Open access • Journal Article • DOI:10.1016/J.ENVSOFT.2017.07.019

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

Tassadit Bouadi, Marie-Odile Cordier, Pierre Moreau, Ren Quiniou ...+2 more authors

Institutions: University of Rennes, Agrocampus Ouest, French Institute for Research in Computer Science and Automation

Published on: 01 Nov 2017 - Environmental Modelling and Software (Elsevier)

Topics: Data warehouse, Online analytical processing and Dimension (data warehouse)

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

Tassadit Bouadi, Marie-Odile Cordier, Pierre Moreau, René Quiniou, Jordy Salmon-Monviola, Chantal Gascuel

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

Highlights

- A data warehouse (DW) as a tool to explore simulated agro-environmental data
- N-Catch as an example of a DW for analyzing N emissions across a catchment
- DWs for catchment N management

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.

Keywords: multidimensional modeling, simulation data, data warehouse, OLAP, water quality, nitrogen, catchment, distributed agro-hydrological model.

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 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 modelling 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; Immon, 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 et al. (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 spatiotemporal 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.

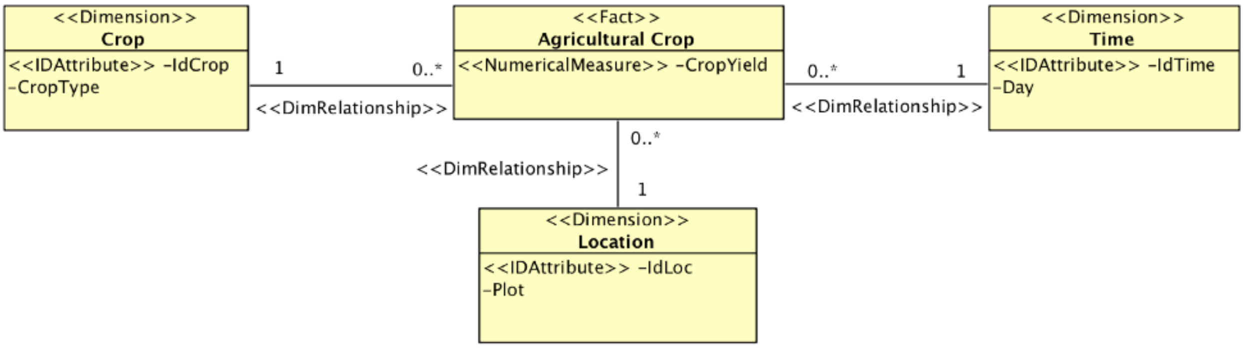


Figure 1 : Example of a multidimensional data model with three dimensions (crop (Id-Crop), time (Id-Time), and location (Id-Loc)) and a measure (crop yield)

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).

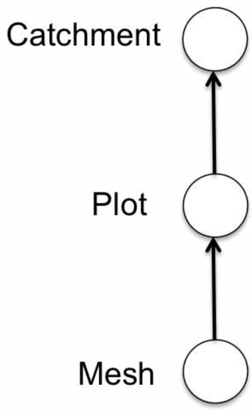


Figure 2 : Example of a hierarchical structure for the dimension "Location"

During data analysis, users exploit the DW by considering different combinations of dimensions and different levels of their hierarchies. To select the right data at the right abstraction level, users express and submit queries to the DW. Generally, such queries often require scanning thousands or millions of records. Consequently, computing a query can be highly time-consuming, while, ideally, the analysis process should be interactive. To reconcile these two conflicting needs (i.e. a quick response to a query to a large database), the concept of a "data cube", the central OLAP concept, was introduced (Gray et al., 1996) to store precomputed partial results. A data cube (or OLAP Cube) is a data abstraction that allows users to view aggregated data according to a set of dimensions related to a pre-determined structure. The cells of a data cube contain the measures selected from the DW according to the dimensions and their hierarchies. A data cube is generally n-dimensional, though examples of data cubes are often 3-dimensional for illustration purposes. As in the previous example, data-cube representation of agricultural crop data with three dimensions (crop, time and location) can be used to calculate a measure, i.e. a crop yield, for a given crop at different levels of time and location with a sum function (Fig. 3).

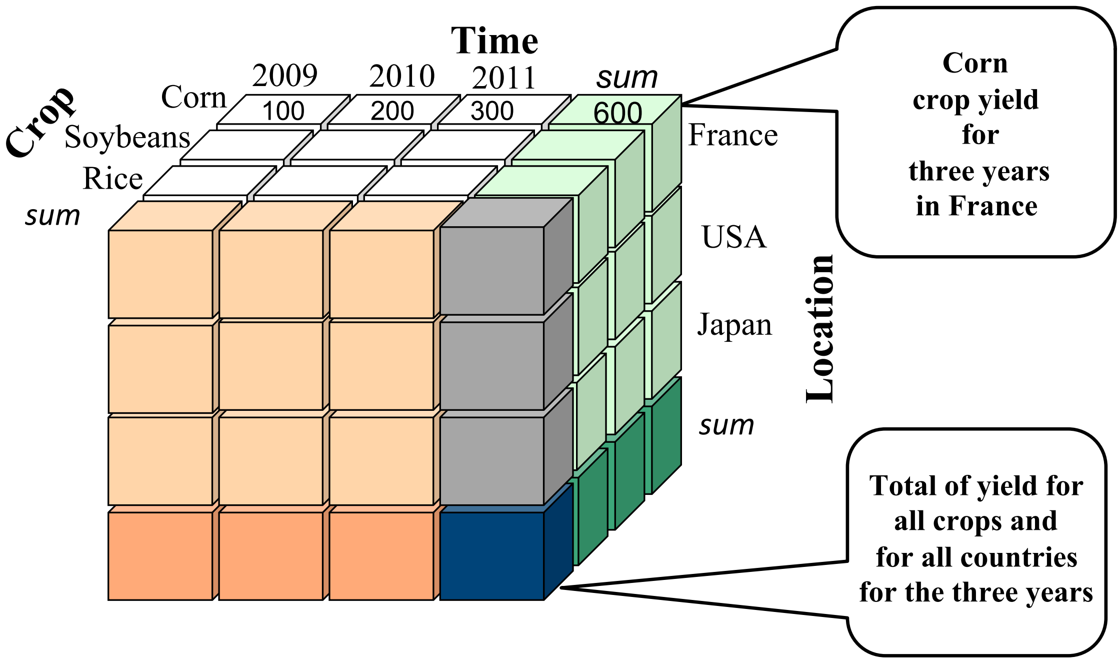


Figure 3 : Example of a data cube for agricultural crops

2.2. Relational platforms of DWs, OLAP analysis and data cubes

DWs contain multidimensional and summarized data. OLAP queries are defined to explore and perform complex analysis by interactively generating different views of these data from different perspectives. OLAP operations aim to select planes from the DW and change the granularity at which data are observed. The main OLAP operations are “roll-up”, to move to a more abstract level on one dimension; “drill-down”, the inverse of roll-up, to move to finer-grained data; “slice and dice”, to focus on data related to subsets of dimension values; and “pivot”, to rotate the axes on which data are viewed to provide an alternative display of the data. For example, two OLAP queries can calculate corn yield for three years in France (Fig. 3); they consist of selecting (i.e. slicing) the value "Corn" in the "Crop" dimension and the value "France" in the "Location" dimension. They can also calculate the sum of yields of all crops for three years in all countries, using roll-up operations on all dimensions of the agricultural crop cube (Fig. 3).

DW and OLAP data cubes are generally implemented in relational platforms consisting of four tiers (Fig. 4): (i) the ETL (Extract, Transform and Load), which extracts and precomputes data from heterogeneous sources and loads them into the DW; (ii) the DW itself, the relational database that stores the finest data; (iii) the OLAP server, which calculates data cubes from the DW; and (iv) the OLAP client, which displays data-cube information using tables, diagrams (e.g. pie charts, histograms) and reports (Boulil et al., 2014). An OLAP client makes it easy to optimize manipulation of the cube to change focal areas, granularity, or analysis functions, for example. In addition, data-mining tools are sometimes available to perform trend analysis or predictions.

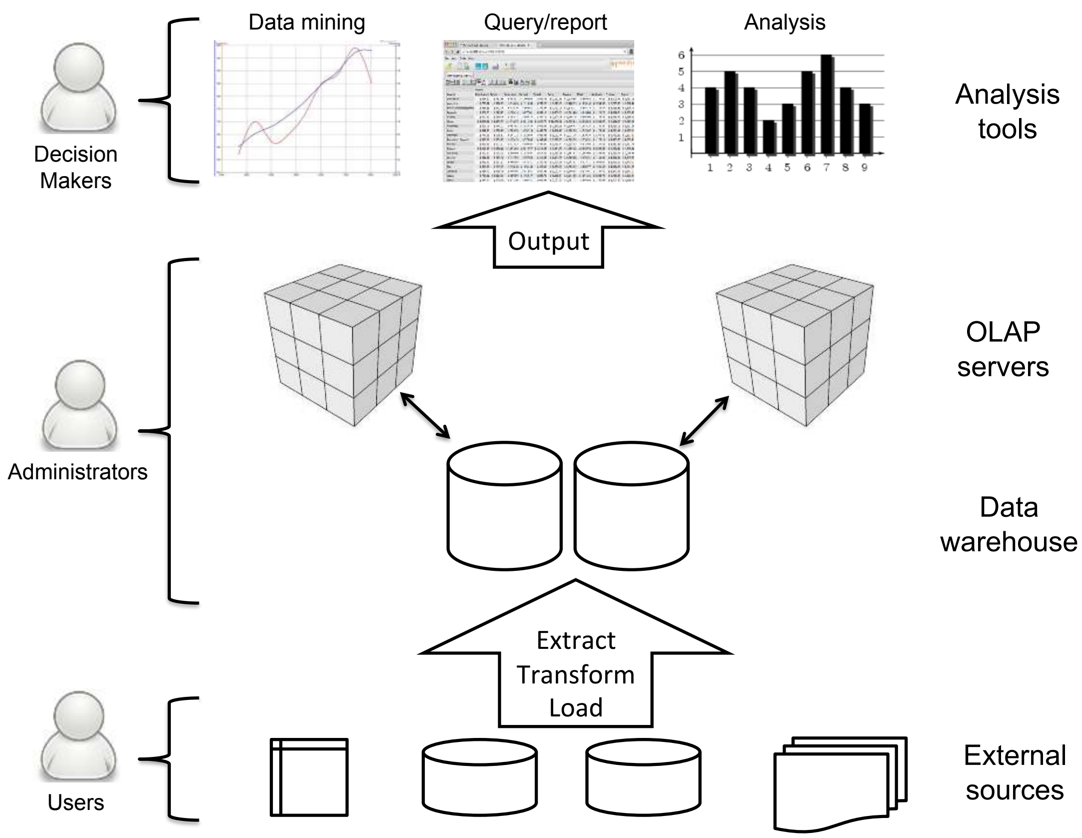


Figure 4 : Typical Data-Warehousing architecture

3. TNT2 model and N-Catch DW design and implementation

3.1. The spatially distributed agro-hydrological model TNT2

TNT2 (Beaujouan et al., 2002) is a process-based and spatially-distributed model designed to capture spatial and temporal interactions of hydrological and agronomic processes throughout catchments. TNT2, developed to study water and N fluxes in small agricultural catchments (<50 km^2), is dedicated to improving understanding of catchment systems by performing virtual experiments (Benhamou et al., 2013; Liao et al., 2016; Salmon-Monviola et al., 2013), including testing effects of numerous and diverse mitigation options proposed by local stakeholders (Moreau et al., 2012b) or public-policy makers (Durand et al., 2006) to decrease nitrate concentrations at the catchment outlet (Durand et al., 2015; Ferrant et al., 2011; Moreau et al., 2012a), as well as to support management and decision-making.

The TNT2 model comprises (i) the generic soil-vegetation atmosphere transfer model STICS (Brisson et al., 1998); (ii) the fully-distributed hydrological model TNT, based on TOPMODEL hypotheses (Beven, 1997); and (iii) a sub-model adapted from NEMIS (Hénault and Germon, 2000; Oehler et al., 2009) that simulates heterotrophic denitrification in soils. More details about hydrological and agronomical sub-model with each water and nitrogen processes are described in (Beaujouan et al., 2002, Salmon-Monviola et al., 2013; Ferrant et al., 2016). To represent the heterogeneity of water and N inputs and pathways, TNT2 performs calculations explicitly on a grid, from cell to cell. The catchment is thus modeled by a cluster of three-dimensional columns, each corresponding to a pixel of a Digital Elevation Model (DEM) of the catchment (Moreau et al., 2013). Each column is divided into two compartments: a soil layer and an underlying shallow aquifer layer. Soil and aquifer porosity is determined by both retention (micro) and drainage (macro) porosities. Water flow in TNT2 is modeled according to main hypotheses derived from TOPMODEL: i) the hydraulic gradient in each cell is constant and controlled by topography, and the gradient is calculated from the DEM at the beginning of the simulation; ii) the hydraulic transmissivity decreases exponentially with depth, and the model is based on a daily water-balance calculation for each cell and computes an explicit cell-to-cell multi-directional routing (water in one cell can flow to several cells), based on a D8 algorithm (Ocallaghan and Mark, 1984). Three outgoing flows are calculated for each cell in TNT2 : i) overland flow, resulting from soil saturation, which is routed to the soil surface of the downslope cells; ii) exfiltration that comes from excess groundwater and is also routed to the surface of the downslope cells; iii) subsurface flow that comes from groundwater and is routed to the downslope cell groundwater. Water balance in river cells is calculated as in other cells, but all the calculated outflows (overland flow, exfiltration and subsurface flow) are assumed to be drained by the river and are routed directly from the river cell to the outlet without any interaction with other cells. The travel time in the stream is ignored. Nitrogen processes in soil and aquifer layer compartment of non-agricultural zones (urban and woodlands) of the catchment are taking account for. A part of surface water and nitrogen flow (overland flow, exfiltration flow) from upslope cells and rainfall input infiltrate in soil of non-agricultural cells. The other part is routed to the outlet. For non-agricultural cell as in other cells, subsurface flow of water and nitrogen, that comes from groundwater is routed to the downslope cell groundwater. Model inputs and parameters include (i) a DEM in raster format; (ii) a map in raster format delineating agricultural plots, roads and the hydrological network; (iii) a map in raster format of homogeneous soil zones; (iv) a map in raster format representing homogenous climate zones, which allow climate gradients to exist within the catchment; (v) daily climate data (i.e., minimum and maximum air temperatures, precipitation, potential evapotranspiration, total solar radiation) for each climate zone; (vi) 20 soil properties for each soil type in the soil-zone map; and (vii) cropping systems for each plot during the simulation period. Cropping systems are defined at two levels (Leenhardt et al., 2010): (i) crops and their succession over time in each plot, called a crop rotation; and (ii) the crop management system (CMS), representing an organized series of cultivation techniques or crop operations (e.g. sowing, mineral and organic N applications, harvesting, grazing, mowing) applied to a crop to obtain a given product. A crop operation has a duration defined by starting and ending dates, as well as a chronological position in the CMS in which it occurs.

Cropping systems were described by temporal sequences, allowing us to manage (i) temporal relations between crop operations and CMSs and between crop types and crop rotations and (ii) multiple occurrences of the same element in the same sequence (e.g. "fertilization" may occur on two different dates in a CMS) (Fig. 5). The model CSAM (Salmon-Monviola et al., 2012) was used to generate cropping systems as input to TNT2. Using CSAM, classification systems for farms and fertilization practices were defined to classify farm-level organization for crop-succession and CMSs, in particular N fertilization. Winter cover crops and multiple agricultural strategies per crop at the farm level can be represented. Cropping systems are modeled with CSAM in three steps: (i) model crop-cover succession in summer with Markov chains (Isaacson et al., 1976) based on empirical data and (ii) in winter with rules based on expert agronomic knowledge and (iii) use, a Knapsack-based algorithm (Salmon-Monviola et al., 2012) to allocate a CMS to plots, with constraints on several CMS of each crop type.

3.2. Design of the N-Catch DW

Building an agro-environmental DW is a complex task: no relevant data should be excluded, in order to avoid incorrect or incomplete information in the DW. Therefore, the architecture, measures, facts, dimensions and ETL of the DW must be carefully designed. We designed the DW using the open-source business intelligence platform Pentaho® (ver. 3.5) (http://www.pentaho.com/), a high-quality conceptual and technical solution. Specifically, we used the Pentaho module Mondrian (http://community.pentaho.com/projects/mondrian/) to design the related data cube. Mondrian is an OLAP engine that enables designing, publishing and querying multidimensional cubes.

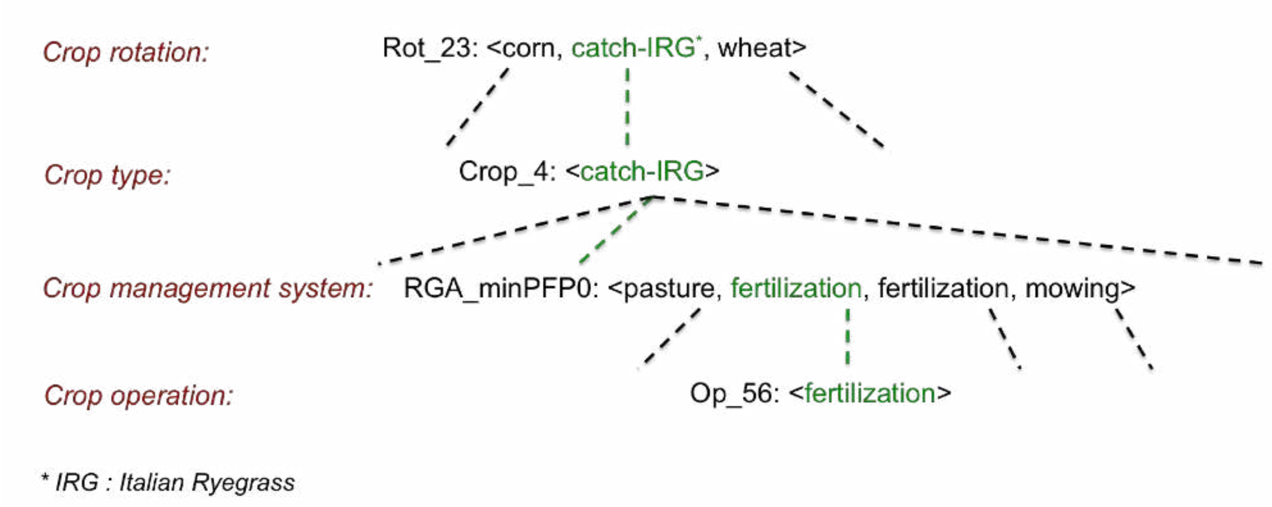


Figure 5 : Description of the agricultural dimension of the N-Catch data warehouse

3.2.1. Measures stored in the N-Catch DW

To study how agricultural practices influence N emissions to the air, groundwater and stream water, measures in the N-Catch fact table are composed of input data (i.e. cropping systems, catchment description, meteorological and hydrological parameters) and extensive output data generated by TNT2 (i.e. water fluxes and N concentration and fluxes at a daily time step for each plot in the catchment). These simulated results are stored in two types of files: (i) N variables predicted by the model for each operation *i.e* sowing, fertilizing, harvesting, grazing, mowing) for crop sequences on each plot throughout the simulation period, and (ii) water and N concentrations and fluxes at a daily time step for each plot. Out of 44 variables, 16 that describe water and N flux in the groundwater and air, water and N stored in the soil and groundwater and N denitrification are stored in N-Catch (Table 1). From these 16 variables, we added two calculated measures that correspond to N flux from soil to groundwater and from soil to air, defined as:

$$NFlux_{Soil\ to\ GW} = N\_Atm - N\_Denit - N\_Volat\_Ferti + N\_Mine\_Manure + N\_Mine\_Grazing + N\_Mine + N\_Fix - \Delta N\_Soilwater + N\_Fertilizer - N\_Plant\ (kg.ha^{-1})$$

$$NFlux_{Soil\ to\ Atm} = N\_Denit + N\_Volat\_Ferti + N\_Volat\_Manure\ (kg.ha^{-1})$$

With ΔN_Soilwater, the stock variation of nitrogen stored in soil water storage (kg.ha^{-1}). Each variable is defined as in Table 1.

A three-dimensional data model of N-Catch includes the dimensions of location, time and agricultural practices, as well as selected measures (Table 1), as detailed below:

Variable	Description
GW	Groundwater table (height from the impermeable layer) (m)
WSC	Soil water storage capacity (mm)
N_Soilwater	Nitrogen stored in soil water storage (kg.ha^{-1})
N_Weathered	Nitrogen stored in weathered layer (kg.ha^{-1})
N_GW	Nitrogen stored in groundwater (kg.ha^{-1})
N_Fix	Atmospheric nitrogen fixed by plants (kg.ha^{-1})
N_Mine	Mineral nitrogen resulting from mineralization (kg.ha^{-1})
N_Denit	Denitrified nitrogen (kg.ha^{-1})
N_Sequestre	Nitrogen stored in organic matter (kg.ha^{-1})
N_Plant	Nitrogen fixed by plants (kg.ha^{-1})
N_Mine_Grazing	Amount of mineral N returns from cows during grazing (kg.ha-1)
N_Mine_Manure	Amount of mineral N from manure (kg.ha-1)
N_Volat_Manure	Amount of N from manure to atmosphere by volatilization (kg.ha-1)
N_Volat_Ferti	Amount of N from fertilizers to atmosphere by volatilization (kg.ha-1)
N_Atm	Amount of nitrogen from atmospheric deposition (kg.ha-1)
N_Fertilizer	Amount of N mineral fertilizer (kg.ha-1)

Table 1: Description of 16 TNT2 output variables

3.2.2. The dimensions of the N-Catch DW

N-Catch has three dimensions: spatial, temporal and agricultural.

Spatial dimension. The spatial dimension (i.e. location) is useful for quantifying interactions such as those between plot location and agricultural practices. In N-Catch, the location dimension has two levels: plot and catchment. TNT2 inputs (e.g. crop, fertilizer amount, operation date) and outputs (e.g. N emissions from soil to air and groundwater) are available at the plot and whole-catchment levels. Because of this dimension, any attribute of plot data (e.g., soil type, surface area) is easily aggregated, using classical aggregation operators (such as COUNT, SUM, etc.) or specific aggregation functions, at the catchment level.

Temporal dimension. The temporal dimension is useful to allow users to analyze effects of changes in the sequence of agricultural activities or events. In N-Catch, the temporal dimension has four levels: day, month, year and simulation period. Simulation results have a daily time step, from which users can aggregate simulated results per month, per year or for the entire simulation period.

Agricultural dimension. The agricultural dimension is a key part of N-Catch, since the main purpose of simulation is to predict impacts of agricultural practices on stream water quality. Cropping system data used as input to TNT2 were stored in the DW. We define four levels from these data: crop operation, CMS, crop type and crop rotation (Fig. 5). As an example of the agricultural dimension, the crop operation "fertilization" appears in the CMS RGA_minPFP0 <pasture, fertilization, mowing>, which belongs to the crop type "IGS-grassland" in the crop rotation <corn, catch-IRG, wheat>.

3.3. Implementing the N-Catch DW

DW tiers are generally managed using a relational database in which data are structured following three main schema types: star, constellation and snowflake modeling (IBM, 1998). The first is the most common and consists of a central fact and dimensions visually represented by a star. In contrast, a constellation model merges multiple star models using common dimensions; it therefore includes several facts and common or specific dimensions. Finally, the snowflake model is an offshoot of the star model: the fact is maintained, while the dimensions are split into several tables according to their hierarchies. The snowflake model is required for flexible querying of complex dimension relations. Because this corresponds to our case study, based on a set of multilevel dimensions with complex hierarchical structures and relations, we chose the snowflake model to design N-Catch (Fig.6).

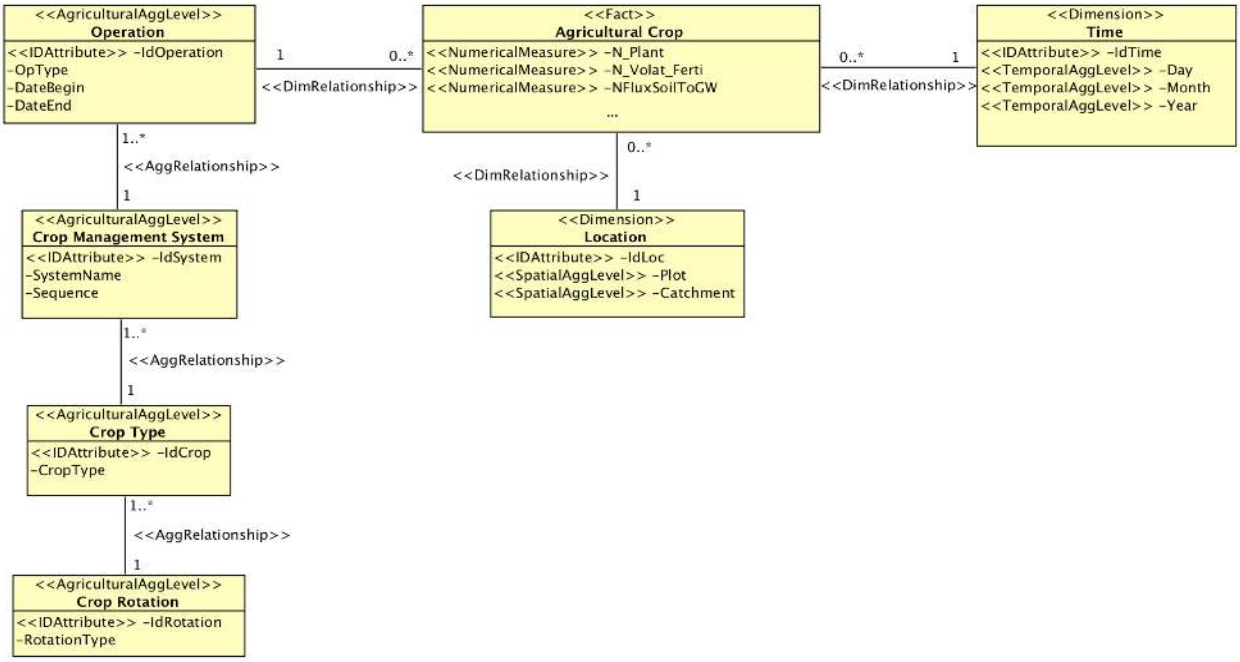


Figure 6 : Multidimensional data model of the N-Catch data warehouse

Once the preliminary steps are completed (i.e., identification of user needs, data collection, multidimensional model design), the ETL step is developed.

Extraction. The design of the data warehouse is driven by user needs and relevance of the simulation data to them. A list of example user requests was made (e.g., "Which plots and cropping systems emit the lowest amounts of N to air and groundwater?"), from which TNT2 input and output variables were identified and loaded into the data warehouse. Finally, 16 daily variables (e.g. mineral fertilizer, organic fertilizer, N volatilization, N exported by crops), three dimensions (location, time and agricultural dimension) and two flows, N emissions from soil to groundwater and to the atmosphere (calculated from the 16 selected variables), were extracted from TNT2 out-put files. Scripts written in the Perl programming language were used to extract the selected data from simulation results.

Transformation. The transformation of extracted data required fairly substantial preprocessing, especially (i) adapting the structure of cropping system input data to the four levels of the agricultural dimension in the DW and (ii) calculating nitrate and water balances from TNT2 output variables. Other types of transformations were performed on the raw data:

decoding, cleaning, normalizing, de-normalizing, harmonizing and merging heterogeneous sources.

Loading. The last step consists of loading and integrating the extracted and transformed data into the data warehouse. To perform this step, we used scripts written in the Perl programming language.

3.4. Visualization and exploration of the data stored in N-Catch

There is a clear need for tools to help end users visualize and explore geo-referenced data within large amounts of simulation results (Boulil et al., 2013; Laniak et al., 2013). Maps display information (e.g. distance between two isolated phenomena, extent of a phenomenon, shape of a phenomenon along a river bank) that would not have been revealed by using any other representation. In the case of N-Catch, end users have expressed the need to analyze whether the spatial distributions of processes show patterns among plots, hillslopes or sub-catchments and to examine the details of a specific region to compare them to those of non-neighboring region. A map used as an exploration and visualization tool becomes a decision-aid tool because it is closer to the reality of end users and requires less abstraction; this increases the effectiveness of N-Catch (*e.g.* perform spatio-temporal analysis as drilling down mapping component of the spatio-temporal data cube, identify potential spatio-temporal patterns, etc.). To this end, to exploit data to their full potential, we developed a spatial component in N-Catch to allow users to view stored simulation data on a map by querying N-Catch, which thus facilitates decision support.

Coupling cartographic components and online analysis requires new tools. Two technologies represent potential candidates: GIS and SOLAP. GIS applications process geographic information, but most of them have few or no data-analysis functions. Indeed, some GIS can be considered digital mapping systems, while analysis is performed by external software or plugins. The aim of SOLAP systems is to combine OLAP tools (e.g. decision support, graphics) with geographic tools (e.g. mapping, geographic aggregators). SOLAP systems can handle three types of spatial dimensions: i) descriptive, for which the spatial references are textual (e.g. location name); ii) geometric, for which each hierarchical level consists of a set of geometric shapes (e.g. polygons, points); and iii) mixed, a combination of the two previous dimensions, allowing for both textual and geometric references.

Processing mixed and geometric dimensions implies redefining OLAP aggregation operators for such data. In the N-Catch DW, the spatial dimension contains only two hierarchical levels (plot and catchment). Ultimately, N-Catch required only cartographic visualization of simulation results; thus, spatio-multidimensional operators were not essential, and coupling of GIS and OLAP was sufficient. We used QGIS (http://www.qgis.org), which is free GIS software shared under the GNU GPL. One of the main advantages of working with QGIS is the ability to integrate plugins that can enhance its features. To facilitate users' interactions with N-Catch and the spatial visualization tool, we used a plugin with a graphical interface that allows users to select data from N-Catch (variable, duration and location) and view them on a map. This plugin allows users to create a new map (i.e. a new data layer) and save it or to select a pre-existing map.

4. Illustrating N-Catch in a case study

4.1. Study site and simulation procedure

The TNT2 model was applied to the Yar catchment, comprising 10 sub-catchments, which discharge into the "Lieue de Grève" bay, located in Brittany, western France (Fig. 7). The Yar flows towards the English Channel and drains an area of 61.5 km², of which 8% is urban, 28% is woodlands and 64% is agricultural land, the latter comprising 4620 agricultural plots. Plot boundaries were based on an aerial survey. In 2007, 194 farms had all or part of their agricultural area in the catchment (Moreau et al., 2012b). Despite moderate nitrate concentrations around 6.8 mg $N\text{-}NO_3.l^{-1}$ (i.e., much lower than the 11.3 mg $N\text{-}NO_3.l^{-1}$ limit defined by national and European directives), the coastal bay at the catchment outlet has experienced macro-algal blooms every summer for the past 40 years. The physiographic context and the hydrodynamic conditions (confined bay, low currents) in the "Lieue de Grève" bay makes it very sensitive to coastal eutrophication. In this bay, algae proliferation is important because they are well fed by continental inputs and marine currents cannot take them offshore. Better understanding of catchment functioning requires detailed analysis of effects of current cropping systems and changes in them on N emissions at different locations and periods, i.e. different spatial and temporal levels. Data used for the case study, which correspond both to input and output of the TNT2 model, span from 1 Sep 1996 to 31 Aug 2008. For this simulation (Moreau et al., 2012b), 16 output variables per plot on 4620 plots were stored for 4380 days (i.e., 20,235,600 records with 16 data fields, with a total database size of 9 GB). Daily weather data were acquired from Météo France.

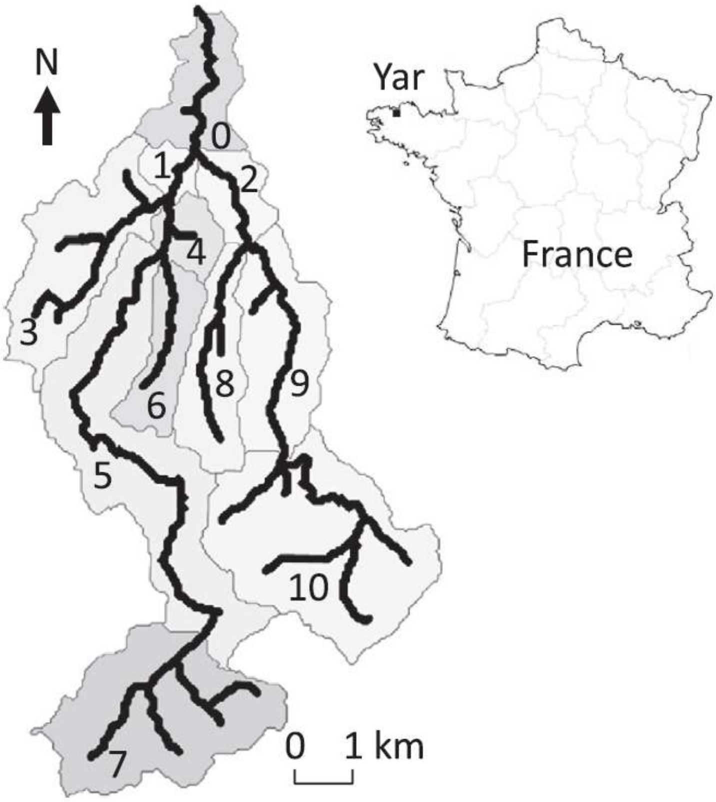


Figure 7 : Plot map of the Yar catchment, Brittany, France

4.2. N-Catch illustration

N-Catch can be used to analyze more deeply, in space and time, effects of agricultural practices on N emissions to stream water and the air. Benefits of the multidimensional structure of N-Catch are illustrated by considering each dimension, based on simulation data from a 4380-day period (i.e. 12 years), and demonstrating that N-Catch can reveal new knowledge within each dimension.

4.2.1. Temporal dimension

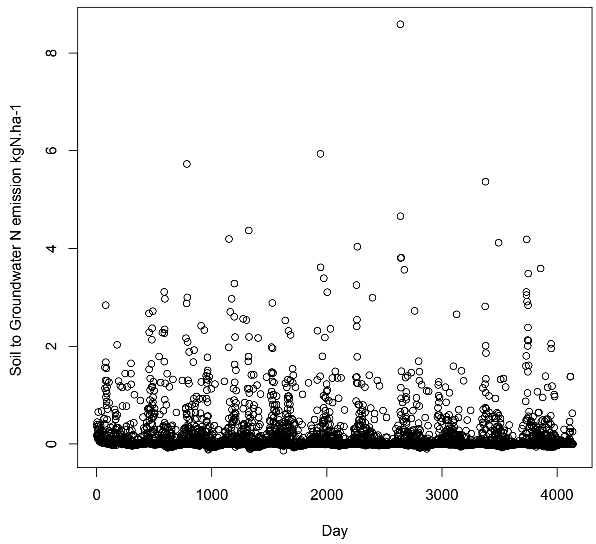
The temporal dimension facilitates data aggregation at different time steps, as well as analysis of temporal dynamics. Dynamics of daily N emissions from soil to groundwater show a seasonal trend with some outliers, while at a monthly time step, well-defined seasonal differences appear, with a net increase in autumn and more frequent outliers during the re-wetting period (Fig.8). At a yearly time step, higher values occur during the beginning of the simulation period, and differences occur among years, with higher flows during 1996-1999 and 2005-2006 due to higher precipitation.

4.2.2. Spatial dimension

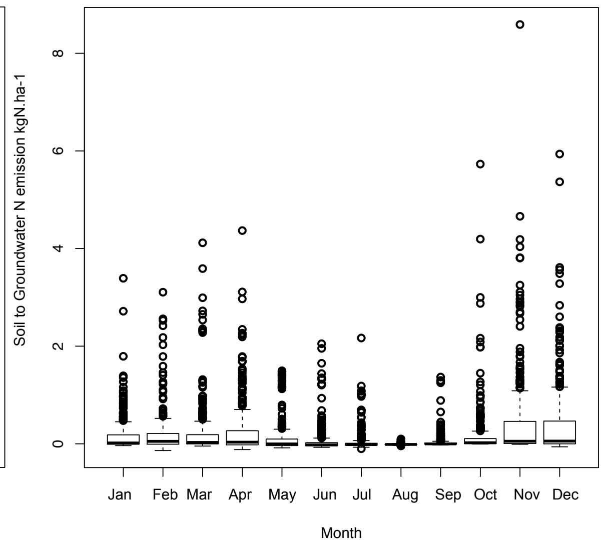
The spatial dimension can be used to analyze spatial relations between measures, for example, between those of permanent grassland (P-Grassland) and temporary grassland (T-Grassland) (Fig. 9). In this example, the aggregation function used is the sum of simulated values over time, divided by the rotation duration, to obtain mean values per year. A high positive correlation ($r = 0.8$) exists between denitrification and N emissions to air, while the correlation between N mineralization and N emissions to groundwater depends on the type of grassland: for the same amount of mineralization, N emissions to groundwater are much lower in permanent grasslands than in temporary grasslands (Fig. 9). A slight positive correlation ($r = 0.3$) is observed between N emissions to groundwater and N mineralization; however, the higher the amount of mineralization, the more variable are the N emissions to groundwater. No correlation was found between fertilization and N emissions from soil to groundwater, but above a certain amount of N fertilization (76 kg $N.ha^{-1}.y^{-1}$), variability in N emissions to groundwater among plots increases (Fig. 9c). The correlation between N emissions from soil to groundwater and that to air is again scattered around two groups according to land cover (Fig. 9d). For the same N emission from soil to air, N emission from soil to groundwater is much lower in permanent grasslands than in temporary grasslands. Such analysis is interesting and has never been performed in previous applications of TNT2.

4.2.3. Agricultural dimension

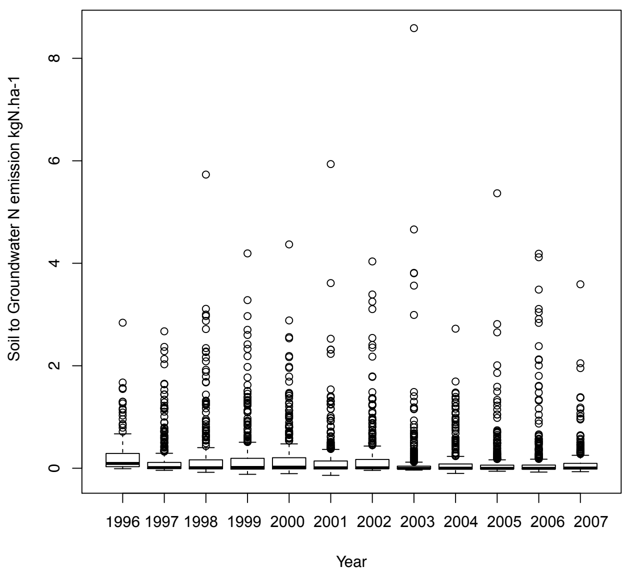
Considering all plots in the catchment and analyzing daily variations in N emissions from soil to groundwater reveals complex relations between agricultural practices and nitrate leaching at the plot scale. The aggregation function used in this case corresponds to mean daily N emission from soil to groundwater per month across the catchment. Differences among crops are small (Fig. 10a), but they do exist, such as those between wheat and corn. Intra-crop variability is surprisingly high. For example, half of the monthly N emissions from soil to groundwater for potatoes vary strongly. This variability can be due to climate, topographic position within the catchment or crop management. Effects of agricultural practices can be seen in variations in mean daily N emission from soil to groundwater per CMS, which integrates crop type and crop operations at the catchment level (Fig. 10b). Differences are much smaller among CMSs than among crop types, which indicates that crops must be considered in their agricultural and environmental contexts. The type of CMS particularly influences the presence of outliers. N-Catch is particularly effective at analyzing variability in effects of crop management, identifying high and low values and searching for factors that explain them, and ultimately proposing recommendations or mitigation options.



(a)



(b)



(c)

Figure 8 : Simulated temporal dynamics of N emissions from soil to groundwater a) per day, b) per month, and c) per year based on a 12- year time series

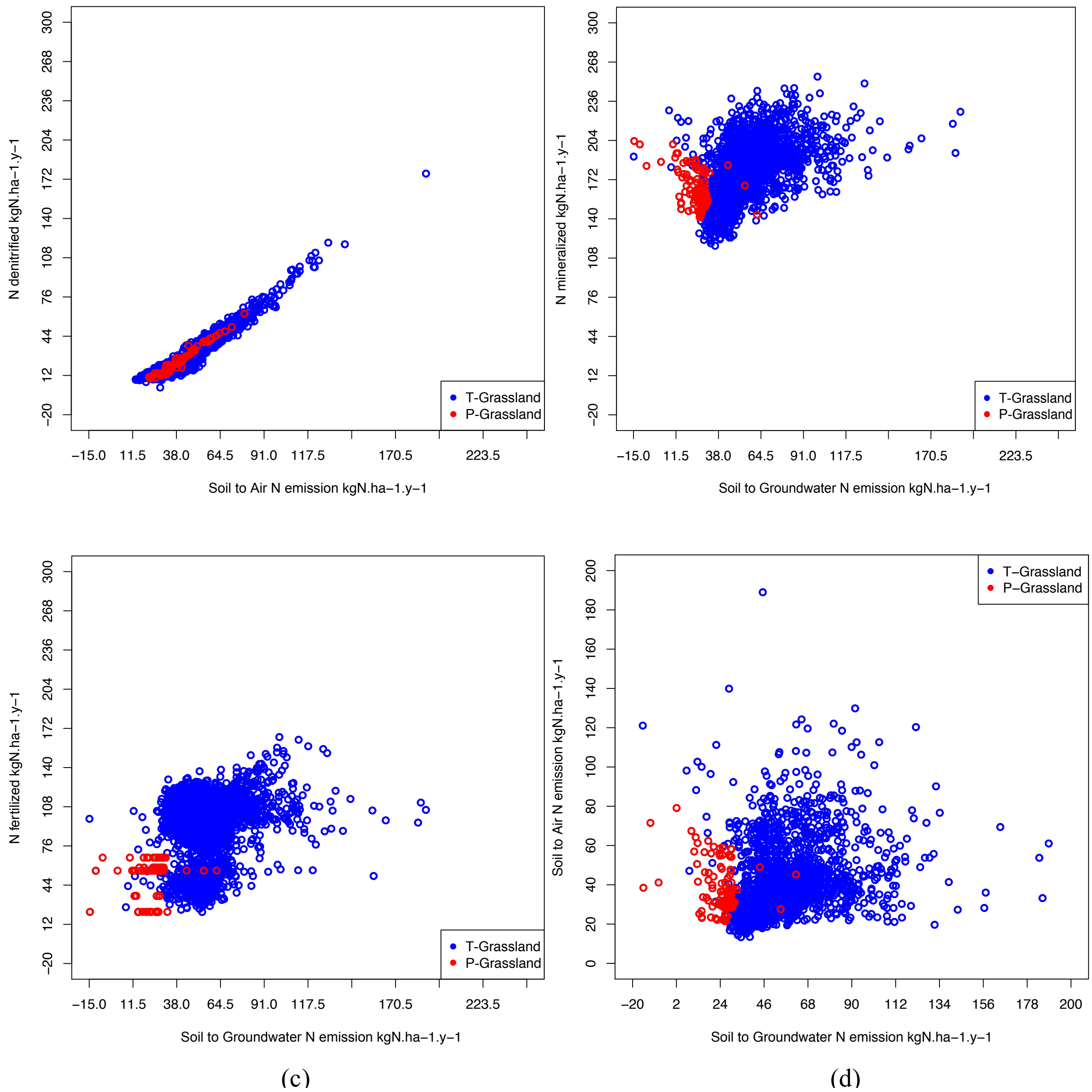
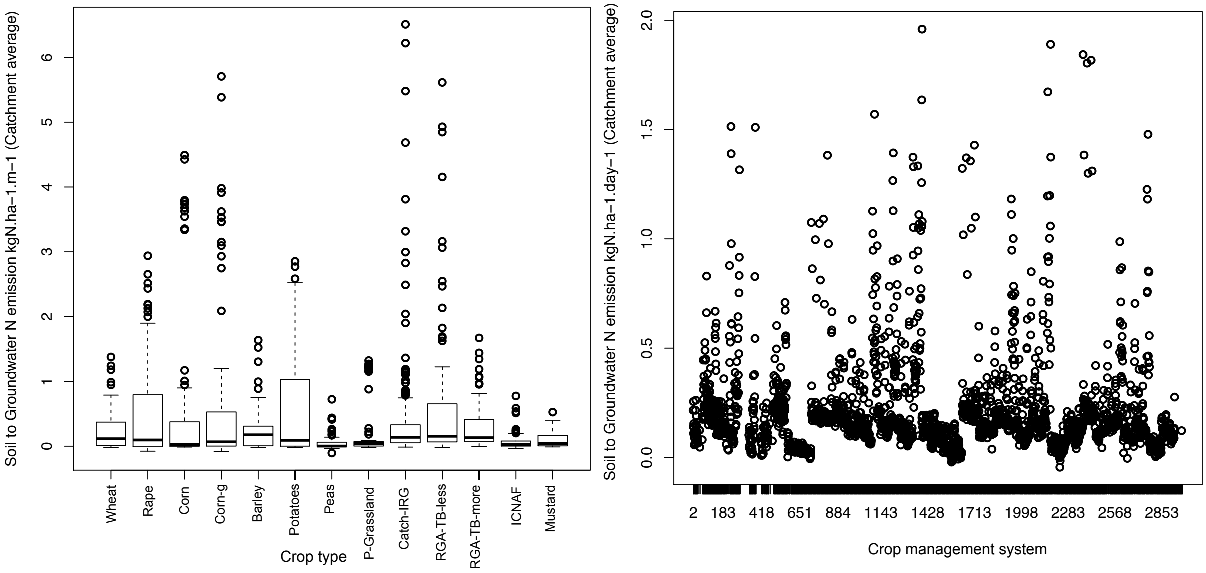


Figure 9 : Process analysis at the plot scale: a) N denitrification versus N emission from soil to air; b) N mineralization versus N emission from soil to groundwater; c) N fertilization versus N emission from soil to groundwater; d) N emission from soil to groundwater versus N emission from soil to air

4.2.4. Coupling N-Catch with QGIS

In the previous examples, displaying plots in a point cloud identified the most polluting plots in the Yar catchment; however, plot ID numbers provide limited information. Users should be able to visualize these plots on a map, for example, to know their topographic position in a catchment (i.e. upper, middle or lower part of the hillslope) and their extent in the catchment. Simple visualization of this spatial query helps users understand the phenomenon in question (i.e. nitrate pollution) by seeing its position and extension within a geographic frame of reference.



(a) (b)

Figure 10 : Effect of agricultural practice dynamics on simulated N emission from soil to groundwater, based on a 12-year time series, for temporary grass-clover grassland with <40% white clover (*RG-WC-less*) or >40% white clover in aboveground biomass (*RG-WC-more*), averaged over the entire catchment and expressed a) per month or b) per crop management system.

For example, one can visualize the sum of daily N emission from soil to groundwater at the plot level for the hydrological year 1997-1998 (Fig. 11) or the mean daily N emission from soil to air for the corn crop during 2003 (Fig. 12). In the former, the highest amounts of N emitted to groundwater are frequently located in patches, probably by farm. In the latter, the highest amounts of N emitted to air are scattered throughout the catchment. This kind of visual display can help users interact with the model and develop hypotheses about catchment functioning and remediation.

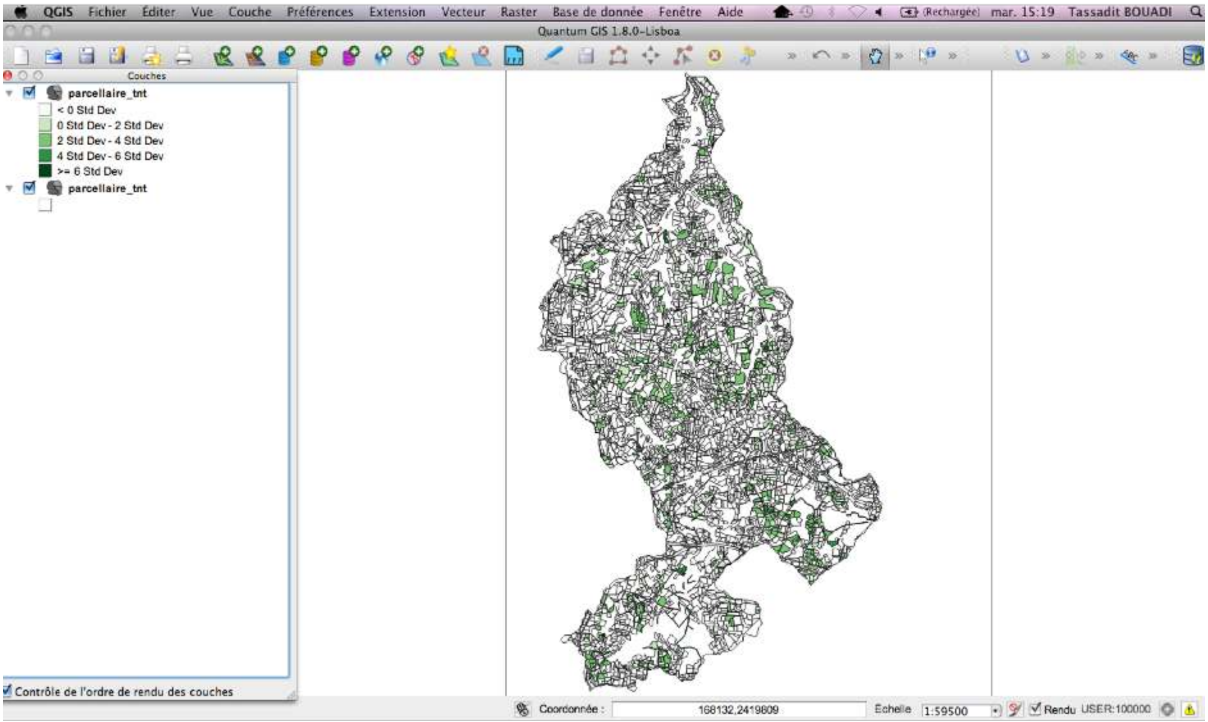


Figure 11: Sum of daily N emission from soil to groundwater at the plot level for the hydrological year 1997-1998 in the Yar catchment, Brittany, France

5. Discussion

5.1. Purpose, scope and generality of the N-catch DW

N-Catch was specifically built to store simulated data from the TNT2 model, better analyze effects of agricultural practices on the landscape-level N cycle at multiple spatiotemporal levels and provide strategic information for decision making.

In table 2, we present some characteristics of the storage size of the N-catch data warehouse. The storage size of the N-Catch DW is increasing with the number of: plots, agricultural operations and days. This is due to its fact table size in *O(p*g*d)* (*p* is the number of plots, *g* the number of agricultural operations, and *d* the number of simulation days) which induces a substantial increase of the storage size of N-Catch. Also, in order to speed up the execution time of queries and data retrieval, we used: (i) database indexes (*i.e.* data structures used to quickly locate and access the data in a database table) to improve execution time of

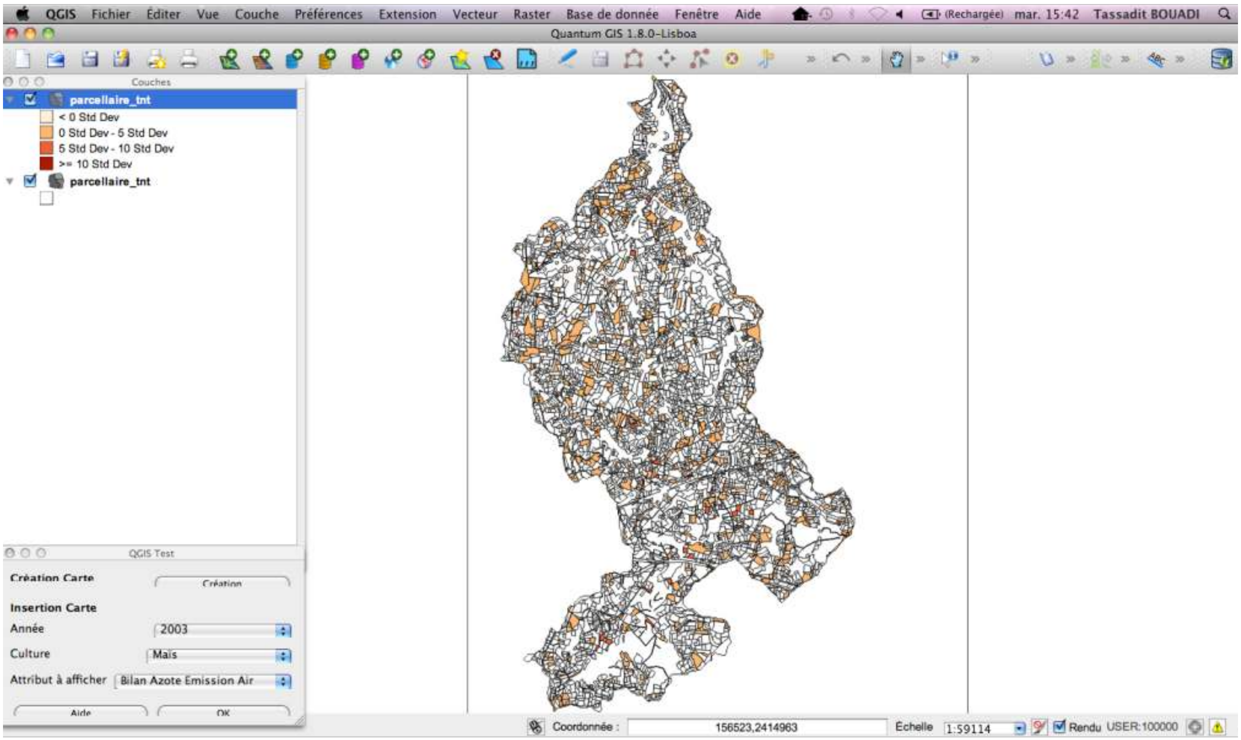


Figure 12: Mean daily N emission from soil to air for maize in 2003 at the plot level in the Yar catchment, Brittany, France

queries by a factor of 5, to vary between 0,01 seconds to 0,4 seconds (a complex query template leads to a significant increase in the query execution time), and (ii) materialized views to improve query performance by pre-calculating complex query templates (*e.g.* expensive join and aggregation operations) and storing the results in the data warehouse.

Stored data	Description
TNT2 inputs/outputs	9000 processed files (8,9 GB)
N-Catch size	9 GB
Database indexes	2,1 GB
Materialized views	2,7 GB

Table 2: Storage size of the N-Catch data warehouse

The main methodological contributions of this work include methods to (i) preprocess, transform and load simulated data to DW; (ii) describe agricultural practices in an original manner by representing hierarchies in cropping systems; and (iii) analyze simulation results by combining spatiotemporal simulation data, data warehousing and online analysis. The QGIS plugin associated with this DW architecture allows end users to visualize and explore geo-referenced data, which has been identified as an important need (Boulil et al., 2013, 2014).

The N-Catch application can be used effectively to explore spatiotemporal dimensions of simulation data; analyze agro-hydrological processes and extract relations between measures and dimensions; provide finer analysis to help with decision making (i.e. improve decision makers' access to data); and summarize environmental information and understand N emissions by using comparative and personalized views of archived data. These methodological contributions can be applied to a variety of agro-environmental issues and situations. More generally, this approach can be used to analyze effects of agricultural practices on water quantity and quality as well as other environmental impacts. Generic steps include identification of user needs, extraction of selected data from the set of simulation data, dimensional modeling, and use of OLAP to access and exploit multidimensional and aggregated data.

This kind of DW tool is particularly effective when common requests can be defined for big-data environments (i.e. when high data fluxes are involved and explored by infrequent and pre-defined requests), which is usually the case for agro-environmental data generated by simulations. Simulation generates numerous data, while queries are often small but sophisticated. This could become even more relevant when using multiple sets of simulation data, in which they would be considered as a set of potential future realities of the catchment (e.g. various land uses, agricultural management practices, and climate scenarios).

Such a set of simulations, whether for the case study or any other agro-environmental system, would benefit from the tools developed. For such a purpose, the original description of the multiple levels of the agricultural dimension could be reused, since it is completely generic. As basic dimensions of any spatially and explicitly distributed model, the spatial and temporal dimensions are therefore easily transposable to any model simulating at landscape or regional levels.

5.2. Limits and potential improvements of the N-Catch DW

The spatial data within N-Catch was limited to plot and catchment levels. This is a limitation for users who exploit the spatial dimension, since TNT2 considers several spatial granularities (e.g. hillslope, riparian areas, sub-catchment, farm, soil type, topographic indexes). It would be useful to increase the number of hierarchies in the spatial dimension and provide users with new spatial navigation capabilities. Integrating a SOLAP system in the DW architecture, in which spatial data (e.g. geometry, topology, description) are represented explicitly within the dimension, could be one way to improve the approach within N-Catch.

N-Catch should be expanded with data-mining or information-retrieval methods, such as skyline queries, to perform advanced analyses to find meaningful patterns and relations in the simulation data (Bouadi et al., 2012). Data-mining techniques are powerful tools for identifying and extracting interesting knowledge from large data collections. To extend OLAP analysis from simple aggregate operations on data cubes for prediction or decision objectives, many studies investigated the combination of OLAP analysis with data mining. Three main approaches are found in the literature: i) modifying OLAP operators (Bentayeb and Favre, 2009; Chen et al., 2000; Goil and Choudhary, 2001; Han, 1997; Sarawagi, 1999; Sathe and Sarawagi, 2001) to simulate data-mining techniques; (ii) adapting multidimensional structures (Chen et al., 2001; Goil and Choudhary, 2001; Pinto et al., 2001) by reorganizing the multidimensional data to make them usable with data-mining methods; and (iii) adapting data-mining algorithms (Bodin-Niemczuk et al., 2008; Giacometti et al., 2008; Sarawagi et al., 1998) to cope with the multidimensional data environment. Coupling data warehousing and skyline queries allows users to navigate along dimensions' hierarchies (i.e. specialize/generalize) while ensuring online calculation of the skyline. Bouadi et al. (2014) developed an efficient approach for simulating the effect of the OLAP "drill-down" operator on the computation of skyline queries. Further discussions with agronomists and agro-hydrologists will test and illustrate their uses, but we can already say that they enable queries to be formulated in the DW by combining conflicting environmental indicators and finding compromise solutions associated with these requests that meet stakeholders' expectations. Subsequently, it would be interesting to design a user-friendly interface for N-Catch that provides users with centralized access to all the functions developed (e.g. OLAP analysis, skyline queries).

N-catch was applied for specific results of TNT2 simulation of the Yar catchment in an offline manner. One challenge would be to couple N-catch and TNT2 online. In this way, simulation results of TNT2 could be loaded automatically into the DW. This integration can be achieved by using tools for manipulating data (Chuffart et al., 2010; Laniak et al., 2013) to manage simulation runs and process and load model output into the DW. To automate creation of the DW, one approach could be to not set an "a priori" process for the DW, but to use algorithms to automatically design DW schemas and OLAP cubes. Sautot et al. (2015) developed such an approach by using hierarchical agglomerative clustering with a metric (similarity index) to automatically build hierarchical dimensions in an OLAP cube. With this similarity index, one can perform hierarchical clustering in heterogeneous datasets that contain qualitative and quantitative variables. Their approach, based on data-mining methods, can supplement expert knowledge during the design of an OLAP cube. With this method, one could build new dimensions based on hierarchies in the data that are initially hidden.

The objectives of such developments would be to completely transform the use of integrated models, which are generally complex and only developed and used in a research context. The general idea is that such models could be used by end users, such as consultants or knowledgeable laypeople, to acquire and transfer the knowledge obtained from the models. An alternative could be to keep such models, as well the corresponding simulation data and responsibility for the procedure, in the research domain and to make the simulation results widely accessible to the larger public through interface and exploration tools, as is the case for environmental data. This development could be considered the first step for a framework leading to an automatic and user-friendly procedure to explore not the model itself, but the model results.

6. Conclusion

A DW is expected to (i) provide dynamic multidimensional analysis, supporting end users with analytical and navigational properties, and (ii) offer a rapid response to queries, regardless of the DW's size and complexity. Such a framework was developed and applied to quantify impacts of agricultural practices on N emissions at multiple spatiotemporal levels by providing relevant ways to identify and analyze where and why N pollution is present in a catchment. The method developed shows how OLAP and the data-cube concept can be useful for analyzing the huge amount of data produced by modeling activities. This design was applied by developing the N-Catch DW, which was built to store and manage simulation data from the agro-hydrological model TNT2, and was illustrated in a case study. The N-Catch DW allows users to explore N emissions in space and time, to more accurately analyze transfer and transformation processes as a function of cropping systems, and to obtain new knowledge that facilitates making specific and detailed decisions in space and time. The approach adopted for developing N-Catch is not specific to this case study. It can be applied to explore the functioning of any agro-environmental system. In particular, the spatiotemporal modeling and agricultural dimension are generic to all models with landscape and regional dimensions. Instead of transferring models, the future of agro-environmental modeling could be to allow users to easily navigate the simulation data of numerous models that were developed to help make decisions about effects of human activities on the environment.

Acknowledgments

This research was supported by grants from the ANR Systerra project ACASSYA (ANR-08-STRA-01). Special thanks go to Patrick Durand for providing the opportunity to use the TNT2 model, and to Sylvain Dousset for QGIS-NCatch coupling.

References

Abdullah, A. (2009). Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool. Computers and Electronics in Agriculture, 69(1):59–72 http://dx.doi.org/10.1016/j.compag.2009.07.003

Beaujouan, V., Durand, P., and Ruiz, L. (2001). Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. Ecological Modelling, 137(1):93–105 http://dx.doi.org/10.1016/S0304-3800(00)00435-X

Beaujouan, V., Durand, P., Ruiz, L., Aurousseau, P., and Cotteret, G. (2002). A hydrological model dedicated to topography-based simulation of nitrogen transfer and transformation: rationale and application to the geomorphology denitrification relationship. Hydrological Processes, 16(2):493–507 http://dx.doi.org/10.1002/hyp.327

Benhamou, C., Salmon-Monviola, J., Durand, P., Grimaldi, C., and Merot, P. (2013). Modeling the interaction be-tween fields and a surrounding hedgerow network and its impact on water and nitrogen flows of a small watershed. Agricultural Water Management, 121:62–72 http://dx.doi.org/10.1016/j.agwat.2013.01.004

Bentayeb, F. and Favre, C. (2009). ROK: Roll-Up with the K-Means clustering method for recommending OLAP queries. In Proceedings of the International Conference on Database and Expert Systems Applications, pages 501–515 http://dx.doi.org/10.1007/978-3-642-03573-9_43

Berrahou, L., Lalande, N., Serrano, E., Molla, G., Berti-Équille, L., Bimonte, S., Bringay, S., Cernesson, F., Grac, C., Ienco, D., et al. (2015). A quality-aware spatial data warehouse for querying hydroecological data. Computers & Geosciences, 85:126–135 http://dx.doi.org/10.1016/j.cageo.2015.09.012

Beven, K. (1997). Topmodel: a critique. Hydrological Processes, 11(9):1069–1085 http://dx.doi.org/10.1002/(SICI)1099-1085(199707)11:9<1069::AID-HYP545>3.0.CO;2-O

Bimonte, S., Miquel, M. (2010). When spatial analysis meets OLAP: multidimensional model and operators. International Journal of Data Warehousing and Mining. 6 (4), 33–60 http://dx.doi.org/10.4018/jdwm.2010100103

Bodin-Niemczuk, A., Ben Messaoud, R., Loudcher Rabaséda, S., and Boussaid, O. (2008). Vers l'intégration de la prédiction dans les cubes OLAP. In Proceedings of the confrence Extraction et Gestion des Connaissances, pages 203–204.

Bouadi, T., Cordier, M.-O., and Quiniou, R. (2012). Incremental computation of skyline queries with dynamic preferences. In Proceedings of the International Conference on Database and Expert Systems Applications, pages 219–233 http://dx.doi.org/10.1007/978-3-642-41221-9_2

Bouadi, T., Cordier, M.-O., and Quiniou, R. (2014). Computing hierarchical skyline queries "on-the-fly" in a data warehouse. In Proceedings of the International Conference on Data Warehousing and Knowledge Discovery, pages 146–158 http://dx.doi.org/10.1007/978-3-319-10160-6_14

Boulil, K., Le Ber, F., Bimonte, S., Grac, C., and Cernesson, F. (2014). Multidimensional modeling and analysis of large and complex watercourse data: an olap-based solution. Ecological Informatics, 24:90–106 http://dx.doi.org/10.1016/j.ecoinf.2014.07.001

Boulil, K., Pinet, F., Bimonte, S., Carluer, N., Lauvernet, C., Cheviron, B., Miralles, A., and Chanet, J.-P. (2013). Guaranteeing the quality of multidimensional analysis in data warehouses of simulation results: application to pesticide transfer data produced by the macro model. Ecological Informatics, 16:41–52 http://dx.doi.org/10.1016/j.ecoinf.2013.04.004

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J., Meynard, J., and Delécolle, R. (1998). STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie, 18(1):311–346 http://dx.doi.org/10.1051/agro:19980501

Chaudhuri, S. and Dayal, U. (1997). An overview of data warehousing and OLAP technology. ACM Sigmod record, 26(1):65–74 http://dx.doi.org/10.1145/248603.248616

Chen, M., Zhu, Q., and Chen, Z. (2001). An integrated interactive environment for knowledge discovery from heterogeneous data resources. Information & Software Technology, 43(8):487–496 http://dx.doi.org/10.1016/S0950-5849(01)00159-8

Chen, Q., Dayal, U., and Hsu, M. (2000). An olap-based scalable web access analysis engine. In Proceedings of the International Conference on Data Warehousing and Knowledge Discovery, pages 210–223 http://dx.doi.org/10.1007/3-540-44466-1_21

Chuffart, F., Dumoulin, N., Faure, T., Deffuant, G., (2010). SimExplorer: programming experimental design on models and managing quality of modelling process. International Journal of Agricultural and Environmental Information Systems 1, 55–68.

Durand, P., Ferchaud, F., Salmon-Monviola, J., Goetschel, F., and Martin, C. (2006). Etude sur l'évolution des paramètres nitrates dans les eaux brutes des bassins versants bretagne eau pure et des autres bassins versants bretons. Démarche Evaluation Programme Bretagne Eau Pure, Rapport INRA, INRA, Paris.

Durand P., Moreau P., Salmon-Monviola J., Ruiz L., Vertes F., Gascuel-Odoux C. (2015). Modelling the interplay between nitrogen cycling processes and mitigation options in farming catchments. Journal of Agricultural Science, 153: 959 http://dx.doi.org/10.1017/s0021859615000258

Ferrant, S., Bustillo, V., Burel, E., Salmon-Monviola, J., Claverie, M., Jarosz, N., Yin, T., Rivalland, V., Dedieu, G., Demarez, V., Ceschia, E., Probst, A., Al-Bitar, A., Kerr, Y., Probst, J.-L., Durand, P., Gascoin, S. (2016). Extracting Soil Water Holding Capacity Parameters of a Distributed Agro-Hydrological Model from High Resolution Optical Satellite Observations Series. Remote Sensing http://dx.doi.org/10.3390/rs8020154

Ferrant, S., Gascoin, S., Veloso, A., Salmon-Monviola, J., Claverie, M., Rivalland, V., Dedieu, G., Demarez, V., Ceschia, E., Probst, J.-L., Durand, P., and Bustillo, V (2014). Agro-hydrology and multi-temporal high-resolution remote sensing: toward an explicit spatial processes calibration, Hydrology and Earth System Sciences, 18, 5219-5237 http://dx.doi.org/10.5194/hess-18-5219-2014.

Ferrant, S., Oehler, F., Durand, P., Ruiz, L., Salmon-Monviola, J., Justes, E., Dugast, P., Probst, A., Probst, J.-L., and Sanchez-Perez, J.-M. (2011). Understanding nitrogen transfer dynamics in a small agricultural catchment: comparison of a distributed (TNT2) and a semi distributed (SWAT) modeling approaches. Journal of Hydrology, 406(1):1–15 http://dx.doi.org/10.1016/j.jhydrol.2011.05.026

Giacometti, A., Marcel, P., and Negre, E. (2008). A framework for recommending OLAP queries. In Proceedings of the ACM international workshop on data warehousing and OLAP, pages 73–80 http://dx.doi.org/10.1145/1458432.1458446

Goil, S. and Choudhary, A. (2001). PARSIMONY: An infrastructure for parallel multidimensional analysis and data mining. Journal of Parallel and Distributed Computing, 61(3):285–321 http://dx.doi.org/10.1006/jpdc.2000.1691

Gray, J., Bosworth, A., Layman, A., and Pirahesh, H. (1996). Data Cube: A relational aggregation operator generalizing Group-By, cross-tab, and Sub-Total. In Proceedings of the International Conference on Data Engineering, pages 152–159 http://dx.doi.org/10.1023/A:1009726021843

Han, J. (1997). OLAP mining: An integration of OLAP with data mining. In Proceedings of the Working Conference on Database Semantics, pages 1–9 http://dx.doi.org/10.1007/978-0-387-35300-5_1

Hénault, C. and Germon, J. C. (2000). NEMIS, a predictive model of denitrification on the field scale. European Journal of Soil Science, 51(2):257–270 http://dx.doi.org/10.1046/j.1365-2389.2000.00314.x

IBM (1998). Data modeling techniques for data warehousing. IBM Corporation International Technical Support Organization.

Inmon, W.H. (2005). Building the data warehouse. John wiley & sons.

Isaacson, D. L., & Madsen, R. W. (1976). Markov chains, theory and applications (Vol. 4). New York: Wiley.

Kimball, R. (1996). The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley & Sons.

Laniak, F.G., Olchin, G., Goodall, J., et al. (2013). Integrated environmental modelling: a vision and roadmap for the future. Environmental Modelling and Software. 39, 3-23 http://dx.doi.org/10.1016/j.envsoft.2012.09.006

Leenhardt, D., Angevin, F., Biarnès, A., Colbach, N., and Mignolet, C. (2010). Describing and locating cropping systems on a regional scale. A review. Agronomy for Sustainable Development, 30(1):131–138 http://dx.doi.org/10.1051/agro/2009002

Liao, W., van der Werf, H. M. G., Salmon-Monviola, J. (2015). Improved Environmental Life Cycle Assessment of Crop Production at the Catchment Scale via a Process-Based Nitrogen

Simulation Model. Environmental Science & Technology 49(18), 10790-10796 http://dx.doi.org/10.1021/acs.est.5b01347

Mahboubi, H., Bimonte, S., Faure, T., Pinet, F. (2010). Data warehouse and OLAP for environmental simulation data. International Journal of Agricultural and Environmental Systems 1, 1–19

Moreau, P., Ruiz, L., Mabon, F., Raimbault, T., Durand, P., Delaby, L., Devienne, S., and Vertès, F. (2012b). Reconciling technical, economic and environmental efficiency of farming systems in vulnerable areas. Agriculture, Ecosystems & Environment, 147: 89–99 http://dx.doi.org/10.1016/j.agee.2011.06.005

Moreau, P., Ruiz, L., Raimbault, T., Vertès, F., Cordier, M., Gascuel-Odoux, C., Masson, V., Salmon-Monviola, J., and Durand, P. (2012a). Modeling the potential benefits of catch-crop introduction in fodder crop rotations in a Western Europe landscape. Science of the Total Environment, 437: 276–284 http://dx.doi.org/10.1016/j.scitotenv.2012.07.091

Moreau, P.; Viaud, V.; Parnaudeau, V.; Salmon-Monviola, J.; Durand, P. (2013). An approach for global sensitivity analysis of a complex environmental model to spatial inputs and parameters: A case study of an agro-hydrological model. Environmental Modelling Software, 47, 74–87 http://dx.doi.org/10.1016/j.envsoft.2013.04.006

Nilakanta, S., Scheibe, K., and Rai, A. (2008). Dimensional issues in agricultural data warehouse designs. Computers and Electronics in Agriculture, 60(2):263–278 http://dx.doi.org/10.1016/j.compag.2007.09.009

O'Callaghan, J.F. and Mark, D.M. (1984) The Extraction of Drainage Networks from Digital Elevation Data. Computer Vision, Graphics and Image Processing, 28, 323-344.

Oehler, F., Durand, P., Bordenave, P., Saadi, Z., and Salmon-Monviola, J. (2009). Modelling denitrification at the catchment scale. Science of The Total Environment, 407(5):1726–1737 http://dx.doi.org/10.1016/j.scitotenv.2008.10.069

Pinet, F., Miralles, A., Bimonte, S., Vernier, F., Carluer, N., Gouy, V., and Bernard, S. (2010). The use of UML to design agricultural data warehouses. In Proceedings of the International Conference on Agricultural Engineering.

Pinto, H., Han, J., Pei, J., Wang, K., Chen, Q., and Dayal, U. (2001). Multi-dimensional sequential pattern mining. In Proceedings of the International Conference on Information and Knowledge Management, pages 81–88 http://dx.doi.org/10.1145/502585.502600

Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., van Griensven, A., van der Zee, S.E.A.T.M. (2010). New challenges in integrated water quality modeling. Hydrological Processes, 24, 3447−3461 http://dx.doi.org/10.1002/hyp.7766

Salmon-Monviola, J., Durand, P., Ferchaud, F., Oehler, F., and Sorel, L. (2012). Modelling spatial dynamics of cropping systems to assess agricultural practices at the catchment scale. Computers and electronics in agriculture, 81:1–13 http://dx.doi.org/10.1016/j.compag.2011.10.020

Salmon-Monviola, J., Moreau, P., Benhamou, C., Durand, P., Merot, P., Oehler, F., and Gascuel-Odoux, C. (2013). Effect of climate change and increased atmospheric co2 on hydrological and nitrogen cycling in an intensive agricultural headwater catchment in western France. Climatic change, 120(1-2):433–447 http://dx.doi.org/10.1007/s10584-013-0828-y

Sarawagi, S. (1999). Explaining differences in multidimensional aggregates. In Proceedings of the International Conference on Very Large Data Bases, pages 42–53

Sarawagi, S., Agrawal, R., and Megiddo, N. (1998). Discovery-driven exploration of OLAP data cubes. In Proceedings of the International Conference on Extending Database Technology: Advances in Database Technology, pages 168–182 http://dx.doi.org/10.1007/BFb0100984

Sathe, G. and Sarawagi, S. (2001). Intelligent rollups in multidimensional OLAP data. In Proceedings of the International Conference on Very Large Data Bases, pages 531–540

Sautot, L., Faivre, B., Journaux, L., Molin, P. (2015). The hierarchical agglomerative clustering with Gower index: A methodology for automatic design of OLAP cube in ecological data processing context. Ecological Informatics 26 (2) 217-230 http://dx.doi.org/10.1016/j.ecoinf.2014.07.011

Vernier, F., Miralles, A., Pinet, F., Carluer, N., Gouy, V., Molla, G., and Petit, K. (2013). Eis pesticides: An environmental information system to characterize agricultural activities and calculate agro-environmental indicators at embedded watershed scales. Agricultural Systems, 122:11–21 http://dx.doi.org/10.1016/j.agsy.2013.07.005

Viaud, V., Durand, P., Merot, P., Sauboua, E., and Saadi, Z. (2005). Modeling the impact of the spatial structure of a hedge network on the hydrology of a small catchment in a temperate climate. Agricultural Water Management, 74(2):135–163 http://dx.doi.org/10.1016/j.agwat.2004.11.010

Wellen, C., Kamran-Disfani, A.R., Arhonditsis, G.B. (2015). Evaluation of the Current State of Distributed Watershed Nutrient Water Quality Modeling. Environmental Science Technology 49(6) 3278-3290 http://dx.doi.org/10.1021/es5049557