



**HAL**  
open science

## 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

### ► To cite this version:

Tassadit Bouadi, Marie-Odile Cordier, Pierre Moreau, René Quiniou, Jordy Salmon-Monviola, et al.. A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management. *Environmental Modelling and Software*, 2017, 97, pp.229 - 242. 10.1016/j.envsoft.2017.07.019 . hal-01597840

**HAL Id: hal-01597840**

**<https://inria.hal.science/hal-01597840v1>**

Submitted on 29 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## 16 **Abstract**

17

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

32

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

35

36

37

## 38 **Software Availability**

39 Name of software : N-Catch (Nitrogen in Catchment data warehouse)

40 Developers : Tassadit Bouadi and Sylvain Dousset

41 Contact : tbouadi@irisa.fr

42 Hardware required : Experiments were performed on an Intel Core i7 CPU at 2.8 GHz and 16  
43 GB of RAM on a Mac OSX platform

44 Software required : The relational database management system MySQL, Quantum GIS (QGIS)  
45 and Microsoft Windows, Mac OSX or Linux operating system

46 Program language : Perl, Python

47 Software availability : source code can be provided through collaborative arrangements

## 48 **1. Introduction**

49

50 Agro-hydrological models have been used extensively by researchers and stakeholders as  
51 the scientific basis for environmental management by estimating nonpoint-source pollution,  
52 identifying source areas, predicting effects of climate and land-use changes and testing the  
53 efficiency of mitigation plans to improve water quality at the catchment level (Ferrant et al.,  
54 2014; Rode et al., 2010; Wellen et al., 2015). Extensive research has focused on improving the  
55 ability of these models to consider the heterogeneity of structures and processes within  
56 agricultural landscapes by representing their spatial distributions.

57 These models generate a large volume of spatiotemporal results of various formats and  
58 semantics. Generally, only daily flux and concentrations at the catchment outlet are analyzed,  
59 even though finer-grained variable, temporal or spatial resolutions are potentially available.  
60 The main reason for this is a technological barrier that prevents efficient data processing. To  
61 address this issue, efficient tools are needed to store, display and analyze this spatiotemporal  
62 information and turn it into useful knowledge that enables better understanding of agro-  
63 environmental systems and adaptations to meet environmental targets.

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

77 To this end, recent studies have shown how a data warehouse (DW) and On Line  
78 Analytical Processing (OLAP) technologies are used to analyze environmental simulations  
79 (Boulil et al., 2013; Mahboubi et al., 2010). A DW is a subject-oriented, integrated, time-  
80 variant, non-volatile collection of data that supports the management decision-making  
81 processes (Chaudhuri and Dayal, 1997; Immon, 2005). DWs are emerging as a key

82 technology for organizations seeking to use their data to keep track of activities and improve  
83 data analysis. DWs are used (i) to provide access to massive data accumulated over time from  
84 many sources and in various formats (computer files, traditional databases, text documents,  
85 etc.) and (ii) to support multi-dimensional data analysis to make strategic and tactical  
86 decisions. DW users can extract trends and variability from the data according to various  
87 criteria to better support decision-making or to discover hidden information.

88 Few DWs have been developed in the agro-environmental sciences (Abdullah, 2009; Boulil  
89 et al., 2013, 2014; Nilakanta et al., 2008; Pinet et al., 2010). They support analyses of  
90 agricultural data along different dimensions. Nilakanta et al. (2008) developed the National  
91 Agricultural Resources Information System DW for the Indian agricultural sector. This DW  
92 provides strategic and periodic information to researchers and planners to facilitate decision  
93 making. Abdullah et al. (2009) developed an OLAP tool, ADSS-OLAP, to analyze mealybug  
94 incidence on cotton crops. The dimensional model of ADSS-OLAP includes different  
95 dimensions for analysing the effect of climate, pesticide and geography. These dimensions are  
96 aggregated as a logical OLAP cube, with a classical multidimensional model. Pinet et al.  
97 (2010) considered ways to use the Unified Modeling Language to design agricultural DWs.  
98 They presented a method for designing a DW and applied it to analyze spatial impacts of  
99 pesticide use in agriculture. Boulil et al. (2013) developed an OLAP system to store and  
100 analyze pesticide transfer data generated at the soil-column scale by a model called MACRO,  
101 to validate the model and compare results of different versions of the model. Boulil et al.  
102 (2014) then applied the OLAP system to analyze data on stream water quality. The  
103 architecture was extended with complex aggregate functions used to define indicators.

104 Recent studies have investigated Spatial OLAP (SOLAP) to study stream water quality in  
105 rivers. SOLAP systems integrate advanced OLAP and Geographic Information Systems (GIS)  
106 functions in a unique framework in which explicit representation of the spatial dimension  
107 allows users to visualize query results on maps and to use topological, metrical and directional  
108 operators when “slicing” multidimensional data (Bimonte and Miquel, 2010). Vernier et al.  
109 (2013) considered a SOLAP system to characterize agricultural activities and calculate agro-  
110 environmental indicators. Similarly, Berrahou et al. (2015) developed a solution that  
111 facilitates spatiotemporal analysis of hydroecological data by considering different levels of  
112 data quality within the system. These studies show that agro-environmental DWs are rarely  
113 developed. The most likely explanation for this is the difficulty in collecting field data in agro-  
114 environmental sciences, since data collection remains a slow and expensive process. In the  
115 past, DWs and OLAP systems were used mainly to analyze observed data. Their use for

116 simulation data is poorly developed, particularly in the field of distributed models, despite  
117 their potential utility.

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

134

## 135 **2. Main concepts of DWs and OLAP technologies**

136

### 137 **2.1. A DW as a multidimensional and hierarchical data model**

138

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

144 Facts represent the subjects of analysis and are described by quantitative “measures”,  
145 which are analyzed at different “granularities”, i.e. at different hierarchical levels of the  
146 dimensions (Berrahou et al., 2015; Sautot et al., 2015). For example, the fact “agricultural  
147 crop” can be represented according to three dimensions (crop, time and location) and  
148 described by the measure “crop yield” (kg. ha<sup>-1</sup>) (Fig. 1). Dimensions represent measure-  
149 analysis criteria and allow measures to be viewed and analyzed from different perspectives.

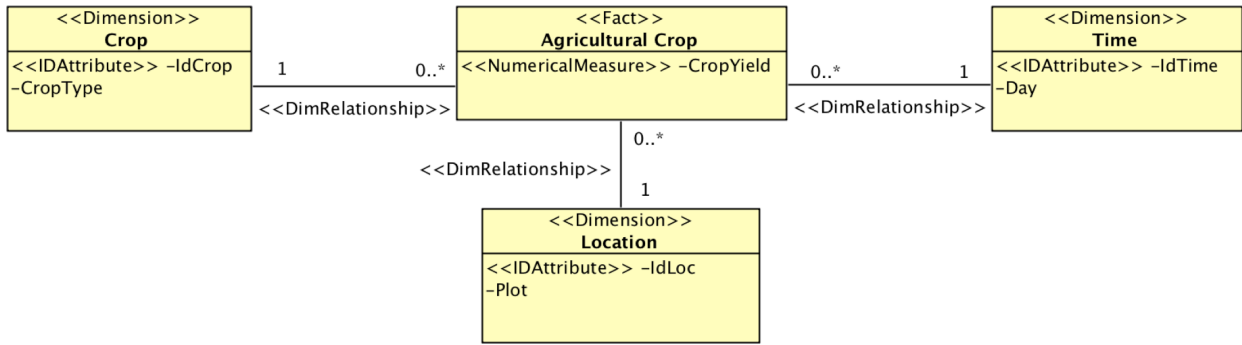


Figure 1 : Example of a multidimensional data model with three dimensions (crop (Id-Crop), time (Id-Time), and location (Id-Loc)) and a measure (crop yield)

150 Measures at upper levels of a hierarchical dimension are obtained by aggregating the  
 151 measures at the next lower level in the hierarchy.  
 152 Aggregation functions (e.g., sum, average, maximum) can be defined for each measure in the  
 153 hierarchical data model. For example, in a hierarchy associated with the dimension  
 154 “Location”, each mesh belongs to a plot, and each plot belongs to a catchment (Fig. 2).  
 155

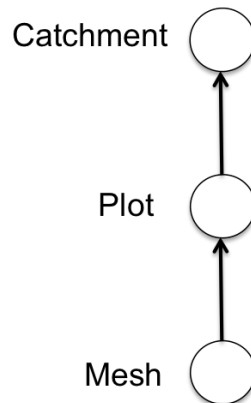


Figure 2 : Example of a hierarchical structure for the dimension “Location”

156  
 157  
 158 During data analysis, users exploit the DW by considering different combinations of  
 159 dimensions and different levels of their hierarchies. To select the right data at the right  
 160 abstraction level, users express and submit queries to the DW. Generally, such queries often  
 161 require scanning thousands or millions of records. Consequently, computing a query can be  
 162 highly time-consuming, while, ideally, the analysis process should be interactive. To reconcile  
 163 these two conflicting needs (i.e. a quick response to a query to a large database), the concept  
 164 of a “data cube”, the central OLAP concept, was introduced (Gray et al., 1996) to store  
 165 precomputed partial results. A data cube (or OLAP Cube) is a data abstraction that allows

166 users to view aggregated data according to a set of dimensions related to a pre-determined  
 167 structure. The cells of a data cube contain the measures selected from the DW according to the  
 168 dimensions and their hierarchies. A data cube is generally n-dimensional, though examples of  
 169 data cubes are often 3-dimensional for illustration purposes. As in the previous example, data-  
 170 cube representation of agricultural crop data with three dimensions (crop, time and location)  
 171 can be used to calculate a measure, i.e. a crop yield, for a given crop at different levels of time  
 172 and location with a sum function (Fig. 3).  
 173

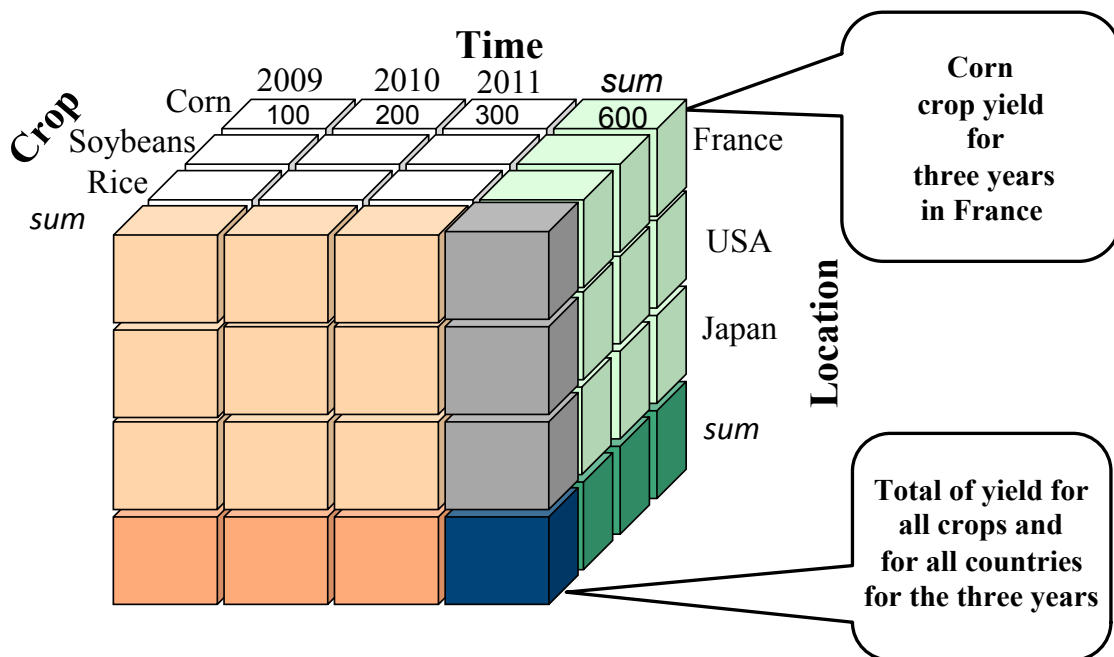


Figure 3 : Example of a data cube for agricultural crops

174

## 175 2.2. Relational platforms of DWs, OLAP analysis and data cubes

176

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



186 the value "France" in the "Location" dimension. They can also calculate the sum of yields of  
 187 all crops for three years in all countries, using roll-up operations on all dimensions of the  
 188 agricultural crop cube (Fig. 3).

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

198  
 199

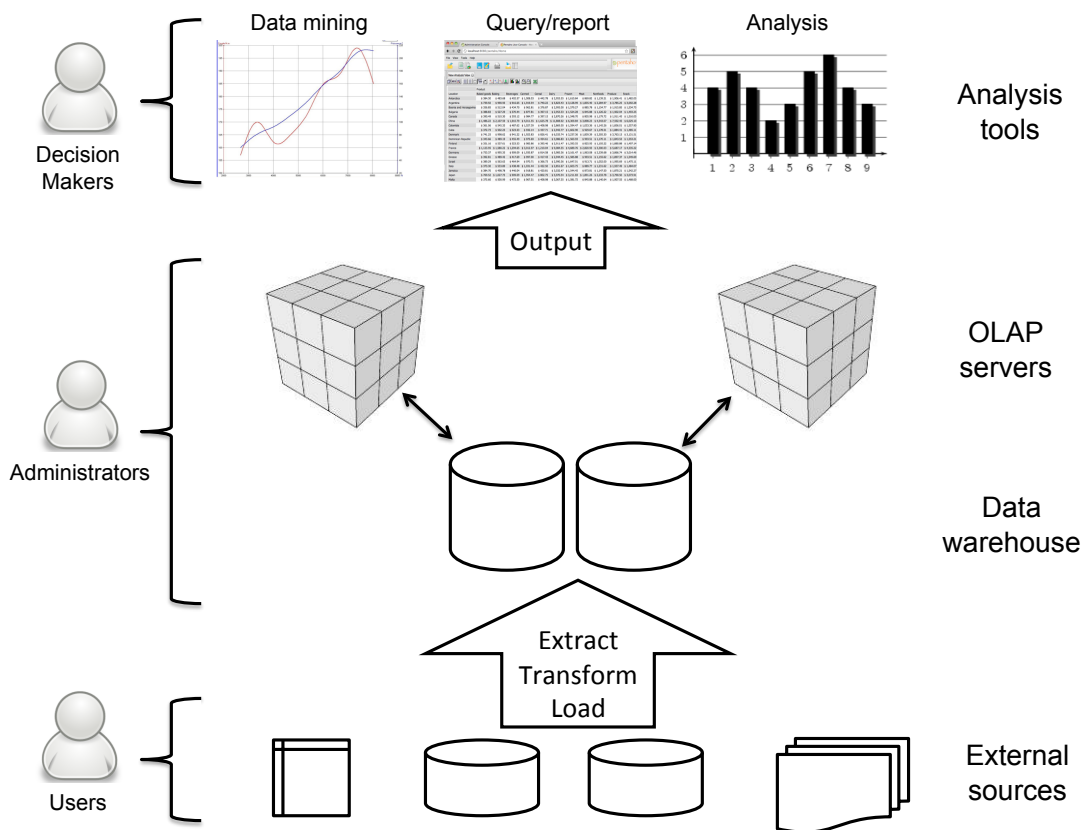


Figure 4 : Typical Data-Warehousing architecture

### 200 **3. TNT2 model and N-Catch DW design and implementation**

201

#### 202 **3.1. The spatially distributed agro-hydrological model TNT2**

203

204 TNT2 (Beaujouan et al., 2002) is a process-based and spatially-distributed model designed  
205 to capture spatial and temporal interactions of hydrological and agronomic processes  
206 throughout catchments. TNT2, developed to study water and N fluxes in small agricultural  
207 catchments (<50 km<sup>2</sup>), is dedicated to improving understanding of catchment systems by  
208 performing virtual experiments (Benhamou et al., 2013; Liao et al., 2016; Salmon-Monviola  
209 et al., 2013), including testing effects of numerous and diverse mitigation options proposed by  
210 local stakeholders (Moreau et al., 2012b) or public-policy makers (Durand et al., 2006) to  
211 decrease nitrate concentrations at the catchment outlet (Durand et al., 2015; Ferrant et al.,  
212 2011; Moreau et al., 2012a), as well as to support management and decision-making.

213 The TNT2 model comprises (i) the generic soil-vegetation atmosphere transfer model  
214 STICS (Brisson et al., 1998); (ii) the fully-distributed hydrological model TNT, based on  
215 TOPMODEL hypotheses (Beven, 1997); and (iii) a sub-model adapted from NEMIS (Hénault  
216 and Germon, 2000; Oehler et al., 2009) that simulates heterotrophic denitrification in soils.  
217 More details about hydrological and agronomical sub-model with each water and nitrogen  
218 processes are described in (Beaujouan et al., 2002, Salmon-Monviola et al., 2013; Ferrant et  
219 al., 2016). To represent the heterogeneity of water and N inputs and pathways, TNT2  
220 performs calculations explicitly on a grid, from cell to cell. The catchment is thus modeled by  
221 a cluster of three-dimensional columns, each corresponding to a pixel of a Digital Elevation  
222 Model (DEM) of the catchment (Moreau et al., 2013). Each column is divided into two  
223 compartments: a soil layer and an underlying shallow aquifer layer. Soil and aquifer porosity  
224 is determined by both retention (micro) and drainage (macro) porosities. Water flow in TNT2  
225 is modeled according to main hypotheses derived from TOPMODEL: i) the hydraulic gradient  
226 in each cell is constant and controlled by topography, and the gradient is calculated from the  
227 DEM at the beginning of the simulation; ii) the hydraulic transmissivity decreases  
228 exponentially with depth, and the model is based on a daily water-balance calculation for each  
229 cell and computes an explicit cell-to-cell multi-directional routing (water in one cell can flow  
230 to several cells), based on a D8 algorithm (Ocallaghan and Mark, 1984). Three outgoing flows  
231 are calculated for each cell in TNT2 : i) overland flow, resulting from soil saturation, which is  
232 routed to the soil surface of the downslope cells; ii) exfiltration that comes from excess  
233 groundwater and is also routed to the surface of the downslope cells; iii) subsurface flow that

234 comes from groundwater and is routed to the downslope cell groundwater. Water balance in  
235 river cells is calculated as in other cells, but all the calculated outflows (overland flow,  
236 exfiltration and subsurface flow) are assumed to be drained by the river and are routed directly  
237 from the river cell to the outlet without any interaction with other cells. The travel time in the  
238 stream is ignored. Nitrogen processes in soil and aquifer layer compartment of non-  
239 agricultural zones (urban and woodlands) of the catchment are taken into account. A part of  
240 surface water and nitrogen flow (overland flow, exfiltration flow) from upslope cells and  
241 rainfall input infiltrate in soil of non-agricultural cells. The other part is routed to the outlet.  
242 For non-agricultural cell as in other cells, subsurface flow of water and nitrogen, that comes  
243 from groundwater is routed to the downslope cell groundwater. Model inputs and parameters  
244 include (i) a DEM in raster format; (ii) a map in raster format delineating agricultural plots,  
245 roads and the hydrological network; (iii) a map in raster format of homogeneous soil zones;  
246 (iv) a map in raster format representing homogeneous climate zones, which allow climate  
247 gradients to exist within the catchment; (v) daily climate data (i.e., minimum and maximum  
248 air temperatures, precipitation, potential evapotranspiration, total solar radiation) for each  
249 climate zone; (vi) 20 soil properties for each soil type in the soil-zone map; and (vii) cropping  
250 systems for each plot during the simulation period. Cropping systems are defined at two levels  
251 (Leenhardt et al., 2010): (i) crops and their succession over time in each plot, called a crop  
252 rotation; and (ii) the crop management system (CMS), representing an organized series of  
253 cultivation techniques or crop operations (e.g. sowing, mineral and organic N applications,  
254 harvesting, grazing, mowing) applied to a crop to obtain a given product. A crop operation has  
255 a duration defined by starting and ending dates, as well as a chronological position in the  
256 CMS in which it occurs.

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

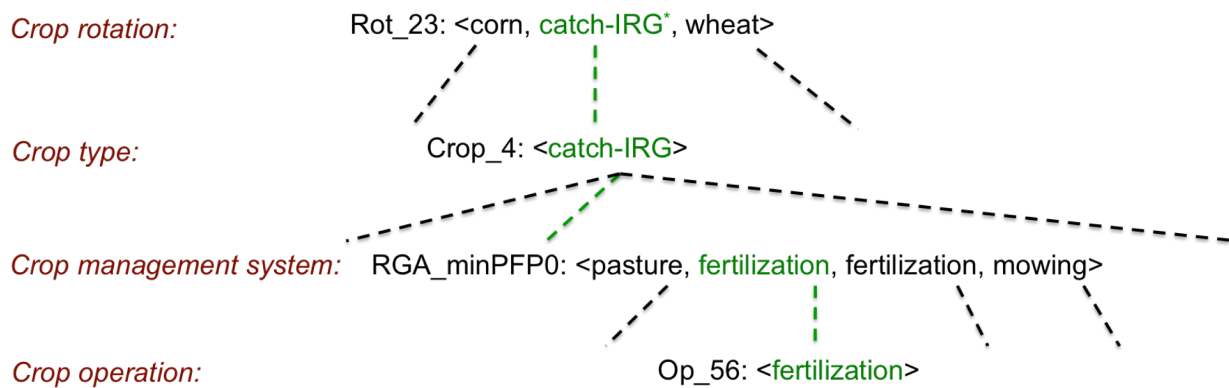
268 algorithm (Salmon-Monviola et al., 2012) to allocate a CMS to plots, with constraints on  
269 several CMS of each crop type.

270

### 271 3.2. Design of the N-Catch DW

272

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



\* IRG : Italian Ryegrass

281

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

282

#### 283 3.2.1. Measures stored in the N-Catch DW

284

285 To study how agricultural practices influence N emissions to the air, groundwater and  
286 stream water, measures in the N-Catch fact table are composed of input data (i.e. cropping  
287 systems, catchment description, meteorological and hydrological parameters) and extensive  
288 output data generated by TNT2 (i.e. water fluxes and N concentration and fluxes at a daily  
289 time step for each plot in the catchment). These simulated results are stored in two types of

290 files: (i) N variables predicted by the model for each operation *i.e* sowing, fertilizing,  
 291 harvesting, grazing, mowing) for crop sequences on each plot throughout the simulation  
 292 period, and (ii) water and N concentrations and fluxes at a daily time step for each plot. Out of  
 293 44 variables, 16 that describe water and N flux in the groundwater and air, water and N stored  
 294 in the soil and groundwater and N denitrification are stored in N-Catch (Table 1). From these  
 295 16 variables, we added two calculated measures that correspond to N flux from soil to  
 296 groundwater and from soil to air, defined as:

$$297 \quad NFlux_{Soil\ to\ GW} = N\_Atm - N\_Denit - N\_Volat\_Ferti + N\_Mine\_Manure + N\_Mine\_Grazing$$

$$298 \quad + N\_Mine + N\_Fix - \Delta N\_Soilwater + N\_Fertilizer - N\_Plant \text{ (kg.ha}^{-1}\text{)}$$

$$299 \quad NFlux_{Soil\ to\ Atm} = N\_Denit + N\_Volat\_Ferti + N\_Volat\_Manure \text{ (kg.ha}^{-1}\text{)}$$

301 With  $\Delta N\_Soilwater$ , the stock variation of nitrogen stored in soil water storage (kg.ha<sup>-1</sup>). Each  
 302 variable is defined as in Table 1.

303 A three-dimensional data model of N-Catch includes the dimensions of location, time and  
 304 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 <sup>-1</sup> )
N_Weathered	Nitrogen stored in weathered layer (kg.ha <sup>-1</sup> )
N_GW	Nitrogen stored in groundwater (kg.ha <sup>-1</sup> )
N_Fix	Atmospheric nitrogen fixed by plants (kg.ha <sup>-1</sup> )
N_Mine	Mineral nitrogen resulting from mineralization (kg.ha <sup>-1</sup> )
N_Denit	Denitrified nitrogen (kg.ha <sup>-1</sup> )
N_Sequestre	Nitrogen stored in organic matter (kg.ha <sup>-1</sup> )
N_Plant	Nitrogen fixed by plants (kg.ha <sup>-1</sup> )
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

### 306 **3.2.2. The dimensions of the N-Catch DW**

307

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

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

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

321 *Agricultural dimension.* The agricultural dimension is a key part of N-Catch, since the main  
322 purpose of simulation is to predict impacts of agricultural practices on stream water quality.  
323 Cropping system data used as input to TNT2 were stored in the DW. We define four levels  
324 from these data: crop operation, CMS, crop type and crop rotation (Fig. 5). As an example of  
325 the agricultural dimension, the crop operation "fertilization" appears in the CMS  
326 RGA\_minPFP0 <pasture, fertilization, mowing>, which belongs to the crop type "IGS-  
327 grassland" in the crop rotation <corn, catch-IRG, wheat>.

328

329

### 330 **3.3. Implementing the N-Catch DW**

331

332 DW tiers are generally managed using a relational database in which data are structured  
333 following three main schema types: star, constellation and snowflake modeling (IBM, 1998).  
334 The first is the most common and consists of a central fact and dimensions visually  
335 represented by a star. In contrast, a constellation model merges multiple star models using  
336 common dimensions; it therefore includes several facts and common or specific dimensions.  
337 Finally, the snowflake model is an offshoot of the star model: the fact is maintained, while the  
338 dimensions are split into several tables according to their hierarchies. The snowflake model is  
339 required for flexible querying of complex dimension relations. Because this corresponds to

340 our case study, based on a set of multilevel dimensions with complex hierarchical structures  
 341 and relations, we chose the snowflake model to design N-Catch (Fig.6).  
 342

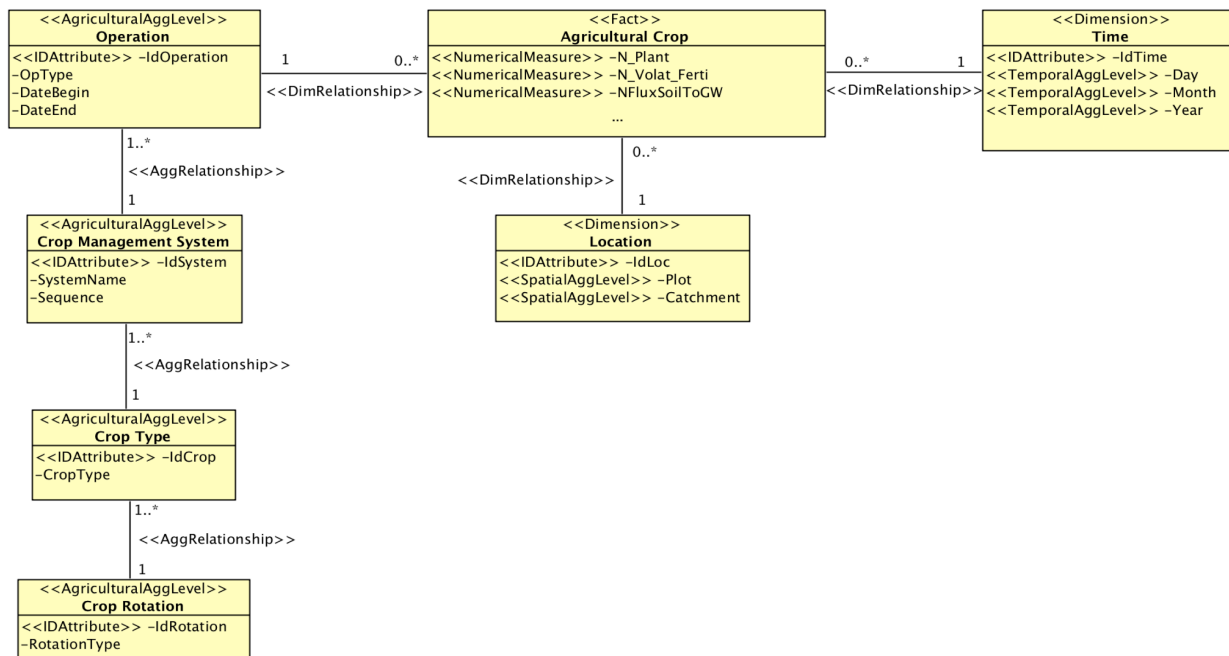


Figure 6 : Multidimensional data model of the N-Catch data warehouse

343  
 344 Once the preliminary steps are completed (i.e., identification of user needs, data collection,  
 345 multidimensional model design), the ETL step is developed.

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

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



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

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

364

### 365 **3.4. Visualization and exploration of the data stored in N-Catch**

366

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

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

391 Processing mixed and geometric dimensions implies redefining OLAP aggregation  
392 operators for such data. In the N-Catch DW, the spatial dimension contains only two



393 hierarchical levels (plot and catchment). Ultimately, N-Catch required only cartographic  
394 visualization of simulation results; thus, spatio-multidimensional operators were not essential,  
395 and coupling of GIS and OLAP was sufficient. We used QGIS (<http://www.qgis.org>), which  
396 is free GIS software shared under the GNU GPL. One of the main advantages of working with  
397 QGIS is the ability to integrate plugins that can enhance its features. To facilitate users'  
398 interactions with N-Catch and the spatial visualization tool, we used a plugin with a graphical  
399 interface that allows users to select data from N-Catch (variable, duration and location) and  
400 view them on a map. This plugin allows users to create a new map (i.e. a new data layer) and  
401 save it or to select a pre-existing map.

402

## 403 **4. Illustrating N-Catch in a case study**

404

### 405 **4.1. Study site and simulation procedure**

406

407 The TNT2 model was applied to the Yar catchment, comprising 10 sub-catchments, which  
408 discharge into the "Lieu de Grève" bay, located in Brittany, western France (Fig. 7). The Yar  
409 flows towards the English Channel and drains an area of 61.5 km<sup>2</sup>, of which 8% is urban, 28%  
410 is woodlands and 64% is agricultural land, the latter comprising 4620 agricultural plots. Plot  
411 boundaries were based on an aerial survey. In 2007, 194 farms had all or part of their  
412 agricultural area in the catchment (Moreau et al., 2012b). Despite moderate nitrate  
413 concentrations around 6.8 mg N-NO<sub>3</sub>.l<sup>-1</sup> (i.e., much lower than the 11.3 mg N-NO<sub>3</sub>.l<sup>-1</sup> limit  
414 defined by national and European directives), the coastal bay at the catchment outlet has  
415 experienced macro-algal blooms every summer for the past 40 years. The physiographic  
416 context and the hydrodynamic conditions (confined bay, low currents) in the "Lieu de Grève"  
417 bay makes it very sensitive to coastal eutrophication. In this bay, algae proliferation is  
418 important because they are well fed by continental inputs and marine currents cannot take  
419 them offshore. Better understanding of catchment functioning requires detailed analysis of  
420 effects of current cropping systems and changes in them on N emissions at different locations  
421 and periods, i.e. different spatial and temporal levels. Data used for the case study, which  
422 correspond both to input and output of the TNT2 model, span from 1 Sep 1996 to 31 Aug  
423 2008. For this simulation (Moreau et al., 2012b), 16 output variables per plot on 4620 plots  
424 were stored for 4380 days (i.e., 20,235,600 records with 16 data fields, with a total database  
425 size of 9 GB). Daily weather data were acquired from Météo France.

426

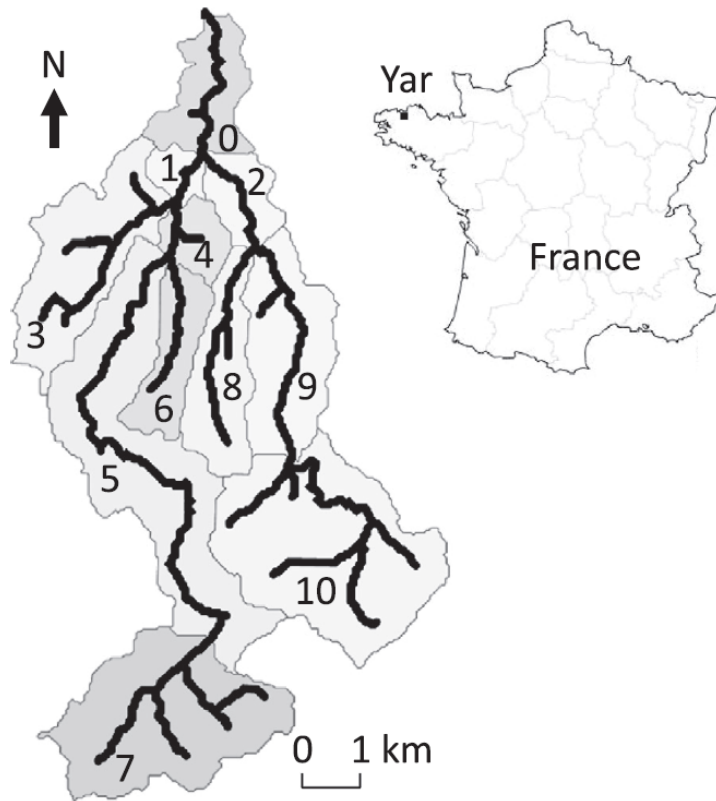


Figure 7 : Plot map of the Yar catchment, Brittany, France

427

## 428 **4.2. N-Catch illustration**

429

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

435

### 436 **4.2.1. Temporal dimension**

437

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

#### 445 **4.2.2. Spatial dimension**

446

447 The spatial dimension can be used to analyze spatial relations between measures, for  
448 example, between those of permanent grassland (P-Grassland) and temporary grassland (T-  
449 Grassland) (Fig. 9). In this example, the aggregation function used is the sum of simulated  
450 values over time, divided by the rotation duration, to obtain mean values per year. A high  
451 positive correlation ( $r = 0.8$ ) exists between denitrification and N emissions to air, while the  
452 correlation between N mineralization and N emissions to groundwater depends on the type of  
453 grassland: for the same amount of mineralization, N emissions to groundwater are much lower  
454 in permanent grasslands than in temporary grasslands (Fig. 9). A slight positive correlation ( $r$   
455  $= 0.3$ ) is observed between N emissions to groundwater and N mineralization; however, the  
456 higher the amount of mineralization, the more variable are the N emissions to groundwater.  
457 No correlation was found between fertilization and N emissions from soil to groundwater, but  
458 above a certain amount of N fertilization ( $76 \text{ kg N}\cdot\text{ha}^{-1}\cdot\text{y}^{-1}$ ), variability in N emissions to  
459 groundwater among plots increases (Fig. 9c). The correlation between N emissions from soil  
460 to groundwater and that to air is again scattered around two groups according to land cover  
461 (Fig. 9d). For the same N emission from soil to air, N emission from soil to groundwater is  
462 much lower in permanent grasslands than in temporary grasslands. Such analysis is interesting  
463 and has never been performed in previous applications of TNT2.

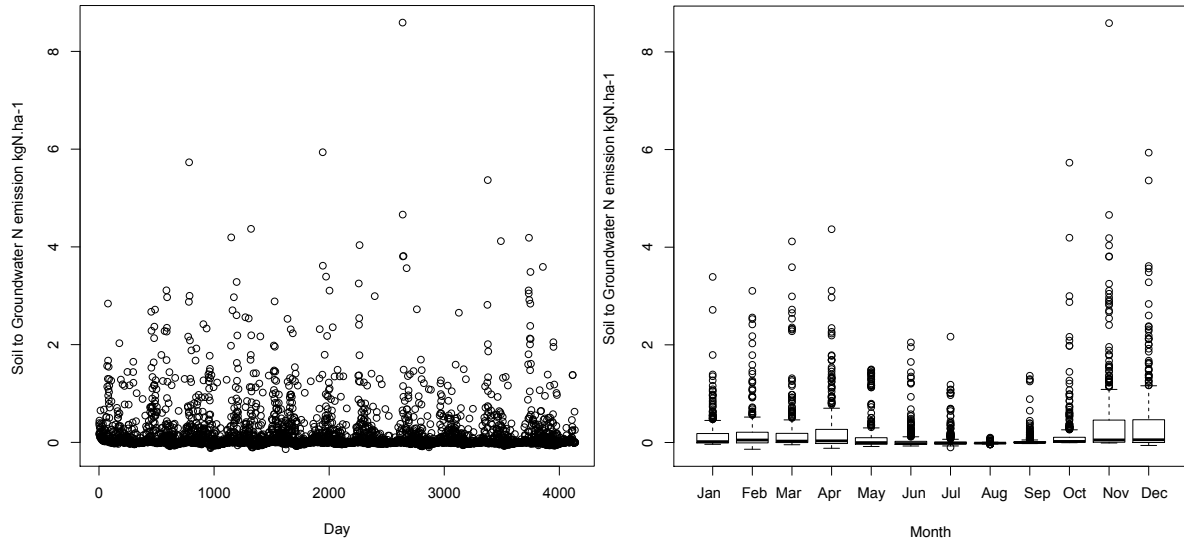
464

#### 465 **4.2.3. Agricultural dimension**

466

467 Considering all plots in the catchment and analyzing daily variations in N emissions  
468 from soil to groundwater reveals complex relations between agricultural practices and nitrate  
469 leaching at the plot scale. The aggregation function used in this case corresponds to mean  
470 daily N emission from soil to groundwater per month across the catchment. Differences  
471 among crops are small (Fig. 10a), but they do exist, such as those between wheat and corn.  
472 Intra-crop variability is surprisingly high. For example, half of the monthly N emissions from  
473 soil to groundwater for potatoes vary strongly. This variability can be due to climate,  
474 topographic position within the catchment or crop management. Effects of agricultural  
475 practices can be seen in variations in mean daily N emission from soil to groundwater per  
476 CMS, which integrates crop type and crop operations at the catchment level (Fig. 10b).  
477 Differences are much smaller among CMSs than among crop types, which indicates that crops  
478 must be considered in their agricultural and environmental contexts. The type of CMS

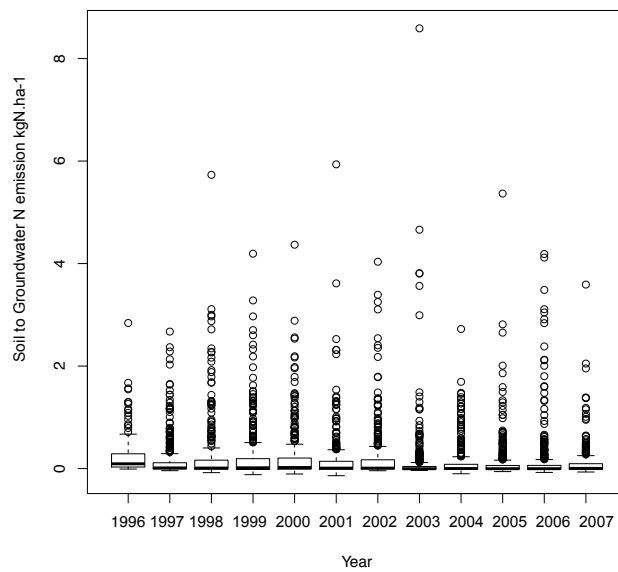
479 particularly influences the presence of outliers. N-Catch is particularly effective at analyzing  
480 variability in effects of crop management, identifying high and low values and searching for  
481 factors that explain them, and ultimately proposing recommendations or mitigation options.  
482



483

(a)

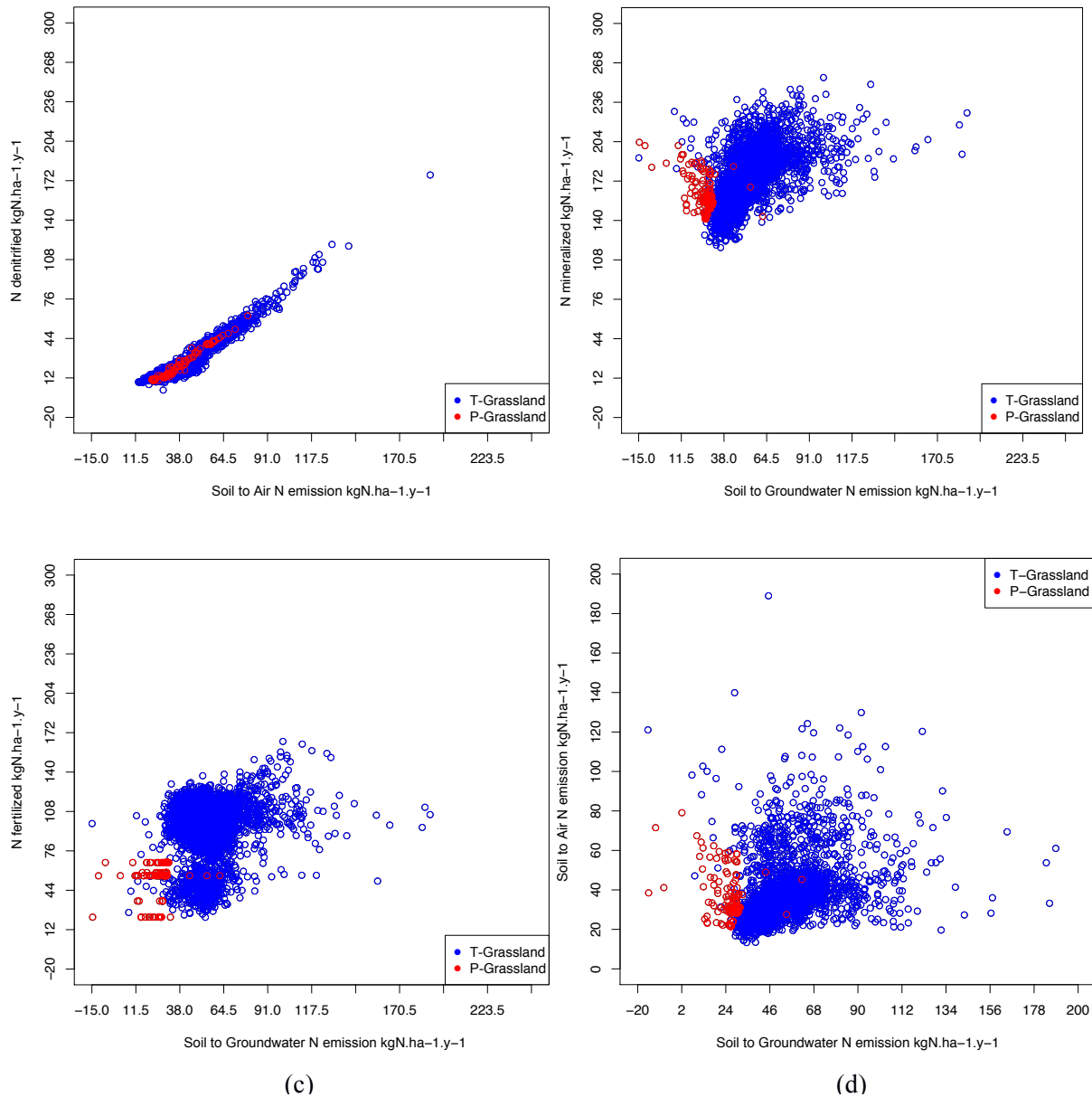
(b)



484

(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



485

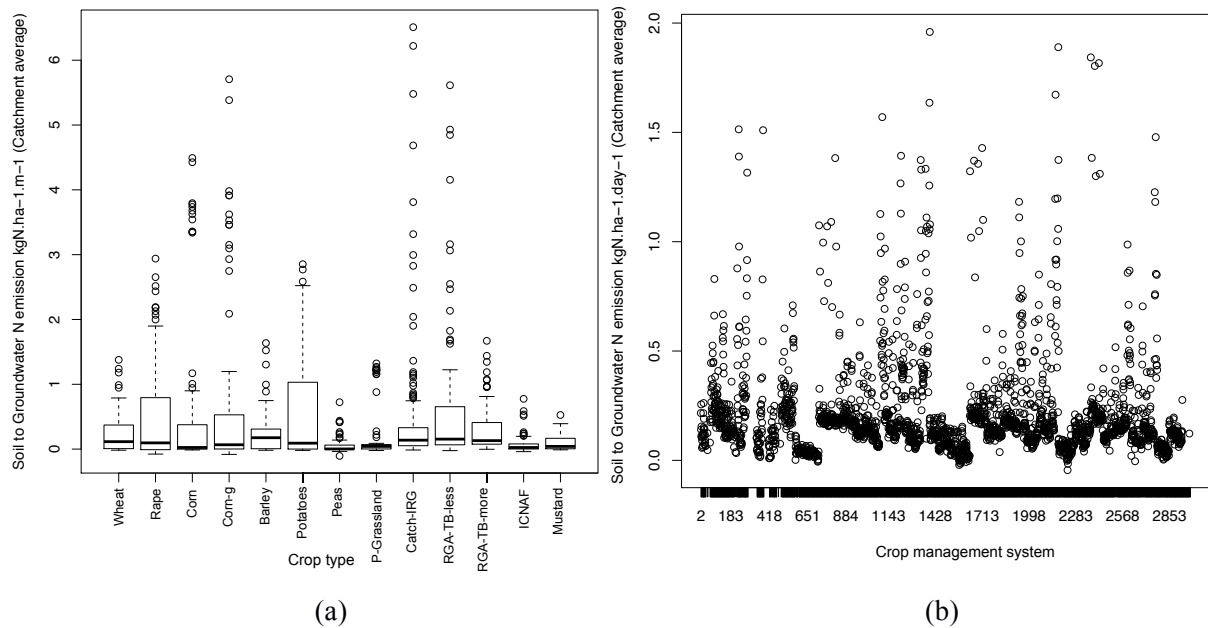
Figure 9 : Process analysis at the plot scale: a) N denitrification versus N emission from soil to air; b) N mineralization versus N emission from soil to groundwater; c) N fertilization versus N emission from soil to groundwater; d) N emission from soil to groundwater versus N emission from soil to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

#### 486 4.2.4. Coupling N-Catch with QGIS

487

488 In the previous examples, displaying plots in a point cloud identified the most polluting  
 489 plots in the Yar catchment; however, plot ID numbers provide limited information. Users  
 490 should be able to visualize these plots on a map, for example, to know their topographic  
 491 position in a catchment (i.e. upper, middle or lower part of the hillslope) and their extent in the  
 492 catchment. Simple visualization of this spatial query helps users understand the phenomenon

493 in question (i.e. nitrate pollution) by seeing its position and extension within a geographic  
 494 frame of reference.



495  
 496  
 497  
 498  
 499  
 500  
 501

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.

502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510

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.

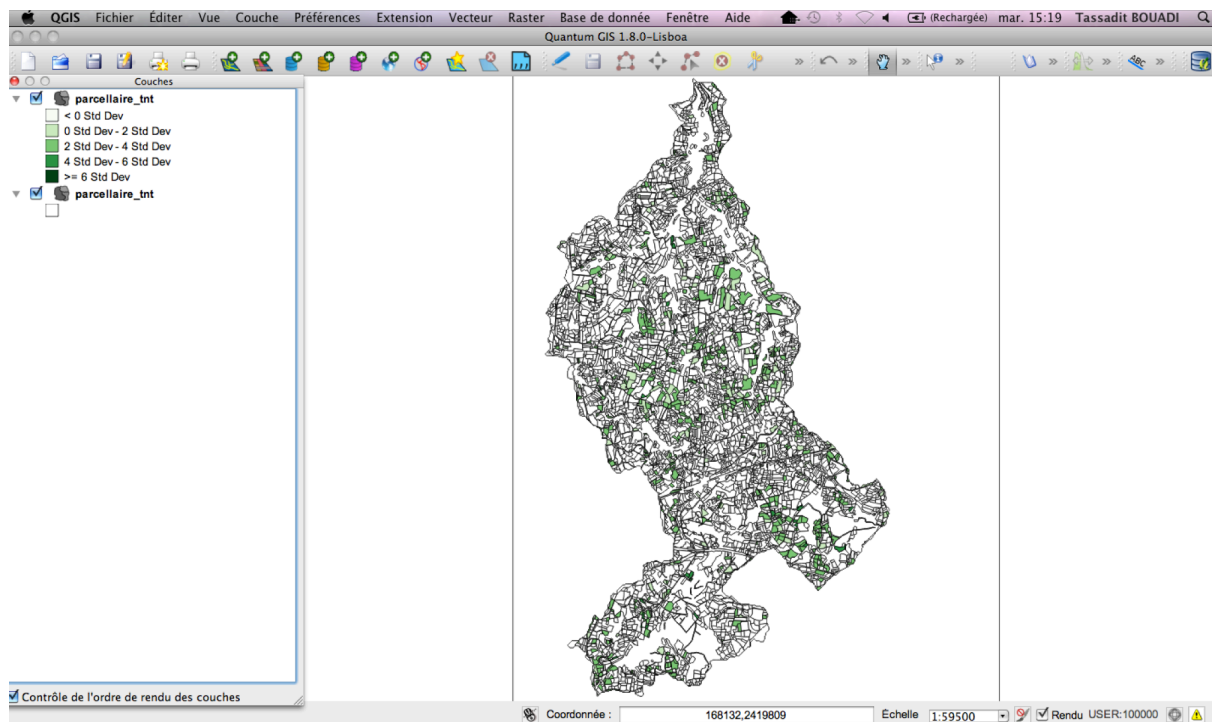


Figure 11: Sum of daily N emission from soil to groundwater at the plot level for the hydrological year 1997-1998 in the Yar catchment, Brittany, France

## 511 **5. Discussion**

512

### 513 **5.1. Purpose, scope and generality of the N-catch DW**

514

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

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

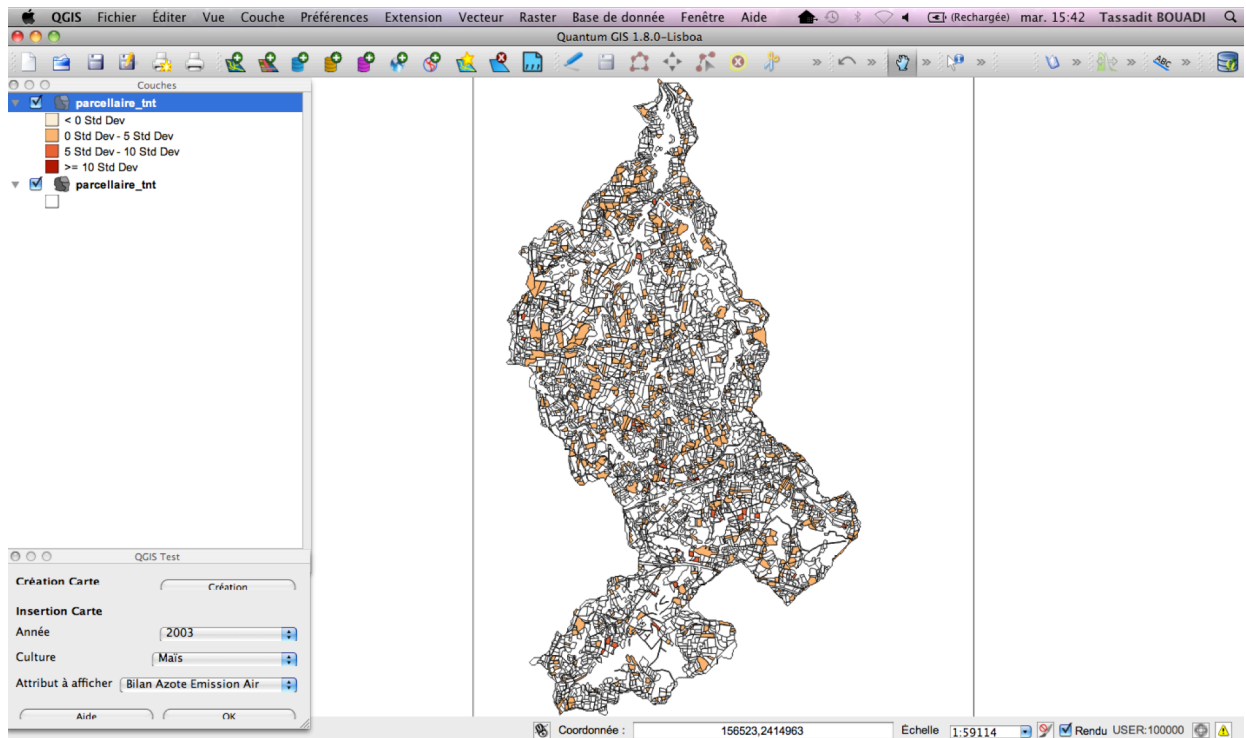


Figure 12: Mean daily N emission from soil to air for maize in 2003 at the plot level in the Yar catchment, Brittany, France

525 queries by a factor of 5, to vary between 0,01 seconds to 0,4 seconds (a complex query  
 526 template leads to a significant increase in the query execution time), and (ii) materialized  
 527 views to improve query performance by pre-calculating complex query templates (*e.g.*  
 528 expensive join and aggregation operations) and storing the results in the data warehouse.  
 529

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



530 The main methodological contributions of this work include methods to (i) preprocess,  
531 transform and load simulated data to DW; (ii) describe agricultural practices in an original  
532 manner by representing hierarchies in cropping systems; and (iii) analyze simulation results  
533 by combining spatiotemporal simulation data, data warehousing and online analysis. The  
534 QGIS plugin associated with this DW architecture allows end users to visualize and explore  
535 geo-referenced data, which has been identified as an important need (Boulil et al., 2013,  
536 2014).

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

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

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

561

## 562 **5.2. Limits and potential improvements of the N-Catch DW**

563

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

571 N-Catch should be expanded with data-mining or information-retrieval methods, such as  
572 skyline queries, to perform advanced analyses to find meaningful patterns and relations in the  
573 simulation data (Bouadi et al., 2012). Data-mining techniques are powerful tools for  
574 identifying and extracting interesting knowledge from large data collections. To extend OLAP  
575 analysis from simple aggregate operations on data cubes for prediction or decision objectives,  
576 many studies investigated the combination of OLAP analysis with data mining. Three main  
577 approaches are found in the literature: i) modifying OLAP operators (Bentayeb and Favre,  
578 2009; Chen et al., 2000; Goil and Choudhary, 2001; Han, 1997; Sarawagi, 1999; Sathe and  
579 Sarawagi, 2001) to simulate data-mining techniques; (ii) adapting multidimensional structures  
580 (Chen et al., 2001; Goil and Choudhary, 2001; Pinto et al., 2001) by reorganizing the  
581 multidimensional data to make them usable with data-mining methods; and (iii) adapting data-  
582 mining algorithms (Bodin-Niemczuk et al., 2008; Giacometti et al., 2008; Sarawagi et al.,  
583 1998) to cope with the multidimensional data environment. Coupling data warehousing and  
584 skyline queries allows users to navigate along dimensions' hierarchies (i.e.  
585 specialize/generalize) while ensuring online calculation of the skyline. Bouadi et al. (2014)  
586 developed an efficient approach for simulating the effect of the OLAP "drill-down" operator  
587 on the computation of skyline queries. Further discussions with agronomists and agro-  
588 hydrologists will test and illustrate their uses, but we can already say that they enable queries  
589 to be formulated in the DW by combining conflicting environmental indicators and finding  
590 compromise solutions associated with these requests that meet stakeholders' expectations.  
591 Subsequently, it would be interesting to design a user-friendly interface for N-Catch that  
592 provides users with centralized access to all the functions developed (e.g. OLAP analysis,  
593 skyline queries).

594 N-catch was applied for specific results of TNT2 simulation of the Yar catchment in an  
595 offline manner. One challenge would be to couple N-catch and TNT2 online. In this way,  
596 simulation results of TNT2 could be loaded automatically into the DW. This integration can  
597 be achieved by using tools for manipulating data (Chuffart et al., 2010; Laniak et al., 2013) to

598 manage simulation runs and process and load model output into the DW. To automate  
599 creation of the DW, one approach could be to not set an “a priori” process for the DW, but to  
600 use algorithms to automatically design DW schemas and OLAP cubes. Sautot et al. (2015)  
601 developed such an approach by using hierarchical agglomerative clustering with a metric  
602 (similarity index) to automatically build hierarchical dimensions in an OLAP cube. With this  
603 similarity index, one can perform hierarchical clustering in heterogeneous datasets that  
604 contain qualitative and quantitative variables. Their approach, based on data-mining methods,  
605 can supplement expert knowledge during the design of an OLAP cube. With this method, one  
606 could build new dimensions based on hierarchies in the data that are initially hidden.

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

617

## 618 **6. Conclusion**

619

620 A DW is expected to (i) provide dynamic multidimensional analysis, supporting end users  
621 with analytical and navigational properties, and (ii) offer a rapid response to queries,  
622 regardless of the DW’s size and complexity. Such a framework was developed and applied to  
623 quantify impacts of agricultural practices on N emissions at multiple spatiotemporal levels by  
624 providing relevant ways to identify and analyze where and why N pollution is present in a  
625 catchment. The method developed shows how OLAP and the data-cube concept can be useful  
626 for analyzing the huge amount of data produced by modeling activities. This design was  
627 applied by developing the N-Catch DW, which was built to store and manage simulation data  
628 from the agro-hydrological model TNT2, and was illustrated in a case study. The N-Catch  
629 DW allows users to explore N emissions in space and time, to more accurately analyze  
630 transfer and transformation processes as a function of cropping systems, and to obtain new  
631 knowledge that facilitates making specific and detailed decisions in space and time. The

632 approach adopted for developing N-Catch is not specific to this case study. It can be applied  
633 to explore the functioning of any agro-environmental system. In particular, the spatiotemporal  
634 modeling and agricultural dimension are generic to all models with landscape and regional  
635 dimensions. Instead of transferring models, the future of agro-environmental modeling could  
636 be to allow users to easily navigate the simulation data of numerous models that were  
637 developed to help make decisions about effects of human activities on the environment.

638

639

## 640 **Acknowledgments**

641

642 This research was supported by grants from the ANR Systerra project ACASSYA (ANR-  
643 08-STRA-01). Special thanks go to Patrick Durand for providing the opportunity to use the  
644 TNT2 model, and to Sylvain Dousset for QGIS-NCatch coupling.

645

## 646 **References**

647

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

651 Beaujouan, V., Durand, P., and Ruiz, L. (2001). Modelling the effect of the spatial distribution  
652 of agricultural practices on nitrogen fluxes in rural catchments. *Ecological Modelling*,  
653 137(1):93–105 [http://dx.doi.org/10.1016/S0304-3800\(00\)00435-X](http://dx.doi.org/10.1016/S0304-3800(00)00435-X)

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

658 Benhamou, C., Salmon-Monviola, J., Durand, P., Grimaldi, C., and Merot, P. (2013).  
659 Modeling the interaction between fields and a surrounding hedgerow network and its  
660 impact on water and nitrogen flows of a small watershed. *Agricultural Water Management*,  
661 121:62–72 <http://dx.doi.org/10.1016/j.agwat.2013.01.004>

662 Bentayeb, F. and Favre, C. (2009). ROK: Roll-Up with the K-Means clustering method for  
663 recommending OLAP queries. In *Proceedings of the International Conference on Database*

664 and Expert Systems Applications, pages 501–515 [http://dx.doi.org/10.1007/978-3-642-03573-](http://dx.doi.org/10.1007/978-3-642-03573-9_43)  
665 [9\\_43](http://dx.doi.org/10.1007/978-3-642-03573-9_43)

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

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

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

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

678 Bouadi, T., Cordier, M.-O., and Quiniou, R. (2012). Incremental computation of skyline  
679 queries with dynamic preferences. In *Proceedings of the International Conference on*  
680 *Database and Expert Systems Applications*, pages 219–233 [http://dx.doi.org/10.1007/978-3-](http://dx.doi.org/10.1007/978-3-642-41221-9_2)  
681 [642-41221-9\\_2](http://dx.doi.org/10.1007/978-3-642-41221-9_2)

682 Bouadi, T., Cordier, M.-O., and Quiniou, R. (2014). Computing hierarchical skyline queries  
683 "on-the-fly" in a data warehouse. In *Proceedings of the International Conference on Data*  
684 *Warehousing and Knowledge Discovery*, pages 146–158 [http://dx.doi.org/10.1007/978-3-](http://dx.doi.org/10.1007/978-3-319-10160-6_14)  
685 [319-10160-6\\_14](http://dx.doi.org/10.1007/978-3-319-10160-6_14)

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

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

694 Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M., Ruget, F., Nicoullaud, B., Gate, P.,  
695 Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S.,  
696 Tayot, X., Plenet, D., Cellier, P., Machet, J., Meynard, J., and Delécolle, R. (1998). STICS:  
697 a generic model for the simulation of crops and their water and nitrogen balances. I. Theory

698 and parameterization applied to wheat and corn. *Agronomie*, 18(1):311–346  
699 <http://dx.doi.org/10.1051/agro:19980501>

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

702 Chen, M., Zhu, Q., and Chen, Z. (2001). An integrated interactive environment for knowledge  
703 discovery from heterogeneous data resources. *Information & Software Technology*,  
704 43(8):487–496 [http://dx.doi.org/10.1016/S0950-5849\(01\)00159-8](http://dx.doi.org/10.1016/S0950-5849(01)00159-8)

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

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

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

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

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

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

729 Ferrant, S., Oehler, F., Durand, P., Ruiz, L., Salmon-Monviola, J., Justes, E., Dugast, P.,  
730 Probst, A., Probst, J.-L., and Sanchez-Perez, J.-M. (2011). Understanding nitrogen transfer  
731 dynamics in a small agricultural catchment: comparison of a distributed (TNT2) and a semi



732 distributed (SWAT) modeling approaches. *Journal of Hydrology*, 406(1):1–15  
733 <http://dx.doi.org/10.1016/j.jhydrol.2011.05.026>

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

737 Goil, S. and Choudhary, A. (2001). PARSIMONY: An infrastructure for parallel  
738 multidimensional analysis and data mining. *Journal of Parallel and Distributed Computing*,  
739 61(3):285–321 <http://dx.doi.org/10.1006/jpdc.2000.1691>

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

744 Han, J. (1997). OLAP mining: An integration of OLAP with data mining. In *Proceedings of*  
745 *the Working Conference on Database Semantics*, pages 1–9 [http://dx.doi.org/10.1007/978-](http://dx.doi.org/10.1007/978-0-387-35300-5_1)  
746 [0-387-35300-5\\_1](http://dx.doi.org/10.1007/978-0-387-35300-5_1)

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

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

752 Inmon, W.H. (2005). *Building the data warehouse*. John Wiley & sons.

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

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

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

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

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

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

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

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

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

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

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

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

787 Oehler, F., Durand, P., Bordenave, P., Saadi, Z., and Salmon-Monviola, J. (2009). Modelling  
788 denitrification at the catchment scale. *Science of The Total Environment*, 407(5):1726–  
789 1737 <http://dx.doi.org/10.1016/j.scitotenv.2008.10.069>

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

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

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



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

803 Salmon-Monviola, J., Moreau, P., Benhamou, C., Durand, P., Merot, P., Oehler, F., and  
804 Gascuel-Oudou, C. (2013). Effect of climate change and increased atmospheric co2 on  
805 hydrological and nitrogen cycling in an intensive agricultural headwater catchment in  
806 western France. Climatic change, 120(1-2):433–447 [http://dx.doi.org/10.1007/s10584-013-](http://dx.doi.org/10.1007/s10584-013-0828-y)  
807 [0828-y](http://dx.doi.org/10.1007/s10584-013-0828-y)

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

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

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

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

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

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

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