versions of the model. Boulil et al. (2014) then applied the OLAP system to analyze data on stream water quality. The architecture was extended with complex aggregate functions used to define indicators.

Recent studies have investigated Spatial OLAP (SOLAP) to study stream water quality in rivers. SOLAP systems integrate advanced OLAP and Geographic Information Systems (GIS) functions in a unique framework in which explicit representation of the spatial dimension allows users to visualize query results on maps and to use topological, metrical and directional operators when "slicing" multidimensional data (Bimonte and Miquel, 2010). Vernier et al. (2013) considered a SOLAP system to characterize agricultural activities and calculate agro-environmental indicators. Similarly, Berrahou et al. (2015) developed a solution that facilitates spatio-temporal analysis of hydroecological data by considering different levels of data quality within the system. These studies show that agro-environmental DWs are rarely developed. The most likely explanation for this is the difficulty in collecting field data in agro-environmental sciences, since data collection remains a slow and expensive process. In the past, DWs and OLAP systems were used mainly to analyze observed data. Their use for simulation data is poorly developed, particularly in the field of distributed models, despite their potential utility.

This work aims to develop methods to store, display and analyze simulation data obtained from an agro-hydrological model, and to design and implement operational tools for researchers and stakeholders to represent, understand, explain and formulate hypotheses about the catchment system they study. Fundamental to this application is development of "what if" management questions meant to evoke possible outcomes in different scenarios (e.g., "Which plots and cropping systems emit the least nitrogen to air and groundwater?"). Since the development of DWs for water-quality issues is quite new, we first present basic concepts of DWs, as well as studies of OLAP, particularly in the agro-environmental domain. Next, we describe the N-Catch DW, dedicated to simulation data generated by the distributed agro-hydrological model TNT2 (Topography-based Nitrogen Transfer and Transformations), a process-based and spatially explicit model that simulates transfer and transformation of nitrogen (N) in agricultural catchments and predicts water and N fluxes at a daily time step at their outlets (Beaujouan et al., 2001, 2002). We describe how the DW was designed and built and we use a case study to demonstrate how exploring simulated data can be used to extract knowledge to help users better understand drivers of N emissions in the environment. We then discuss the generality of the case study.

2. Main concepts of DWs and OLAP technologies

2.1. A DW as a multidimensional and hierarchical data model

DWs and OLAP systems are widely recognized as decision-support systems for analysis of huge volumes of data generated by a multidimensional model, which defines the concepts of "facts" and "dimensions" (Kimball, 1996). DWs allow users to deliver highly aggregated data from heterogeneous sources to respond to complex queries and perform analyses, and in this way, discover implicit properties.

Facts represent the subjects of analysis and are described by quantitative "measures", which are analyzed at different "granularities", i.e. at different hierarchical levels of the dimensions (Berrahou et al., 2015; Sautot et al., 2015). For example, the fact "agricultural crop" can be represented according to three dimensions (crop, time and location) and described by the measure "crop yield" (kg. ha^{-1}) (Fig. 1). Dimensions represent measure-analysis criteria and allow measures to be viewed and analyzed from different perspectives. Measures at upper levels of a hierarchical dimension are obtained by aggregating the measures at the next lower level in the hierarchy.

Aggregation functions (e.g., sum, average, maximum) can be defined for each measure in the hierarchical data model. For example, in a hierarchy associated with the dimension "Location", each mesh belongs to a plot, and each plot belongs to a catchment (Fig. 2).

During data analysis, users exploit the DW by considering