Nitrates/nitrites concentration estimation in wastewater samples using transmittance curve models optimized by evolutionary computation techniques

Pablo López-Espí, Sancho Salcedo-Sanz, Ángel M. Pérez-Bellido, Irene de Bustamante and Francisco López-Ferreras

ABSTRACT

This paper presents an application of evolutionary computation algorithms in the estimation of the concentration of nitrates/nitrites in water. Specifically, we start from the measured transmittance curves of a water sample and a model of the curve consisting of a mixture of polynomial, Fermi and Gaussian functions. An evolutionary programming algorithm is then used to obtain the optimal parameters of the model which minimize the distance between the measured and the modeled transmittance curves. This process allows us to separate the modeled transmittance curve into several components, one of them associated with the nitrate/nitrite concentration. We can extract the nitrate/nitrite concentration of the water sample using this component of the transmittance model. We test our proposal on several laboratory samples and on three real samples measured in different locations around Madrid, Spain.

Key words | evolutionary computation, evolutionary programming, nitrates/nitrites concentration estimation, ultraviolet spectroscopy

Pablo López-Espí
Sancho Salcedo-Sanz (corresponding author)
Ángel M. Pérez-Bellido
Francisco López-Ferreras
Departament of Signal Theory and Communications,
Universidad de Alcalá de Henares,
Campus Universitario,
E-28871, Alcalá de Henares Madrid,
Spain
Tel.: +34 91 885 6731
Fax: +34 91 624 8749
E-mail: *sancho.salcedo@uah.es*

Irene de Bustamante
Department of Geology,
Universidad de Alcalá de Henares,
Campus Universitario,
E-28871, Alcalá de Henares Madrid,
Spain

INTRODUCTION

Nitrogen is the most common pollutant found in groundwater. In many cases, nitrates are found in dangerous concentrations mainly due to agricultural activities: nitrogen is a vital nutrient to enhance plant growth. This fact has motivated the intensive use of nitrogen-based fertilizers to increase the productivity of crops in many regions of the world (Almasri & Kaluarachchi 2004, 2007). Thus, it is known that agricultural activities are the main source of elevated nitrate concentrations in groundwater (Almasri & Kaluarachchi 2007). The problem is that, in many cases, application of nitrogen-rich fertilizers exceeds the plant's demand, so the nitrogen reaches the groundwater. It has been reported that elevated nitrates concentrations in drinking water can cause different serious health problems in children and adults (Wolfe & Patz 2002), so the detection and control of these contaminants are very important tasks in which a huge amount of research is being carried out. Different methods have been proposed to accurately estimate the concentration of nitrates in water. We briefly summarize them in the next subsection, before stating the objective and structure of this paper.

doi: 10.2166/hydro.2010.016

Measurement techniques for nitrates concentration estimation

Ionic chromatography (IC) is one of the most used methods to estimate nitrates concentration. In IC, a process of ionic interchange is used to separate molecules with different charges. IC allows us to determine very low concentration values of contaminants, so it is really useful in water quality studies (Holm *et al.* 1997). Another interesting property is that IC is able to detect different ions from nitrates: however, the measurement time is relatively large, depending on the type of sample (Holm *et al.* 1997).

Another simple form of contaminant concentration analysis is the use of colorimetric techniques. These methods use several components which dye the sample when it contains nitrate. The process is practical, but different studies have shown that the color intensity depends on the salinity of the sample. In addition, this process may produce a toxic residue which must be eliminated by special procedures (Gross *et al.* 1999).

Also, selective electrodes for nitrate are used to measure its concentration. This method is based in the setting of a material for the ionic interchange deposit in PVC membranes. The measurement range of this method is high, but it is very sensitive to interference, especially by chlorine ions, which implies the necessity of a frequent calibration process (Karlsson *et al.* 1995).

Spectrometry has been commonly applied to the identification or determination of the concentration of contaminants in water (Thomas *et al.* 1990*a*,*b*; Beemster & Kahle 1995; Karlsson *et al.* 1995; Holm *et al.* 1997; Gross *et al.* 1999; Johnson & Coletti 2002). Specifically, several methods to measure the nitrates/nitrites content of water have been proposed using ultraviolet spectrophotometry: In Gross *et al.* (1999) a method based on the light absorption at 220 nm produced by the presence of nitrates in water is proposed. However, in this work it is recognized that organic matter and some metallic ions can cause interference in the determination of nitrate concentration. A preprocessing step by treating the sample with persulfate is proposed in order to eliminate the organic matter. In Gross *et al.* (1999) this method was used to determine the nitrate content in natural and wastewaters.

In Thomas *et al.* (1990*a*,*b*) an adaptation of the organic matter response is proposed. This method consist of modeling the organic matter spectral response as a polynomial, and then the nitrate concentration is determined by multilineal regression. This method is interesting since no reactants are needed to eliminate the organic matter and it is not necessary to filter the samples, except in the cases when the sample is very turbid.

In Karlsson *et al.* (1995) the nitrate concentration analysis was performed using UV spectroscopy, measuring from 180 to 820 nm, and using a method of partial adjusting using mean squares. The authors applied this method to obtain good quality measures of nitrate concentrations in the water of Stockholm's zone. Using a similar idea, in Beemster & Kahle (1995) a method to measure nitrate content in water was proposed based on concentration patterns, constructed by means of a combination of real and pure water with known nitrate concentrations. From these concentration patterns, a multi-variant analysis was used to obtain the real concentration of the sample.

As has been shown, the majority of the previous works which use a measurement of absorption at one or a few wavelengths suffer from difficulties caused by the interaction of organic matter or other contaminants which interfere with the truly nitrate/nitrite response. For example, sulfates and ions of iron, lead or mercury present absorption in the same range as nitrate and nitrite (about 220–230 nm). The treatment of the water sample with reactants in order to eliminate the interfering substances, or modeling the different interfering substances's responses, are the main methods used up until now to obtain accurate measurements of nitrate/nitrite concentrations.

Objective and structure

In this paper we propose a novel method to measure the nitrate/nitrite concentration in water samples. Since high values for nitrate and nitrite concentrations usually refer to wastewater samples, the proposed method is especially well suited for this kind of sample, but it can also be applied to any other water sample. The method we propose consists of using a model of the transmittance curves of nitrate/nitrites and an evolutionary programming algorithm to obtain the optimal parameters of the model. First, the total transmittance curve of the sample is modeled as the product of a number of individual contaminants, in which any number of contaminants can be included. We consider that each contaminant response can also be modeled with a mix of a Fermi and a Gauss distribution (except for the organic matter, which can be described using a second-order polynomial, because it does not reach a maximum in the studied wavelengths). Then, the parameters of each contaminant model must be optimized. This optimization is performed in such a way that the contributions of the individual contaminant responses are as close as possible to the total measured response (the quadratic error function between the measured and the optimized curves is used as the objective function to be

minimized). In order to do this optimization, a class of Evolutionary Programming (EP) algorithm has been applied (Yao *et al.* 1999). EP is an algorithm of the evolutionary computation family, which has been frequently used in continuous optimization. It is especially well suited to difficult or ill-defined problems, since the method does not require any characteristics of the objective function (continuity, derivability) that are needed in the case of other techniques, such as gradient-based algorithms, for example. In addition, EP, as in other evolutionary techniques, is robust, in the sense that it can tackle different optimization problems (with different classes of constraints or peculiarities) with very few changes in their structure. Specifically, the EP algorithm has been implemented to obtain the optimal parameters of the Fermi–Gauss transmittance model considered in this paper to obtain the nitrate/nitrite concentration of water samples.

The rest of this paper is structured as follows: the next section introduces the model for the transmittance curves of nitrates/nitrites, using a mix of Fermi and Gauss functions. The third section describes the Evolutionary Programming algorithms used in this paper in order to optimize the model's parameters. Then the performance of our proposal is shown by analyzing the nitrates/nitrites concentration of several laboratory samples and also three different real transmittance curves measured in aquifers of Madrid, Spain. The final section closes the paper, giving some final remarks.

A MODEL FOR NITRATE/NITRITE TRANSMITTANCE CURVES

It is well known that water contaminants derived from nitrogen (nitrates and nitrites) absorb energy in the ultraviolet range, with a maximum of absorption around 220–230 nm. Therefore, spectral analysis for different concentrations of nitrates/nitrites in pure water provides pattern transmittance curves, which might be used as a template in order to obtain the nitrates/nitrites content of a sample. The problem is that several contaminants may interfere with the spectral response of nitrates/nitrites, so a direct comparison of the patterns with real samples may lead to over-estimations of the nitrates/nitrites concentration. As has been mentioned above, organic matter, sulfates and different ions of iron, lead or mercury present absorptions at similar wavelength ranges as nitrogen contaminants. In addition, turbidity is also a source of error in these measurements. The possibilities to accurately measure the concentration of nitrates/nitrites are then to use a method to eliminate the different interference contaminants and turbidity influence, or modeling the spectral response of nitrates/nitrites in such a way that the interference can be corrected using adjusting or statistical methods. This paper is based on the latter approach, by defining a novel model for the transmittance curves.

The first step of our method consists of accurately modeling the transmittance shape of nitrate/nitrite solution in pure water using a set of known functions. As an example consider Figure 1, which shows an example of a measure of the transmittance curve for a concentration of nitrites of 100 ppm. We have chosen a Fermi and a Gauss function to obtain this model. Selection of these two type of functions is not random: the processes of spontaneous light emission and absorption in light sources based on semiconductors follow a Fermi function, whereas light sources are usually modeled using Gauss functions.

The Fermi function has the following expression:

$$f_F(x) = A_1\sqrt{x}\exp\left[-(x/\sigma_1)\right] \tag{1}$$

where constants A_1 and σ_1 control the maximum/minimum value and the curve width at half-amplitude. On the other hand, a Gauss-type function has the following expression:

$$f_G(x) = A_2\exp\left[-(x/\sigma_2)^2\right] \tag{2}$$

where constants A_2 and σ_2 control the maximum/minimum value and the curve width at half-amplitude.

Note that the shape of the Fermi/Gauss function from their minimum value to the left/right is similar to the shape of the transmittance curves of nitrite, shown in Figure 1. Since the transmittance curve is not symmetric with respect to the wavelength of maximum absorption (minimum of the transmittance curve), we can model the first part of the transmittance curve, until the minimum, as a Fermi function. The second part of the curve, from the minimum, can be modeled as a Gauss function. Thus, the nitrates/nitrites transmittance curve can be modeled using the following expression:

$$T(\lambda) = \begin{cases} 100 - \left(1 - \frac{100 - T_{\min}}{100\sqrt{\sigma_1}\exp(-0.5)}(\lambda_{\min} + \sigma_1/2 - \lambda)\exp\left[-\left(\frac{\lambda_{\min} + \sigma_1/2 - \lambda}{\sigma_1}\right)\right]\right) & \text{if } \lambda \le \lambda_{\min} \\ 100 - \left(1 - \frac{100 - T_{\min}}{100}\exp\left[-\left(\frac{\lambda - \lambda_{\min}}{\sigma_1}\right)^2\right]\right) & \text{if } \lambda > \lambda_{\min} \end{cases} \quad (3)$$

Using this model, we can reconstruct the shape of the curve given in Figure 1: Figure 2 shows this reconstruction, depicting the complete Fermi and Gauss functions (the model only considers these functions to or from $\lambda_{\min}$, see Equation (3)). We have measured different series of nitrate/nitrite concentrations in pure water and then modeled them with our method. By doing this process we have obtained the patterns shown in Figure 3(a) for nitrites and Figure 3(b) for nitrates. We can use these patterns of nitrate/nitrite in pure water for comparison purposes in measured water samples, as we will show later.

The complete model for a real case, in which the water sample contains nitrate/nitrite and other contaminants such as organic matter, or metallic ions, is the following.

We consider that the transmittance of a real sample is formed by the product of different components. One of them, organic matter, can be modeled as a second-degree polynomial (Thomas *et al.* 1990*a*), whereas the rest of the contaminants will produce minima of transmittance at different wavelengths, which can be modeled using the Fermi–Gauss approach shown before. Thus, our complete

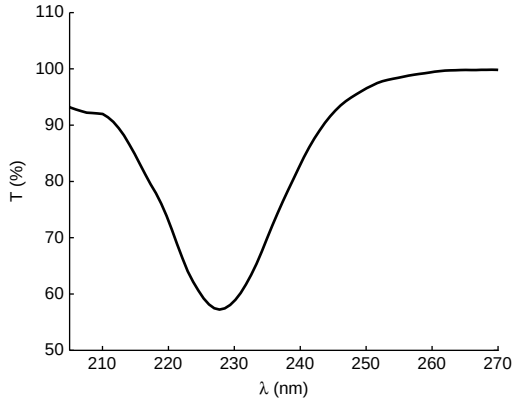

Figure 1 | Detail of a measured transmittance curve of pure water with 100 ppm of nitrite.

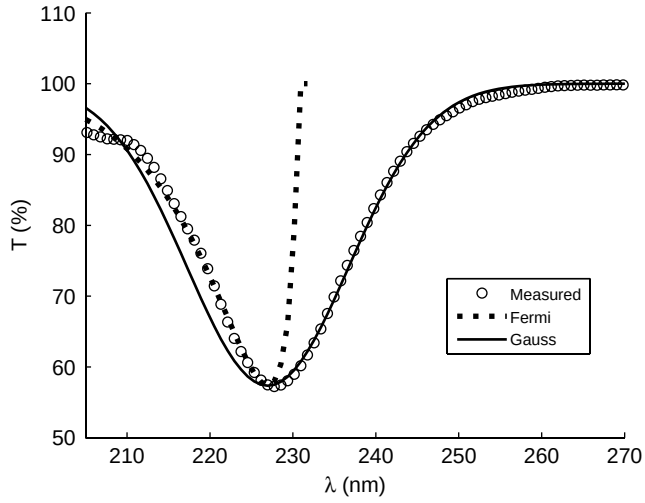

Figure 2 | Optimized transmittance curve model of the transmittance using the proposed method for the measure shown in Figure 1.

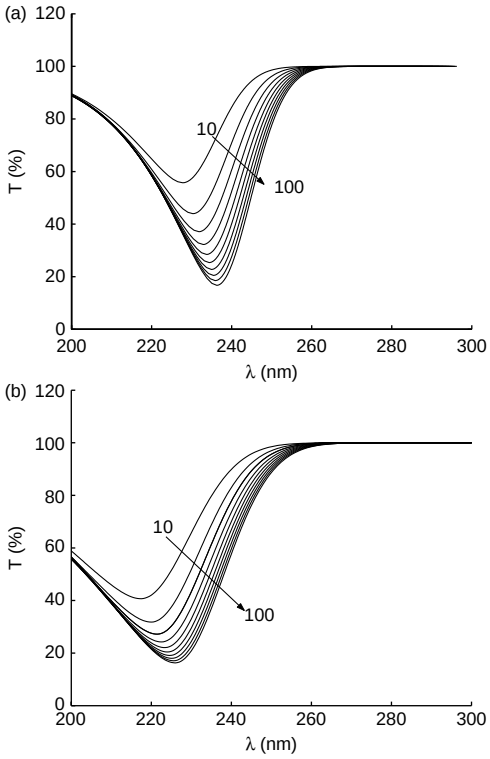

Figure 3 | Patterns of transmittance curves obtained with the proposed model using different measures of nitrite/nitrate concentrations in pure water: (a) nitrite patterns; (b) nitrate patterns.

model for the transmittance curve of a real water sample is:

$$\varphi(\lambda) = (a + b\lambda + c\lambda^2)\prod_{i=1}^{N_c} T(\lambda) \quad (4)$$

where N_c is the total number of contaminants to be estimated. Note that $\phi(\lambda)$ also depends on a number of parameters (a, b, c, λ_{min1}, T_{min1}, σ_{11}, σ_{21}, λ_{min2}, T_{min2}, σ_{12}, σ_{22}, …, $\lambda_{min N_c}$, $T_{min N_c}$, σ_{1N_c}, σ_{2N_c}) (see Equation (4)), which must be optimized in such a way that the model curve should be as close as possible to the measured spectral response of the sample. Mathematically, this problem can be stated as follows.

Let ϕ^* be the measured spectral response of the water sample under study. We must obtain the optimal parameters (a, b, c, λ_{min1}, T_{min1}, σ_{11}, σ_{21}, λ_{min2}, T_{min2}, σ_{12}, σ_{22}, …, $\lambda_{min N_c}$, $T_{min N_c}$, σ_{1N_c}, σ_{2N_c}) in such a way that a given objective function is minimized:

$$\min(\varphi(\lambda) - \varphi^*(\lambda))^2 \quad (5)$$

Note that, in general, we do not know the exact number of contaminants in a water sample N_c, but these can be estimated by the number of minima in its spectral response at the wavelengths of interest. In the next section we propose to use an Evolutionary Programming algorithm in order to solve the problem of the optimal estimation of the model's parameters, stated above.

ESTIMATION OF THE MODEL'S PARAMETERS THROUGH EVOLUTIONARY COMPUTATION

This section introduces algorithms which will be used in our work. First, a brief introduction to the different evolutionary algorithms specific for continuous optimization problems is given. Then, the so-called Classical Evolutionary Programming (CEP) approach is described. Using this description, two more different algorithms of Evolutionary Programming are introduced, the Fast Evolutionary Programming (FEP) and the Improved Fast Evolutionary Programming (IFEP). The application of these algorithms to the curve-fitting problem using the Fermi–Gauss model is explained at the end of this section.

Evolutionary computation techniques in continuous optimization problems

Evolutionary algorithms (EAs) (Fogel *et al.* 1966; Holland 1975; Schwefel 1981; Goldberg 1989; Koza 1992; Bäck & Schwefel 1993; Bäck *et al.* 1993; Fogel 1994; Birru *et al.* 1999; Yao *et al.* 1999; Beyer & Schwefel 2002; Lee & Yao 2004) are robust problem-solving techniques based on natural evolution processes. They are population-based techniques which codify a set of possible solutions to a given problem and evolve it through the application of the so-called *evolutionary operators*. These algorithms have also been applied to a large number of different optimization problems, both continuous and discrete, in the field of hydroinformatics (Babovic & Abbott 1997*a*,*b*; Babovic & Keijzer 2000; Ostfeld & Preis 2005; Ostfeld & Salomons 2005; Savic 2005; Giustolisi & Savic 2006; Tufail & Ormsbee 2006; Solomatine & Ostfeld 2008). Mainly four streams of evolutionary computation algorithms can be identified (Bäck *et al.* 1993): Evolutionary Programming (EP), developed by Fogel *et al.* (1966) and Fogel (1994), Genetic algorithms (GAs), due to Holland (1975) and Goldberg (1989), Evolution Strategies (ES) developed by Schwefel (1981) and Beyer & Schwefel (2002) and Genetic Programming (GP), developed by Koza in the early 1990s (Koza 1992). Usually, the GA approach is applied to discrete optimization problems, GP has been used in different contexts, including systems design and classification problems (Koza 1992; Babovic & Keijzer 2000), and EP and ES have been mainly applied to continuous optimization problems, with good results. The EP and ES algorithms have several points in common, and a few differences. Maybe the most important difference is that the EP algorithm was conceived as a pure mutation-based technique, whereas in the ES there exists an operator of recombination or crossover. Some other differences exist (Bäck *et al.* 1993), such as the use of a kind of tournament selection in EP, or a different calculation of the standard deviation of each individual, based on exponential functions in the case of the ES and in the square root of a linear transformation of the fitness value in the case of the EP (see Bäck *et al.* (1993) for details).

In a more recent paper, Yao *et al.* introduced the so-called Classical Evolutionary Programming (Yao *et al.* 1999), which is an algorithm based both in EP and ES.

The CEP is an evolutionary algorithm based on Gaussian mutations (it does not include a recombination operator). The updating of the individuals is, however, carried out based on the ES model, and finally a tournament-type selection operator is included. The CEP approach can be modified by including a different type of mutation, based on a Cauchy probability distributed function, to obtain a different algorithm called Fast Evolutionary Programming (Yao *et al.* 1999). The possibility of including yet another type of mutation has also been explored in subsequent papers (Yao *et al.* 2004). In our paper, we explore the optimization ability of the CEP and FEP (and also of the combination of both types of mutation in the same algorithm called the Improved Fast Evolutionary algorithm IFEP) in the optimization of the Fermi–Gauss model from different objective (observed) data.

The CEP algorithm

The CEP algorithm is used to optimize a given function $f(\boldsymbol{x})$, i.e. obtaining $\boldsymbol{x}_o$ such that $f(\boldsymbol{x}_o) < f(\boldsymbol{x})$, with $\boldsymbol{x} \in$ [lim_inf, lim_sup]:

(i) Generate an initial population of μ individuals (solutions). Let t be a counter for the number of generations and set it to $t = 1$. Each individual is taken as a pair of real-valued vectors $(\boldsymbol{x}_i, \boldsymbol{\sigma}_i)$, $\forall i \in \{1, \ldots, \mu\}$, where $\boldsymbol{x}_i$ are objective variables and $\boldsymbol{\sigma}_i$ are standard deviations for Gaussian mutations.

(ii) Evaluate the fitness value for each individual $(\boldsymbol{x}_i, \boldsymbol{\sigma}_i)$ (using the problem's objective function).

(iii) Each parent $(\boldsymbol{x}_i, \boldsymbol{\sigma}_i)$, $\{i = 1, \ldots, \mu\}$ then creates a single offspring $(\boldsymbol{x}_i', \boldsymbol{\sigma}_i')$ as follows:

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \sigma_i' \boldsymbol{N}_1(0,1) \quad (6)$$

$$\sigma_i' = \sigma_i \exp(\tau' N(0,1) + \tau \boldsymbol{N}(0,1)) \quad (7)$$

where $N(0,1)$ denotes a normally distributed one-dimensional random number with mean 0 and standard deviation 1, and $\boldsymbol{N}(0,1)$ and $\boldsymbol{N}_1(0,1)$ are vectors containing random numbers of mean 0 and standard deviation 1, generated anew for each value of i. The parameters τ and τ' are commonly set to $(\sqrt{2\sqrt{n}})^{-1}$ and $(\sqrt{2n})^{-1}$, respectively (Yao *et al.* 1999), where n is the length of the individuals.

(iv) If $x_i(j) >$ lim_sup then $x_i(j) =$ lim_sup and if $x_i(j) <$ lim_inf then $x_i(j) =$ lim_inf.

(v) Calculate the fitness values associated with each offspring $(\boldsymbol{x}_i', \sigma_i'), \forall i \in \{1, \ldots, \mu\}$.

(vi) Conduct pairwise comparison over the union of parents and offspring: For each individual, p opponents are chosen uniformly at random from all the parents and offspring. For each comparison, if the individual's fitness is better than the opponent's, it receives a "win".

(vii) Select the μ individuals out of the union of parents and offspring that have the most "wins" to be parents of the next generation.

(viii) Stop if the halting criterion is satisfied, and if not, set $t = t + 1$ and go to Step (iii).

Table 1 | Performance comparison between the EP (CEP, FEP and IFEP) and SaNSDE algorithms in the different laboratory samples analyses. The values displayed are in terms of the objective function given by Equation (5). Values in bold stand for the best values ever obtained by any of the tested algorithms

Sample (ppm)	CEP	FEP	IFEP	SaNSDE (Yang *et al.* 2008)
Best value				
10	1.295	1.143	0.806	**0.781**
20	1.173	**0.922**	0.925	**0.922**
30	1.253	1.206	**1.117**	2.426
40	1.493	1.462	**1.347**	1.353
50	0.689	0.746	**0.624**	1.762
60	1.254	1.264	**1.172**	1.186
70	0.959	0.816	**0.751**	1.801
80	0.998	0.859	0.681	**0.676**
90	0.974	1.012	0.849	**0.836**
100	0.817	0.794	**0.688**	1.517
Mean value				
10	6.836	4.168	3.331	2.689
20	3.267	2.543	1.873	2.289
30	9.384	4.952	4.613	2.660
40	12.538	8.054	4.668	2.671
50	13.252	8.743	5.687	2.587
60	16.824	11.093	6.152	2.338
70	16.422	10.177	4.766	2.532
80	19.704	10.894	6.308	3.553
90	20.288	9.824	4.503	2.601
100	18.848	13.243	4.443	2.053

FEP and IFEP algorithms

A second version of the algorithm is the so-called Fast Evolutionary Programming (FEP). The FEP was described and compared with the CEP in Yao *et al.* (1999). The FEP is similar to the CEP algorithm, but it performs a mutation following a Cauchy probability density function, instead of a Gaussian-based mutation. The one-dimensional Cauchy density function centered at the origin is defined by

$$f_1(x) = \frac{1}{\pi}\frac{1}{t^2 + x^2} \tag{8}$$

where $t > 0$ is a scale parameter. See Yao *et al.* (1999) for further information about this topic. Using this probability density function, we obtain the FEP algorithm by substituting step (iii) of the CEP by the following equation:

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \sigma_i \boldsymbol{\delta} \tag{9}$$

where $\boldsymbol{\delta}$ is a Cauchy random variable vector with the scale parameter set to $t = 1$.

Finally, in Yao *et al.* (1999) the *Improved FEP* (IFEP) is also proposed, where the best result obtained between the Gaussian mutation and the Cauchy mutation is selected to complete the process. This is the global search technique used in this paper.

Figure 4 | Location of the three studied aquifers. All them are in Comunidad de Madrid, in the central zone of Spain.

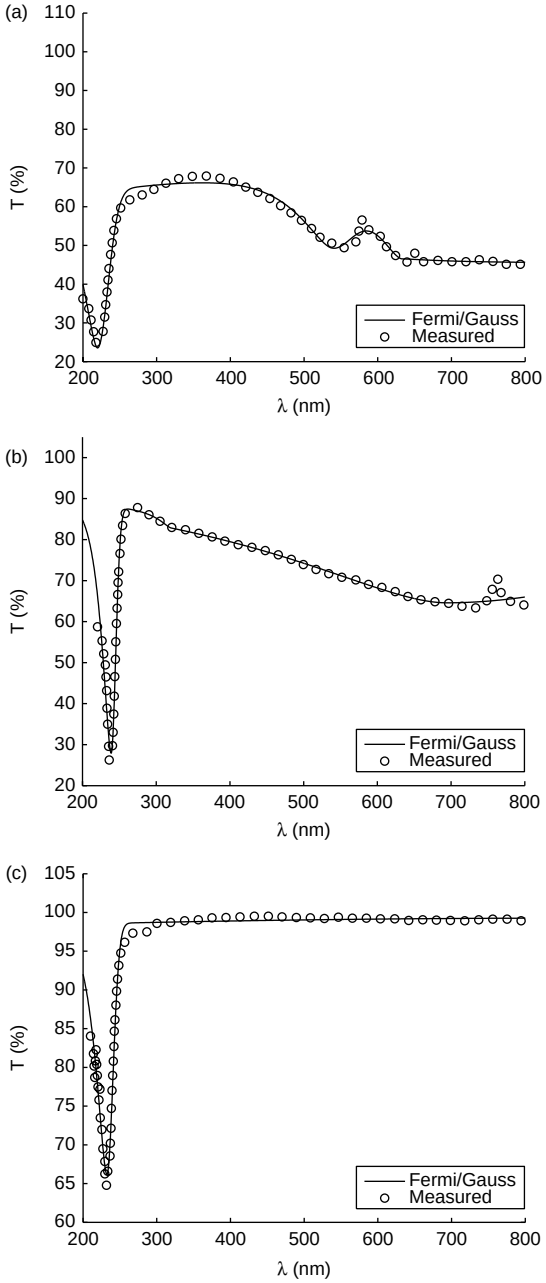

Figure 5 | Real transmittance curves measured in the considered aquifers. (a) Transmittance curve obtained in Redueña; (b) transmittance curve obtained in Loranca; (c) transmittance curve obtained in Fuenlabrada.

Application of EP algorithms in transmittance curves modeling

The EP algorithms shown above (CEP, FEP and IFEP) can be directly applied to estimate the optimal parameters of our transmittance model with very few changes: First, the

individual $\boldsymbol{x} = x_1, \ldots, x_N, \sigma_1, \ldots, \sigma_N$ is in this application $\boldsymbol{x} = (a, b, c, \lambda_{\min 1}, T_{\min 1}, \sigma_{11}, \sigma_{21}, \ldots, \lambda_{\min N_c}, T_{\min N_c}, \sigma_{1N_c}, \sigma_{2N_c}, \sigma_a, \sigma_b, \ldots, \sigma_{\sigma_2 N_c})$ (all the parameters to be minimized and their variance in the EP). The objective function of the problem is given by Equation (5), as described in the Fermi–Gauss model above.

EXPERIMENTS AND RESULTS

In this section we show the performance of the approaches described in this paper for curve fitting, using the Fermi–Gauss model of the transmittance curves. Two different types of experiments are carried out: first, experiments in laboratory samples, where a comparison with an alternative algorithm for curve fitting is shown. Second, we show the performance of our proposal in several samples from aquifers in Madrid, Spain in order to experiment with a real application of the method. In all the experiments carried out in this paper, an optic spectrophotometer Stellarnet, based on a diffraction grating, has been used. The obtained data have been treated using the software Spectrawiz 2.1. This spectrophotometer allows measures in the range of wavelengths 200–800 nm., with a precision of 0.5 nm. Before

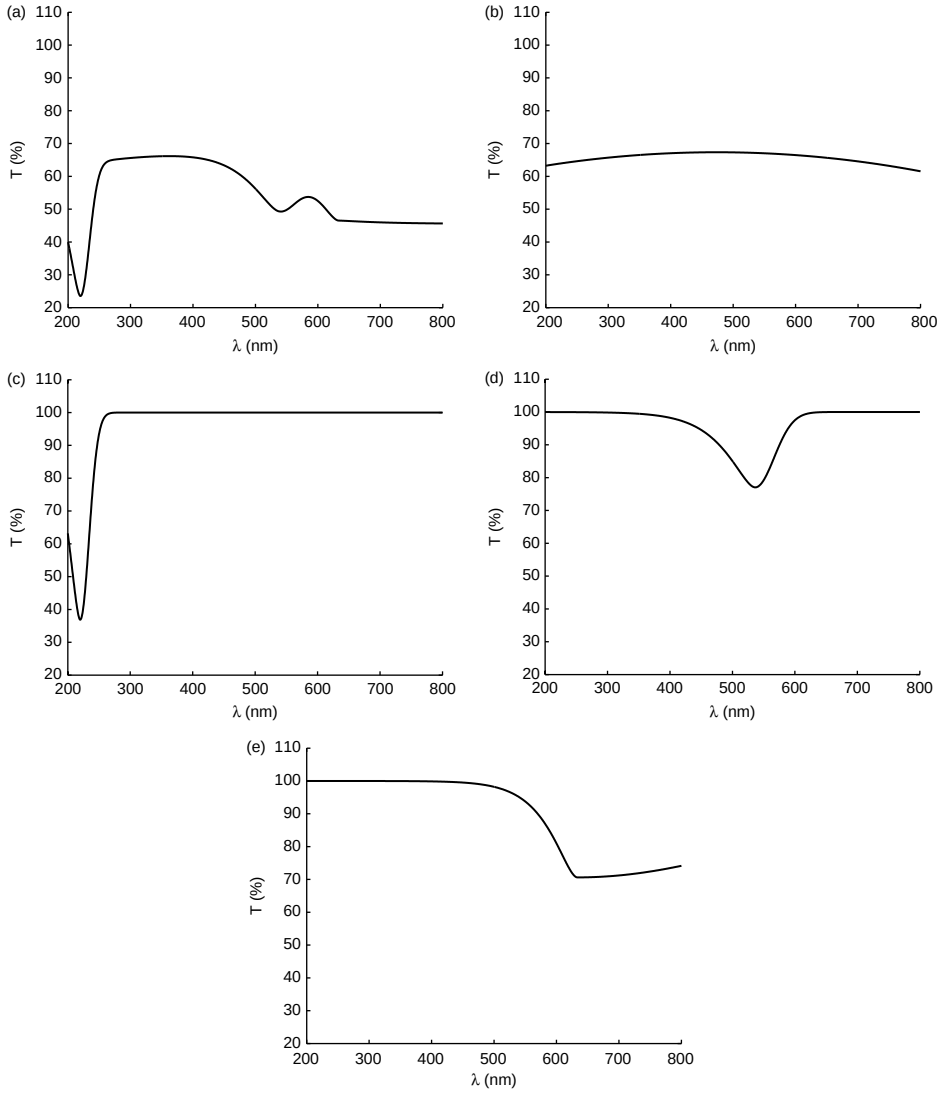

Figure 6 | Modeled curve and its components for the measured transmittance curve of Redueña.

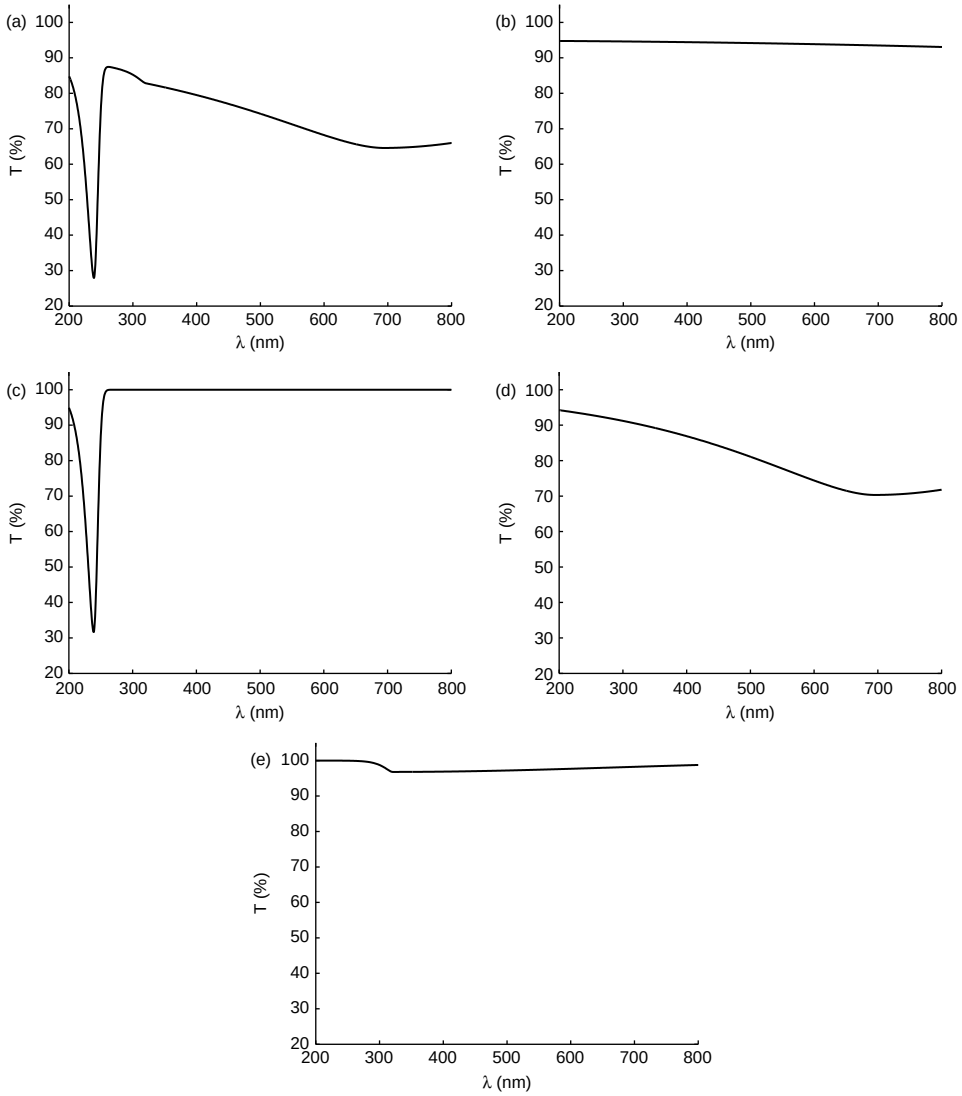

Figure 7 | Modeled curve and its components for the measured transmittance curve of Loranca.

performing the measures, the spectrophotometer must be calibrated using an additional pattern.

Experiments in laboratory samples

First, we test the performance of the proposed approach in the analysis of the transmittance curve of 10 water samples prepared in the laboratory. Note that each water sample consists of pure water with a given (known) concentration of nitrates, and we have 10 different concentrations (from 10 ppm to 100 ppm, in steps of 10 ppm). For each sample, we have measured more than 2,000 points of transmittance for different values of the wavelength (from 200–800 nm). Then, we adjust the proposed transmittance model to this measure, using the EP algorithm previously described. Since a simple contaminant is present in the sample, the transmittance curve can be modeled just by using seven parameters (three from organic matter and four from the Fermi–Gauss model). In order to compare the performance of the EP algorithms used in this paper, we have solved the curve-fitting problems of laboratory samples using a class of Differential Evolution algorithm, specifically the Self-adaptive Differential Evolution with Neighborhood Search (SaNSDE), which has provided good results in difficult

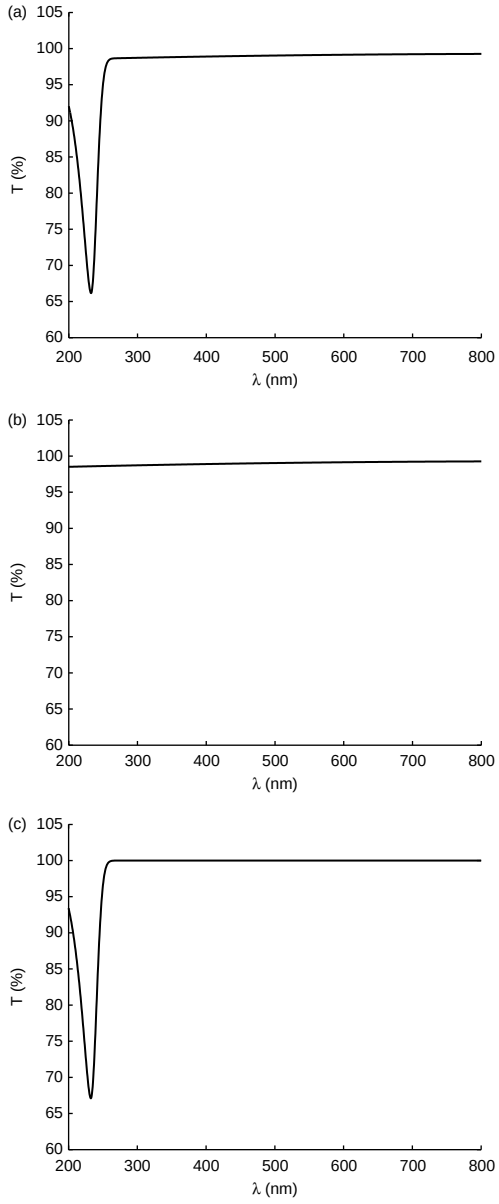

Figure 8 | Modeled curve and its components for the measured transmittance curve of Fuenlabrada.

continuous optimization problems. Details on this algorithm can be seen in Yang *et al.* (2008).

Table 1 shows the results obtained by the compared algorithms (CEP, FEP, IFEP and SaNSDE) in terms of the objective function given by Equation (5), in laboratory water samples with different nitrate concentrations, from 10 to 100 ppm. Note that all the tested algorithms are able to accurately estimate the nitrate concentration (ppm) in the laboratory samples considered, though there are small differences in the quadratic error obtained by each technique Equation (5). This could be important in real samples in which there are several contaminants per sample. Specifically, the results show that, in general, the IFEP and the SaNSDE algorithms obtain quite similar results, outperforming the CEP and FEP approaches in terms of error measure. Note, however, that the IFEP algorithm obtains the best values in samples 30, 40, 50, 60, 70 and 100 ppm, whereas in the 20 ppm sample the FEP and the SaNSDE obtain the best value, and in samples 10, 80 and 90 ppm the SaNSDE obtains values slightly better than the IFEP. These results show that the IFEP is a good approach to implement curve fitting with the Fermi–Gauss model to model transmittance curves. The next step is to show the application of this algorithm in real samples.

Experiments in real samples

In order to complete this experimental study, we test the IFEP approach in the analysis of three real spectroscopy transmittance measures obtained in three different zones of Madrid, Spain. Specifically, we have performed a measure in a wastewater land-application plant sited in Redueña, to the north of Madrid, a measure in a well in Fuenlabrada, to the south of Madrid, and another measure in another well in Loranca, also to the south of Madrid. Figure 4 shows the locations of the places where spectroscopy measures were performed. All the results obtained by the IFEP algorithm have been validated using standard techniques for water analysis (APHA, AWWA & WPCF 1998), specifically ionic chromatography, obtaining similar results.

Once the measured transmittance curves were measured, they were processed with our method, using an IFEP algorithm with a population of 100 individuals for 2000 generations. Figure 5(a)–(c) show the transmittance measures obtained in Redueña, Loranca and Fuenlabrada, respectively. In the case of Redueña and Loranca we have to consider three minima in the measured transmittance curves. This means that we have to optimize 15 parameters (three from the polynomial representing organic matter and four to model each transmittance minimum, following Equation (3). The case of the Fuenlabrada measure is different, since only one minimum is shown in the curve.

In this case we have to optimize seven parameters in order to model the curve.

Figure 6(a)–(e) show the modeled Fermi–Gauss for the Redueña transmittance measure (a) and the corresponding components, organic matter and turbidity influence (b), nitrates (c) and two extra contaminants with a maximum of absorption at 550 nm and 620 nm (d) and (e), respectively. Note that the component curve of nitrates (c) is the one in which we are interested, and it coincides with the nitrate pattern curve of 15 ppm (see Figure 3(b)), which is the concentration of nitrates in the Redueña sample.

A similar analysis can be done for the Loranca transmittance curve. Figure 7(a)–(e) show the modeled Fermi–Gauss for the Loranca transmittance measure (a), and the rest of the components in (b), (c), (d) and (e). We also locate a maximum of absorbance at 680 nm, corresponding to a different contaminant ((c)). Note that we have considered three maxima of absorbance for this sample, so the method locates a small one at 310 nm. This one, however, may be forced by order decision of running the algorithm with three maxima of absorbance instead of two. Note, however, that this decision does not affect the measurement of the nitrites/nitrates in the sample, which is the final objective of this work.

Finally, the analysis of the transmittance curve for a well in Fuenlabrada can be seen in Figure 8(a)–(c). In this case we only found nitrites and no other contaminant was located in this curve.

CONCLUSIONS

In this paper we have proposed a novel model for estimating the concentration of nitrates/nitrites in water. Our approach considers that transmittance curves of samples containing nitrates/nitrites can be modeled using a product of polynomial, Fermi and Gauss functions. We adjust the parameters of the model using an evolutionary programming algorithm in such a way that the differences between the model curve and the real (measured) curve should be a minimum. We have tested our approach in several laboratory samples, where we have compared the performance of different approaches to the problem, and in three real transmittance curves from samples obtained in different aquifers of Madrid. This application of our method has discovered moderate to high concentrations of nitrate/nitrites in these real cases, mainly in samples from a wastewater land-application plant. Our method can be extended to different types of contaminants and models of transmittance/absorbance curves. In future work, this paper opens the possibility to use other types of evolutionary-based approaches in the detection of nitrate concentration in water samples. In this sense, the possibility of using multi-objective optimization in order to process samples with several pollutants could be very interesting.

ACKNOWLEDGEMENTS

This work has been partially supported by Comunidad de Madrid, Universidad de Alcalá, through Project CCG07-UAH/TIC-1894. AMP-B is supported with a doctoral fellowship by the European Social Fund and Junta de Comunidades de Castilla la Mancha, in the framework of the Operating Programme ESF 2007–2013.

REFERENCES

Almasri, M. N. & Kaluarachchi, J. J. 2004 Assessment and management of long-term nitrate pollution of ground water in agriculture-dominated watersheds. *J. Hydrol.* **295**, 225–245.

Almasri, M. N. & Kaluarachchi, J. J. 2007 Modeling nitrate contamination of groundwater in agricultural watersheds. *J. Hydrol.* **343**, 211–229.

APHA, AWWA & WPCF 1998 *Standard Methods for the Examination of Water and Wastewater*, 20th edition. American Public Health Association Press, Washington, DC, USA.

Babovic, V. & Abbott, M. B. 1997 The evolution of equation from hydraulic data: part I: theory. *J. Hydraul. Res.* **35** (3), 397–410.

Babovic, V. & Abbott, M. B. 1997 The evolution of equation from hydra-ulic data: part II: applications. *J. Hydraul. Res.* **35** (3), 411–430.

Babovic, V. & Keijzer, M. 2000 Genetic programming as a model induction engine. *J. Hydroinf.* **2** (1), 35–60.

Bäck, T. & Schwefel, H. P. 1993 An overview of evolutionary algorithms for parameter optimization. *Evol. Comput.* **1** (1), 1–23.

Bäck, T., Rudolph, G. & Schwefel, H. P. 1993 Evolutionary programming and evolution strategies: similarities and differences. In *Proc. of the Second Annual Conference on Evolutionary Programming*, pp. 11–22. Evolutionary Programming Society, San Francisco, CA, USA.

Beemster, J. & Kahle, S. 1995 *On-line Analysis of Nitrate and Nitrite in Wastewater without the Use of Reagents through Multiple Wavelength Absorbance Spectrometry*. ChemScan internal report. Available at: http://www.chemscan.com/technical/asa-63.html

Beyer, H. G. & Schwefel, H. P. 2002 Evolution strategies–a comprehensive introduction. *Nat. Comput.* **1** (1), 3–52.

Birru, H., Chellapilla, K. & Rao, S. 1999 Local search operators in fast evolutionary programming. In *Proc. IEEE Conference on Evolutionary Computation*, pp. 1506–1513. IEEE Press, Piscataway, NJ, USA.

Fogel, D. B. 1994 An introduction to simulated evolutionary optimization. *IEEE Trans. Neural Networks* **5** (1), 3–14.

Fogel, L. J., Owens, A. J. & Walsh, M. J. 1966 *Artificial Intelligence Through Simulated Evolution*. John Wiley & Sons, New York.

Giustolisi, O. & Savic, D. A. 2006 A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinf.* **8** (3), 207–222.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.

Gross, A., Boyd, C. & Seo, J. 1999 Evaluation of the ultraviolet spectrophotometry method for the measurement of total nitrogen in water. *J. World Aquacult. Soc.* **30** (3), 388–393.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Michigan, USA.

Holm, T., Kelly, W., Sievers, L. & Webb, D. 1997 A comparison of ultraviolet spectrophotometry with other methods for the determination of nitrate in water. *Spectroscopy* **12** (9), 38–48.

Johnson, K. & Coletti, L. 2002 In-situ ultraviolet spectrophotometry for high resolution and long-term monitoring of nitrate, bromide and bisulfide in the ocean. *Deep-sea Res. Part I–Ocenanogr. Res.* **49** (7), 1291–1305.

Karlsson, M., Karlberg, B. & Olsson, R. 1995 Determination of nitrate in municipal waste-water by UV spectroscopy. *Anal. Chim. Acta* **312** (1), 107–113.

Koza, J. 1992 *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.

Lee, C. Y. & Yao, X. 2004 Evolutionary programming using mutations based on the Levy probability distribution. *IEEE Trans. Evol. Comput.* **8** (1), 1–13.

Ostfeld, A. & Preis, A. 2005 A data driven model for flow and contaminants runoff predictions in watersheds. In *River Basin Restoration and Management* (ed. A. Ostfeld & J. M. Tyson), *Water and Environmental Management Series*, pp. 62–70. IWA Publishing, London.

Ostfeld, A. & Salomons, S. 2005 A hybrid genetic-instance based learning algorithm for CE-QUAL-W2 calibration. *J. Hydrol.* **310**, 122–142.

Savic, D. 2005 Evolutionary computing in hydro-geological systems. In *Encyclopedia of Hydrological Sciences* (ed. M. G. Andersen), Vol. 1, pp. 1–20. John Wiley & Sons, New York.

Schwefel, H. P. 1981 *Numerical Optimization of Computer Models*. John Wiley & Sons, New York.

Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *J. Hydroinf.* **10** (1), 3–22.

Thomas, O., Gallot, S. & Mazas, N. 1990*a* Ultraviolet multiwavelength absorbtiometry (Uvma) for the examination of natural-waters and wastewaters I: general considerations. *Fresenius J. Anal. Chem.* **338** (3), 234–237.

Thomas, O., Gallot, S. & Mazas, N. 1990*b* Ultraviolet multiwavelength absorbtiometry (Uvma) for the examination of natural-waters and wastewaters II: determination of nitrate. *Fresenius J. Anal. Chem.* **338** (3), 238–240.

Tufail, M. & Ormsbee, L. E. 2006 A fixed functional set genetic algorithm (FFSGA) approach for function approximation. *J. Hydroinf.* **8** (3), 193–206.

Wolfe, A. & Patz, J. A. 2002 Reactive nitrogen and human health: acute and long-term implications. *Ambio* **31** (2), 120–125.

Yang, Z., Tang, K. & Yao, X. 2008 Self-adaptive differential evolution with neighborhood search. In *Proc. of the IEEE Congress on Evolutionary Computation*, pp. 1110–1116. IEEE Press, Piscataway, NJ, USA.

Yao, X., Liu, Y. & Lin, G. 1999 Evolutionary programming made faster. *IEEE Trans. Evol. Comput.* **3** (2), 82–102.

First received 29 February 2008; accepted in revised form 22 June 2009. Available online 6 January 2010