Stochastic hydro-economic modeling for optimal management of groundwater nitrate pollution from agricultural under hydraulic conductivity uncertainty.

S. PEÑA-HARO\textsuperscript{a,b}, M. PULIDO-VELAZQUEZ\textsuperscript{b}, C. LLOPIS-ALBERT\textsuperscript{b},

\textsuperscript{a} Institute of Environmental Engineering, ETH Zurich, Wolfgang-Paulistrasse 15, CH-8093 Zürich, Switzerland. e-mail: pena@ifu.baug.ethz.ch
\textsuperscript{b} Departamento de Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Abstract: In decision-making processes, reliability and risk-aversion play a decisive role. This paper presents a framework for stochastic optimization of control strategies for groundwater nitrate pollution from agriculture under hydraulic conductivity uncertainty. The main goal is to analyze the influence of uncertainty in the physical parameters of a heterogeneous groundwater diffuse pollution problem on the results of management strategies, and to introduce methods that integrate uncertainty and reliability in order to obtain strategies of spatial allocation of fertilizer use in agriculture. A hydro-economic modeling approach is used for obtaining the allocation of fertilizer reduction that complies with the maximum permissible concentration in groundwater while minimizes agricultural income losses. The model is based upon nonlinear programming and groundwater flow and mass transport numerical simulation, condensed on a pollutant concentration response matrix. The effects of the hydraulic conductivity uncertainty on the allocation of nitrogen reduction among agriculture pollution sources is analyzed using four formulations: Monte Carlo simulation with pre-assumed parameter field, Monte Carlo optimization, stacking management, and mixed-integer stochastic model with predefined reliability. The formulations were tested in an illustrative example for 100 hydraulic conductivity realizations with different variance.
The results show a high probability of not meeting the groundwater quality standards when deriving a policy from just a deterministic analysis. To increase the reliability several realizations can be optimized at the same time. By using a mixed-integer stochastic formulation, the desired reliability level of the strategy can be fixed in advance. The approach allows deriving the trade-offs between the reliability of meeting the standard and the net benefits from agricultural production. In a risk-averse decision-making, not only the reliability of meeting the standards counts, but also the probability distribution of the maximum pollutant concentrations. A sensitivity analysis was carried out to assess the influence of the variance of the hydraulic conductivity fields on the strategies. The results have shown that larger the variance, greater the range of maximum nitrate concentrations and the worst-case (or maximum value) that could be reached for the same level of reliability.

**Keywords** groundwater; fertilizer allocation; nitrates; uncertainty; optimization; stochastic management model

1. Introduction

Agricultural activities are often the main source of elevated nitrate concentrations in groundwater (e.g., Oyarzun et al., 2007). Moreover, in the last decades the nitrate concentrations in groundwater increased due to the intensive use of fertilizers in agriculture (e.g., Candela et al., 2008). The need of controlling of groundwater diffuse pollution has given rise to the development of an extensive legal framework in several countries. In Europe, the requirements for agricultural nonpoint pollution in Europe are being ruled by a series of European Directives. The Nitrates Directive (Directive 91/676/EEC), which was established in 1991 to reduce nitrate water pollution from agricultural sources, involves the declaration of Nitrate Vulnerable Zones in which constraints are placed on inorganic fertilizer and organic slurry application rates. The Drinking Water Directive (80/778/EEC and its 98/83/EC revision) sets a maximum allowable concentration for nitrate of 50 mg/l, while the EU Water Framework
Directive (Directive 2000/60/EC; WFD), enacted in 2000, establishes a legal framework to protect and restore clean water across Europe and ensure its long-term sustainable use. The WFD includes groundwater in its river basin management planning, and sets clear milestones for groundwater bodies in terms of delineation, economic analysis, characterization (analysis of pressures and impacts), monitoring, and the design of programs of measures to ensure a good status of quantity and chemical groundwater status by 2015. In addition, significant upward trends in the concentration of pollutants should be identified and reversed (Directive 2006/118/EC, Groundwater Directive).

In order to control and improve groundwater quality, it is necessary to implement often costly management decisions, and here computer models has a basic role for simulating the impact of different policies and get insight into the best options according to the objectives and constraints of our problem. Modeling of nitrate contamination of groundwater in agricultural watersheds has mostly been addressed in a deterministic way (e.g., Martínez and Albiac, 2004; Almasri and Kaluarachchi, 2005; Candela et al., 2008; Peña-Haro et al., 2009). However, because of the heterogeneous nature of most groundwater bodies and the inherent uncertain, the errors involved in the predictions of future pollutant concentration can be considerable. Stochastic models may provide additional insight into the risk and probability of achieving groundwater standards.

One of the most difficult issues in groundwater management modeling is dealing adequately with the effect of model uncertainty in optimal decision making (Wagner and Gorelick, 1987). The uncertainty stems from a wide variety of factors ranging from partial knowledge about aquifer properties, its boundary conditions, land use practices, on-ground pollutant loading, soil characteristics, depth to water table, flow and transport parameters affecting pollutant fate and
transport in groundwater, to economic, regulatory and political factors. The effect of these uncertainties on groundwater management at contaminated sites has been widely reported in the literature, mostly for pumping remediation strategies (Freeze and Gorelick, 1999). The main approaches to deal with these uncertainties can be divided into classic chance-constrained programming and Monte Carlo-based methods. Chance-constrained programming allows for constraints’ violations up to preassigned probability levels, based on the derivation of deterministic equivalents of the chance-contraints (Charnes et al., 1958; Charnes and Cooper, 1963). This often involves an a priori assumption of the statistical distribution of the random variable (e.g. Tung, 1986; Wagner and Gorelick, 1987). For cases involving numerical models of complex hydrogeology, an alternative is to generate a set of equally likely multiple realizations of the hydraulic conductivity field, using then Monte Carlo analysis to assess uncertainty regarding the achievement of the environmental objectives with the optimal strategy (e.g., Wagner and Gorelick, 1989; Morgan et al., 1993; Feyen and Gorelick, 2004). Monte Carlo methods can be further subdivided into the following simulation-optimization techniques: stacking management models, Monte Carlo optimization, and mixed-integer stochastic optimization with predefined reliability. All these approaches will be subsequently discussed in the methodology section.

Most previous applications of these four approaches have focuses on “pump and treat” alternatives for optimal remediation of contaminated aquifers. Most of these studies deal with uncertainty on the hydraulic conductivity or the regional boundary conditions (e.g., Wagner and Gorelick, 1989; Feyen and Gorelick, 2004), although other sources of uncertainty have been also considered (e.g. Van den Brink et al., 2008)
This paper presents a stochastic hydro-economic modelling framework for analyzing fertilizer management strategies to control groundwater nitrate pollution under groundwater parameter uncertainty. It does not intend to discuss the choice of different policy instruments for efficient pollution control, topic for which an extensive literature already exists (e.g., Shortle and Griffin, 2001; Batie and Horan, 2004). Instead, the main contribution of this research is to analyze the influence of uncertainty in the physical parameters of a heterogeneous groundwater diffuse pollution problem on the results of fertilizer management strategies, and to introduce methods that integrate uncertainty and reliability in order to obtain strategies of spatial allocation of fertilizer use in agriculture (fertilizer standards) to meet the groundwater nitrate concentration limits required by law (e.g., EU Water Framework Directive).

The paper is organized as follows. First, we describe the proposed hydro-economic framework and analyze four different approaches (based on Monte Carlo analysis of multiple stochastic realizations) to deal with uncertainty in the pollutant concentration predictions due to uncertain in the spatial variability of the hydraulic conductivity. Then, a 2D synthetic case study is used to illustrate the application of the methodology.

2. Methods

The heterogeneity of hydraulic conductivity field has a strong influence on the migration and evolution in time and space of the pollutant concentration in groundwater and therefore on the optimal fertilizer application. The $K$ of an aquifer can vary spatially by several orders of magnitude (see, e.g., Salamon et al, 2007). To the important variability of the parameter we have to add the lack of data in most practical cases. Given the uncertainty in the conductivity, our groundwater flow and mass transport predictions, based on the conductivity fields, will be uncertain. Therefore, the uncertainty of the $K$ spatial variability should be incorporated into the decision process in order to derive a strategy to control groundwater nitrate pollution with
certain reliability. This paper presents a systematic stochastic framework, using four different formulations, to explicitly incorporate the effects of uncertainty through to the design of reliable groundwater quality schemes. The stochastic hydro-economic modeling framework has been designed for determining groundwater nitrate pollution from agriculture, considering the uncertainty in the conductivity field and the reliability in the optimal strategy designed. All the stochastic formulations are based upon the deterministic framework presented by Peña-Haro et al. (2009). A brief description of the method is provided in the next section.

The stochastic approaches for dealing with uncertainty require the generation of multiple equiprobable spatial $K$ fields (realizations), which can be obtained by means of an appropriate geostatistical approach (such as interpolation methods, sequential Gaussian or indicator simulation, conditional $K$ fields obtained from inverse models, etc.). Obviously, the uncertainty in the results will be strongly influenced by the variance of the hydraulic conductivity probability distribution and the spatial correlation structure. Therefore, the aquifer should be characterized as adequately as possible in order to obtain reliable results. Moreover, a sensitivity analysis with regard the uncertain parameters should accompany a work like this.

2.1. Deterministic hydro-economic management model

The deterministic management model for groundwater pollution control was formulated in Peña-Haro et al. (2009). A holistic optimization model is used to determine the spatial and temporal fertilizer application rate that maximizes the net benefits in agriculture constrained by the quality requirements in groundwater at specified control sites. In accordance with the WFD, the maximum concentrations at these control sites are the policy targets, which are defined by imposing legal upper bounds on the concentration level of specified pollutants in water, based on specific criteria such as adequate margins of safety for human or ecological health. A
coupled agronomic and flow and transport-groundwater modeling approach is used to quantify
the relationship between emissions (i.e., nitrogen loading rates) and groundwater quality
impacts at regulatory control sites. Specifically, Pena-Haro et al. (2009) compute unit response
functions for each source-well pair, which is generated by simulating long-term nitrate
concentration evolution at the control sites in response to uniform source loading with unit
stress. The integration of the response matrix in the constraints of the management model
allows simulating by superposition the evolution of groundwater nitrate concentration over
time at different points of interest throughout the aquifer resulting from multiple pollutant
sources distributed over time and space. Linearity of the system is required to apply
superposition; therefore groundwater flow has to be considered as steady-state. The approach
explicitly simulate the fate and transport of nitrates within the aquifer in the optimization
model, unlike methods that use black-box statistical models such as artificial neural networks
or genetic algorithms to relate on-ground nitrogen loadings with nitrate concentrations (Almasri
and Kaluarachchi, 2007; Aly and Peralta, 1999; Ritzel et al., 1994).

The benefits in agriculture were determined through crop prices and crop production functions,
being the management model for groundwater pollution control formulated as follows:

\[
\text{Max } \Pi = \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{1}{(1+r)^t} A_s \left( p_s \cdot Y_{s,y} - p_n \cdot N_{s,y} - p_w \cdot W_{s,y} \right)
\]  

(subject to:

\[
\sum_s R M_{cst,sxy} \cdot c_{r,sxy} \leq q_{cst}, \forall c, t, y
\]  

where \( \Pi \) is the objective function to be maximized and represents the present value of the
net benefit from agricultural production (€) defined as crop revenues minus fertilizer and water
variable costs (fixed costs are not included); \( A_s \) is the area cultivated for crop located at source
\( s \); \( p_s \) is the crop price (€/kg); \( Y_{s,y} \) is the production yield of crop located at source \( s \) at planning
year \( y \) (kg/ha), that depends on the nitrogen fertilizer and irrigation water applied; \( p_n \) is the nitrogen price (€/kg); \( N_{s,y} \) is the fertilizer applied to crop located at source \( s \) at year \( y \) (kg/ha), \( p_w \) is the price of water (€/m\(^3\)), and \( W_{s,y} \) is the water applied to crop located at source \( s \) at each planning year \( y \) (m\(^3\)); \( r \) is the annual discount rate, \( RM \) is the unitary pollutant concentration response matrix where each column is the nitrate concentration for each crop area (\( s \)) times de number of years within the planning horizon (\( y \)), the number of rows equals the number of control sites (\( c \)) times the number of simulated time steps (\( t \)) in the frame of the problem; \( q \) is a vector of water quality standard imposed at the control sites over the simulation time (kg/m\(^3\)); \( cr \) is a vector representing the nitrate concentration recharge (kg/m\(^3\)) reaching groundwater from a crop located at source \( s \), which is obtained dividing the nitrate leached over the water that recharges the aquifer. Both nitrate leached and crop production are represented by polynomial regression equations depending on the water and fertilizer use (see Peña-Haro et al., 2009). These equations can be derived from the results of agronomic simulations models like EPIC (Williams, 1995; Liu et al., 2007; Peña-Haro et al., 2010).

This modeling approach was developed under several assumptions:

- No crop rotation, changes in farm management practices or changes in crop patterns are considered. This issue is very important for irrigation districts and crops in which farmers may react to input regulations with changes in crop patterns and crop rotation practices. Rotation with crops like alfalfa is a useful management practice for controlling the soil nitrate content (e.g. Toth and Fox, 1998). Changes in management practices and cropping patterns are less likely in the short run than changes in the input levels (Helfand and House, 1995).

- The data on leaching corresponds to average water application rates. No dynamic changes of irrigation applications and rainfall over time are considered.
Only restrictions on fertilizer use are considered; irrigation cutting could be also a way of decreasing nitrate leaching.

The cost of the policies for controlling nitrate pollution is simplified as the direct costs to the users, in terms of net income losses. Transaction costs associated with introducing and maintaining a policy instrument are not considered, although they might be significant in certain cases.

As mentioned, this formulation assumes fixed water applications and crop locations; therefore, the word “optimal” is used hereinafter to refer just to the fertilization rates resulting from the optimization problem defined for controlling groundwater nitrate pollution and not to better irrigation plans or the most environmentally appropriate locations for growing crops.

The optimization problem is coded in GAMS, a high-level modeling system for mathematical programming problems (GAMS, 2008a).

2.2. Stochastic hydro-economic approaches

The framework allows considering four different stochastic approaches to analyze groundwater quality management under parameter uncertainty:

2.2.1. Reliability of deterministic optimization. Monte Carlo simulation with pre-assumed ("true") parameter field.

The objective is to evaluate the reliability of the optimal fertilizer application for an aquifer with a pre-assumed heterogeneous hydraulic conductivity field. This is carried out by assuming one of the multiple $K$ fields generated as the “true” hydraulic conductivity field (e.g., Bark et al., 2003; Ko and Lee, 2008), and determining the corresponding optimal fertilizer application. The reliability of meeting the standard (or probability of not failure) and the uncertainty of the
pre-assumed optimal application are evaluated by simulating the resulting fertilizer allocation for the series of random fields stochastically generated, and testing whether the maximum concentrations are reached or not.

2.2.2. Uncertainty on optimal fertilizer application. Monte Carlo optimization

Monte Carlo management models solve the nonlinear simulation-optimization problem individually for each one of a series of multiple equiprobable realizations obtained using an appropriate geostatistical model. Because of its simplicity, this approach has been widely applied to the design of optimal groundwater remediation strategies (e.g., Gorelick 1983; Wagner and Gorelick, 1989; Freeze and Gorelick 1999; Feyen and Gorelick, 2004; Lacroix et al., 2005; Ko and Lee, 2008; Van den Brink et al., 2008). In this approach, a series of individual optimization problems are solved, each for a single realization of hydraulic conductivity. Each one of the fertilizer applications obtained represents a random sampling from the cumulative density function (CDF) of optimal fertilizer application rates. Therefore, the results of the Monte Carlo hydro-economic modeling can be used to characterize the probability distribution of the optimal fertilizer application rates.

2.2.3. Multiple realizations or stacking management approach

In the multiple realization or stacking approach the nonlinear simulation-optimization problem is simultaneously solved for a set of different scenarios representing uncertainty, e.g., by using a sampling of hydraulic conductivity realizations generated using geostatistical techniques (e.g., Wagner and Gorelick, 1989; Aly and Peralta, 1999; Feyen and Gorelick, 2004 and 2005; Ko and Lee, 2009). However, this approach does not allow a priori definition of the system reliability. The reliability is determined through post-optimization Monte Carlo analysis on a
much larger set of realizations that were used in the stack. The mathematical formulation of the
multiple realization groundwater quality management model consist of maximize (1) subject to:

\[
\sum_x RM_{(x,t,xy)} \cdot q_{(x,t)} \leq q_{(x,t)} \quad \forall i, c, t, y
\]

where an additional component (i) is added to the RM matrix considered in the deterministic
hydro-economic management model. This component is made up of as many elements as
realizations of the random conductivity field are simultaneously considered in the management
model. That is, the optimization problem is solved for \(i = 1, \ldots, s_n\), where i represents a hydraulic
conductivity realization, and \(s_n\) is the stack size, i.e, the number of hydraulic conductivity
realizations included in the stochastic management model. The optimization problem retains
the same number of decision variables as the deterministic model, but the number of
concentration constraints is increased by a factor of \(i\). The reliability is determined through
post-optimization Monte Carlo analysis on a much larger set of realizations that were used in
the stack.

### 2.2.4. Mixed-integer stochastic optimization with predefined reliability

Morgan et al. (1993) introduced a mixed-integer approach to solve the problem of optimal
groundwater remediation design with a certain degree of reliability. The approach combines the
advantages of the simulation-optimization models with those of the chance-constrained models.
In this case, the user selects the desired degree of reliability, which is accomplished by
allowing a certain number of the Monte Carlo realizations to fail. Other authors have also
applied this technique to groundwater remediation (e.g., Ritzel et al., 1994;, Dhar and Datta,
2007; Ng and Eheart, 2008), which has also been termed as mixed-integer-chance-constrained
programming (MICCP) (Morgan et al., 1993).
We have reformulated the approach presented by Morgan et al. (1993) to deal with nitrate pollution abatement in order to meet certain groundwater quality standards, like the ones ruled by the EU Water Framework Directive. The proposed stochastic management problem was defined as finding the optimal fertilizer allocation (for a certain crop distribution) that maximizes the welfare from crop production that meet the groundwater quality constraints with a certain reliability.

The chance-constrained problem is reformulated as a Mixed Integer Non-Linear Programming (MINLP). As in Morgan et al. (1993), the stochastic nature of the conductivity field is analyzed through Monte Carlo realizations, and multiple realizations make up the constraint sets of the optimization model (in this case, represented by pollutant concentration response matrices, as in Peña-Haro et al., 2009). The desired reliability of the system is predetermined by fixing the number of constraints that may be violated, which is done by replacing equations (1) and (2) with equations (4) to (7).

The stochastic method is formulated as follows:

\[ \text{Max } \Pi = \sum_{s} \sum_{y} \frac{1}{(1+r)^{y}} A_i \left( p_s \cdot Y_{s,y} - p_u \cdot N_{s,y} - p_w \cdot W_{s,y} \right) - M \cdot \sum_i F_i \]  \hspace{1cm} (4)

subject to:

\[ \sum R M_{(cxt,sxy)} \cdot c r_{(sxy)} - M \cdot f_{(cxt)} \leq q_{(cxt)} \hspace{0.5cm} \forall i,c,t,y \]  \hspace{1cm} (5)

\[ \sum f_{(cxt)} \leq M \cdot F_i \hspace{0.5cm} \forall i \]  \hspace{1cm} (6)
\[ \sum_{i} F_i \geq NF \quad i = 1, \ldots, NR \quad (7) \]

where:

- \( M \) is a large positive number;
- \( i \) is the hydraulic conductivity realization number,
- \( NR \) is the total number of realizations;
- \( f \) is a matrix of binary integer variables, where \( i \) is the realization number, \( c \) refers to control site, \( t \) stands for simulated time step, and \( y \) is the planning year. The matrix represents the individual failures, and its components take the value 1 if the quality standard is exceeded at any time in any control site, and 0 otherwise;
- \( F \) is a binary vector with \( i \) elements showing realization failures. It takes the value 1, thus representing a failure, if the quality standard is exceeded in at least one time step at any control site for a certain realization \( i \), and 0 otherwise.
- \( NF \) is the number of realization failures that are allowed, defined in accordance with the desired reliability level, \( R \), which is given by,

\[ R = 1 - \frac{\sum_{i} F_i}{NR} \quad (8) \]

Therefore, reliability is maintained by constraining the number of failures allowed. Note that with this formulation for each realization \( i \), a failure \((F_i=1)\) is considered when the quality standard is not met, independently on how many times or in how many control sites the quality standard is exceeded. Therefore, for a single realization \( i \), \( f \) may exceed the quality standard in several times steps or control sites, thus leading to define \( F_i \) as a failure, i.e., \( F_i = 1 \). Finally, failures are penalized in the objective function as shown in (4).
Unlike the classic chance-constrained applications (e.g., Tung, 1986; Wagner and Gorelick 1987; McSweeny and Shortle, 1990; Wagner 1999), this formulation considers uncertainty in the response matrix coefficients and does not require a priori definition of the distribution, as it is required in the “classic” chance-constrained programming (Charnes et al., 1958; Charnes and Cooper, 1963), in which the problem has been usually solved by transforming the probabilistic constraints to deterministic equivalents given knowledge. The deterministic equivalent-based methods can be solved by linear or nonlinear programming methods (e.g., Charnes and Cooper, 1963; Kataoka, 1963). Nevertheless, they entail a series of drawbacks such as requiring assumptions of parameter distributions that may induce to errors (for mathematical convenience, the most widely used statistical model is the normal distribution, e.g., Tung, 1986; Wagner and Gorelick, 1987), or to be unsuitable for complex nonlinear problems where the deterministic equivalent may be difficult or even impossible to establish. Furthermore, they become cumbersome whenever reliability is defined as the probability of meeting a set of constraints simultaneously, rather than one or more (Ng and Eheart, 2008). All these problems are avoided with the approach here presented. However, the chance-constrained approach has also the advantage of requiring less computational effort that the multiple realization model or the mixed-integer models.

3. Illustrative example

The aquifer system configuration is the same that the one used by Peña-Haro et al. (2009), which apply the deterministic formulation to a 2D homogeneous synthetic aquifer. In this case, however, we consider heterogeneous hydraulic conductivity. The aquifer has impermeable boundaries and steady-state flow from top to bottom of the domain. The finite difference grid is $500 \times 500$ meters, and the domain has 58 rows and 40
columns. A confined aquifer has been modeled with a thickness of 10 meters, effective porosity of 0.2, and dispersivity of 10 meters. The natural recharge is 500 m$^3$/ha. There are 70 stress periods, each of one year (365 days). Seven different crop zones (pollution sources or just sources in our model formulation) with five different crops are considered. For each crop a quadratic production function and a leaching function have been defined (Peña-Haro et al., 2009). Each source is related to a crop as shown in Figure 1. Three control sites with concentration upper bounds of 50 mg/l of nitrates, as established by the EU water legislation, are imposed.

The four different stochastic formulations described in section 2 were considered in order to analyze the effect of parameter uncertainty on “optimal” groundwater management and reliability in meeting the water quality standards. A 40 year planning horizon was considered for each scenario, with a constant annual fertilizer application during the 40 years. All the optimization models are coded in GAMS (GAMS, 2008a). The nonlinear optimization models were solved using CONOPT (Drud, 1985), which is based on the Generalized Reduced Gradient algorithm designed for large programming problems. The optimization problem reformulated as a MINLP is also coded in GAMS, using the SBB solver (GAMS, 2008b), combination of the standard Branch and Bound method and a standard nonlinear programming solver (CONOPT in this case). The prepossessing of all the required information in the format required by the GAMS code for the optimization models, including the simulations of the $K$ fields and the generation of the pollutant concentration response matrices, has been automated by means of a “batch” file.
3.1. Simulation of K fields

The different stochastic optimization management formulations require the generation of multiple K fields. The simulation of these K fields, in the 2D synthetic case stated above, has been performed by means of a sequential Gaussian simulation using the computer code GCOSIM3D (Gómez-Hernández and Journel, 1993). The stochastic structure is assumed to be common for all simulated K fields, which simplifies the analysis avoiding the uncertainty on the stochastic structure. Because of this, all K fields are equally likely realizations, and therefore, are plausible representations of reality because they are conditional to the same data and display the same degree of spatial variability.

The stochastic structure has been defined by using a spherical variogram with a range approximately equal to 1/5 of the aquifer size, 0.5 of nugget effect, and sill of 4. The effect of different degrees of heterogeneity of the parameters in the aquifer has been studied. Specifically, a sensitivity analysis by considering two different variances of the hydraulic conductivity distribution has been carried out, both with a normal distribution with mean 40 m/day and with variances of 15 m²/day² (referred as “case 1”) and 60 m²/day² (“case 2”). A hundred realizations were generated for each case. We assume that this set is large enough to provide a significant representation of the variability of the parameter. Figure 2 illustrates the K field for the realization #1, while Figure 3 shows the frequency distribution and univariate statistics for all K realizations.

[Figure 2]

[Figure 3]
3.2. Pollutant concentration response matrices

Once the different conductivity fields were generated the pollutant concentration responses from unit recharge rates at the sources were simulated. The pollutant response matrix describes the influence of pollutant sources upon concentrations at the control sites over time. The simulated time horizon corresponds to the time for the solute to pass all the control sites, and it is independent of the length of the planning period. To construct the pollutant concentration response matrix the flow and transport governing equations must be solved. MODFLOW (McDonald and Harbaugh, 1988), a finite difference groundwater flow model, and MT3DMS (Zheng and Wang, 1999), a solute transport model were used. A pollutant concentration response matrix was generated for each k field realization.

3.3. Reliability of the deterministic optimization. Monte Carlo simulation.

The purpose is to assess the probability of meeting the quality standard for a policy that has been designed without taken into account hydraulic conductivity uncertainty. For this case, one of the realizations (realization 14) was chosen as the “true” K field. The resulting optimal fertilizer application is then tested on the random fields generated to check the reliability of meeting the water quality standard (Monte Carlo simulation). Figure 4 shows the reliability or probability of not exceeding certain nitrate concentration level for the two cases with k fields with different variances, obtained from the maximum concentration values simulated at each conductivity field for the optimal fertilizer application of the “true” parameter field. The reliability level of the pre-assumed optimal policy for meeting the quality standard was only a 14% for case 1 (i.e., only in 14 realizations out of the 100 simulated nitrate concentrations did not exceed the limit of 50 mg/l), and 24% for case 2. It is clear, however, that this reliability levels will highly depend on the realization chosen to find the optimal management (the chosen “true” field). With a larger variance, although the reliability of meeting the standard is higher
the range of probable maximum concentrations increases, what can be relevant for the design of risk-averse policies.

Reducing the fertilizer application rate, the reliability level can be increased up to 100%. For case 2, the mean application rate has to be reduced by 20% to obtain a global 100% reliability when checked with the 100 realizations. This result was obtained by lowering the constraining quality standard (to 30.6 mg/l), as proposed by Ko and Lee (2008) for the analysis of the optimal remediation design of a contaminated aquifer. Although this fertilizer management achieves 100% reliability, it is important to note that this strategy is not necessarily the “optimal” policy for 100% reliability. This fact will be further discussed in the section of the mixed-integer stochastic approach. The alternative with a 24% reliability level produces a total annual net benefit of 20.8 M€. With 100% reliability (20% fertilizer reduction), the total annual net benefits is reduced to 19.7 M€.

3.4. Uncertainty on optimal fertilizer application. Monte Carlo optimization

In this formulation, the uncertainty is considered by solving the hydro-economic optimization model for each of the 100 individual realizations and comparing the corresponding results. This approach can be used to characterize the probability distribution of the optimal fertilizer application rates (Figure 5). The mean for the case 1 (variance of 15 m³/day²) is 138.3 kg/ha, the standard deviation is 2.9, and the mean fertilizer rates range from 131.4 to 148.5 kg/ha. However, it cannot be assured that all these strategies would have a high probability of meeting the standard, what limits the applicability to make decisions. In order to estimate the reliability of meeting the objectives of any of the specific strategies that we obtain, we have to simulate the strategy with the complete set of realizations (post-optimality Monte Carlo simulation). For
example, for the strategy that corresponds to the mean fertilizer application, the reliability of meeting the standard is 33%.

For case 2 (variance of 60 m$^2$/day$^2$), the mean value is similar (138.9 kg/ha), while the standard deviation goes up to 5.3. The reliability of the strategy corresponding to the mean rate is 35%.

The results show more dispersivity and a broader range of possible values of the mean fertilizer rates obtained from a single-realization optimization, and therefore, a greater variability of the economic impact of the strategy for a larger variance in the $K$ fields.

[Figure 5]

3.5. Stacking approach for optimal fertilizer allocation

For this formulation, the hydro-economic management model is solved only once, simultaneously for the complete stack of 100 realizations of the random conductivity field; therefore, only one optimal fertilizer application is obtained. Chang (1993) investigated the number of realizations to be included in the staking in order to achieve a certain level of reliability, using a Bayesian framework. He obtained the following relationship between stack size (number of realizations, NR) and reliability (R):  

$$R = \frac{NR + 1}{NR + 2}$$ (9)  

Feyen and Gorelick (2004) concluded that the previous relationship obtained by Chan (1993) overestimates the reliability for different stack sizes, and presents a formula that provides expected reliability as a function of the number of realizations in the stack and the variance of the log hydraulic conductivity $\sigma^2$ as:
For our case, the application of equations, (9) and (10) yields the same reliability for the 100 realizations, 99%. Therefore, the size of the stack is considered big enough to assess reliability levels.

As expected, the groundwater quality standard is not exceeded when simulating the optimal strategy for all the realizations of the stack (Figure 6). The reliability will be therefore 100%, assuming the 100 set of realizations as a representative measure of k variability. The total net benefit of the optimal solution with a 100% reliability is higher for case 1 (20.22 M€/year) than for the case with a larger variance (19.89 M€/year), since in the latter the fertilizer application has to be lower in order to meet the standards. However, the use of this approach does not allow for prespecification of the desired system reliability. Since the reliability of the system management is not explicitly considered in the optimal solution, the method can lead to conservative (and more expensive) solutions.

Figure 7 shows the influence of the stack size in the reliability of the resulting strategies. Post-optimization Monte Carlo reliability analyses were carried out by simulating each optimal solution against the set of a hundred different hydraulic conductivity realizations. As expected, the reliability of the optimal solution increases with the stack size. The values of reliability versus stack size are in agreement with the findings of other authors (Wagner and Gorelick, 1989; Chan 1993 and Ko and Lee, 2009) for the optimal remediation design to control
groundwater pollution, and Feyen and Gorelick (2004) for controlling groundwater outflow in wetlands. The results also show that a high reliability can be achieved with a stack of a reduced number of realizations.

3.6. Mixed-integer stochastic approach with predefined reliability

In this formulation, the stochastic nature of the conductivity field is considered in the decision-making process by integrating the complete set of Monte Carlo realizations through the response matrix of the optimization management model. The method guarantees the optimal solution for a pre-specified reliability level by using simultaneously all the generated realizations and fixing the number of constraints that may be violated. Different reliability levels were tested. The range of the maximum concentration values that are reached decreases with increasing reliability of meeting the standard, and a steeper slope of the probability curve is observed (Figure 8). The worst-case (upper value) of the maximum nitrate concentrations increases with decreasing reliability (Figure 9). The larger the variance, the greater the range and the worst-case (maximum concentration values). These results tell us that with a high variance, a risk-averse decision-maker would prefer a more costly strategy with higher reliability of meeting the standard than in the case of low variance, in order to reduce the risk of a reaching a high nitrate concentration exceeding by far the standard (which will implies higher economic impacts in terms of environmental and resource costs).

[Figure 7]

[Figure 8]
The objective function (the total net benefit) increases nonlinearly with decreasing reliability (Fig. 10). This implies that a larger amount of net benefit has to be sacrificed when a more risk-averse management is considered.

For the same reliability level, the total net benefit is greater for a lower variance. A high variance also implies that some critical realizations further limit the fertilizer application rate for that reliability level. As the reliability level gets lower, the total net benefit for both $K$ variance fields gets closer, since the fertilizer rate moves toward the optimal application that yields the maximum benefits. For each realization, the influence of the different sources upon the concentration at the control sites is different, and the corresponding benefits from crop production will differ. Table 1 shows the percentage of fertilizer reduction that produces the maximum crop yield that is required to meet the groundwater quality standards for different levels of reliability. These results are relevant for the design of optimal land use policies to control groundwater nitrate pollution. From the table we can see that no fertilizer reduction is need in certain areas, while in the other areas the reduction has to be greater in order to achieve a higher reliability of meeting the standards. The pattern of the spatial fertilizer reduction is maintained for the different reliability levels, showing the robustness of the solution.

4. Discussion
The four stochastic modeling approaches aforementioned have been applied to analyze how uncertainty of hydraulic conductivity leads to different reliability levels of meeting the quality standards. Eventually, this is translated into different optimal fertilizer application rates, and therefore, different net benefits (or reduction of income losses). The four approaches tackle this problem from different points of view. Some important insights can be drawn from the results above presented.

First, we have assessed the reliability of the policy derived from the deterministic optimization for a pre-assumed parameter field. The chosen K field is not necessarily true, and therefore, the obtained optimal fertilization scheme could succeed or fail in meeting the groundwater concentrations standards when applied to random K fields by means of Monte Carlo simulations. As it has been shown, this formulation may lead to low reliability levels. Hence, this formulation is not recommended to derive reliable policies (especially in very heterogeneous aquifers) and should be discarded in the decision making process. Although we can artificially reduce the constraining quality standard in order to achieve a higher reliability in meeting the 50 mg/l of groundwater nitrate concentration, it has been proved that this solution does not necessarily yield the maximum for the objective function (total net benefits).

The Monte Carlo optimization approach can be used to characterize the probability distribution of the mean optimal fertilizer application rates. A post-optimality Monte Carlo simulation is required to estimate the reliability of meeting the standards. Results from these post-simulations have shown that the mean value of the probability distribution can lead, again, to low reliability levels regardless of the variance of the K fields. The different strategies of fertilizer application rates may have a high probability of not meeting the standard, what limits the applicability to make decisions. On the other hand, a choice of a more restrictive fertilizer application (e.g., the lower quartile value of the distribution) could result in too conservative (and more expensive) solution.
Contrary, in the stacking approach, the fertilizer standard resulting from the optimization model fulfills the quality standards for all the realizations (Figure 6); therefore, the relativity level is equal to 100%, assuming that the set of K realizations used in the stacking is large enough to provide a significant representation of the parameter variability. The literature has provided formulas that relate the number of realizations to include in the staking in order to achieve a certain level of reliability. Our results are in line with those presented by other works related to pumping remediation of aquifer pollution (e.g., Chan, 1993; Feyen and Gorelick, 2004). By means of a post-optimization Monte Carlo analysis, the results show that high reliability levels (greater than 90%) can be reached with a small stack sizes (Figure 7). However, since the reliability of the system management is not explicitly considered in the optimal solution, the method can lead to conservative and less economic efficient solutions.

This problem is overcome by resorting to a mixed-integer stochastic approach with an a priori defined reliability level, allowing a certain number of simulations to fail the standards. The higher the predefined reliability level and the lower variance, the lower the minimum concentration that can be reached (Figure 8 and Figure 9). In addition, the lower the variance, the higher the benefits (Figure 10). As a result, this approach leads to less costly and more reliable solutions than in the staking approach, guaranteeing the “optimal” strategy of spatial fertilizer application (maximum total benefit) for a fixed reliability level.

5. Conclusions

A stochastic hydro-economic modeling framework for optimal management of groundwater pollution under K uncertainty has been presented. A holistic optimization model determines the spatial and temporal fertilizer application rate that maximizes the net benefits in agriculture constrained by the groundwater nitrate concentration standards at various control sites. The stochastic management framework presented allows to derive least-cost fertilizer plans in order
to meet the groundwater quality standards ruled by the EU Water Framework Directive or any other water legislation under conditions of parameter uncertainty. As shown in the results, parameter uncertainty leads to different management policies with clear implications in reliability levels, costs and benefits. The study of the least-cost alternative for meeting the environmental objectives is also important in order to justify potential time and objective derogation when disproportionate costs are identified (WFD, art. 4).

Four different formulations (Monte Carlo simulation with preassumed parameter field, Monte Carlo optimization, stacking approach, and mixed-integer stochastic optimization with predefined reliability level) have been applied in order to analyze the influence of the uncertainty of the spatial variability of the hydraulic conductivity upon the optimal management of groundwater nitrate pollution from agricultural sources. All the approaches use a Monte Carlo-type analysis involving a series of realizations of the uncertain parameter, in order to assess reliability and uncertainty of different fertilizer application strategies. These results represent an upper bound or benchmark comparison to possible second-best solutions for controlling nitrate pollution, like economic taxes or incentives either on inputs or ambient standards.

The framework has been applied to a controlled 2D synthetic aquifer system, offering insights into the impacts of uncertainty in the optimal management strategies. Given the uncertainty in the pollutant concentration predictions due to uncertain spatial variability of the hydraulic conductivity, the solution of the optimization of a single realization does not guarantee a high reliability in meeting the groundwater quality standards. A stochastic analysis that considers uncertainty in the performance of the system allows providing more reliable management strategies than deterministic models.

In order to increase the reliability, we can simultaneously optimize for a sampling or stack of hydraulic conductivity realizations (stacking approach). The reliability of the optimal solution
increases with the stack size. However, this approach does not allow for pre-specification of the desired system reliability, and the method can lead to too conservative solutions.

In decision-making processes, reliability and risk-aversion play a decisive role. By using a mixed-integer stochastic formulation, an a priori reliability level of the strategy can be explicitly fixed. As the mixed-integer stochastic model includes the complete set of realizations, it guarantees the best optimal strategy (maximum total net benefit) for that level of reliability, as shown by the results. This approach also allows deriving the trade-off curve between the reliability level and the net benefits.

In a risk-averse decision-making, not only the reliability of meeting the standards counts, but also the probability distribution of the maximum pollutant concentrations. A risk-averse decision-making is specially justified when dealing with well-capture zones for drinking water supply (health risk) or sensitive areas of groundwater dependent ecosystems. A sensitivity analysis was conducted to assess the influence of the variance of the hydraulic conductivity fields on the optimal strategies. The results have shown that the larger the variance, the greater the range of maximum nitrate concentrations and the worst-case (or maximum value) that could be reached for the same level of reliability of meeting the standard.

In the reliability versus net benefit trade-off, for the same reliability level, the total net benefit is greater when the variance is lower. Note that by assuming uncertainty in the random function (e.g., Llopis-Albert and Capilla, 2009) or by considering higher variances of the $K$, a greater influence in the results than in the analyzed cases should be expected.

The uncertainty can be reduced by improving the site characterization, providing more realistic and reliable management schemes. For that purpose, a promising extension of the present work is the integration of a stochastic inverse model in the described framework, in which the stochastic simulations are constrained to data such as hydraulic conductivity, piezometer head, solute concentrations, travel times or secondary data obtained from expert judgment and
geophysical surveys. The influence of the K uncertainty is only analyzed for a fertilizer standards policy. There is a broad range of policies for controlling nitrates in the literature (standards or economic instruments on inputs, emissions or ambient concentrations) (Shortle and Griffin, 2001). A further extension of this work is to incorporate these different policies into the hydro-economic formulation in order to compare their effectiveness in controlling nitrate pollution as second-best solutions.

Besides groundwater hydraulic conductivity, there are many other sources of uncertainty, ranging from partial knowledge about the aquifer properties and boundary conditions, land use practices, on-ground nitrogen loading, nitrogen soil dynamics, soil characteristics, depth to water table, to the diverse economic, regulatory and political factors. The analysis of uncertainty and risk can be also extended to the derived health risk problem (Lichtenberg et al., 1989; Innes and Cory, 2008). Further research is required in order to represent the diversity of potential on-farm management decisions and other policy options rather than fertilizer use, and to extend the analysis to other sources of uncertainty.

Finally, the method can be extended to consider other sources of nitrate pollution such as animal farming, landfills, and septic tanks. Although the method and tools are suitable for simulating the effects of these sources on nitrate concentration at the control sites, further research would be required for modeling the economics of abating the pollution from these other sources.

References


Figure 1. Aquifer system
Figure 2. $K$ field for realization #1 and variances of 15 (left) and 60 (right)
Figure 3. Frequency distribution and univariate statistics for all K realizations and variances of 15 (left) and 60 (right)
Figure 4. Reliability (probability of not exceeding the maximum nitrate concentration) of the optimal fertilizer application for realization 14 with variances of 15 and 60.
Figure 5. Probability distribution of the mean fertilizer application, case 1.
Figure 6. Maximum nitrate concentration vs. reliability of not exceeding the nitrate concentration (post Monte Carlo simulation).
Figure 7. Mean reliability (post Monte Carlo simulation) vs. stack size.
Figure 8. Probability of not exceeding the nitrate concentration for different reliability levels.
Figure 9. Reliability vs. upper value of maximum nitrate concentrations
Figure 10. Trade-off between reliability and benefits
Table 1. Percentage of spatial fertilizer reduction for different levels of reliability

<table>
<thead>
<tr>
<th>Crop Area</th>
<th>Reliability</th>
<th>100%</th>
<th>80%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>3.29%</td>
<td>2.58%</td>
<td>2.12%</td>
<td></td>
</tr>
<tr>
<td>s2</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>s3</td>
<td>43.07%</td>
<td>29.22%</td>
<td>22.17%</td>
<td></td>
</tr>
<tr>
<td>s4</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>s5</td>
<td>17.99%</td>
<td>14.00%</td>
<td>11.48%</td>
<td></td>
</tr>
<tr>
<td>s6</td>
<td>2.75%</td>
<td>2.15%</td>
<td>1.76%</td>
<td></td>
</tr>
<tr>
<td>s7</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td></td>
</tr>
</tbody>
</table>