Data assimilation for the management of environmental and radiological data in case of nuclear accident

N. Pérot

Institute of Protection and Nuclear Safety (IPSN), Division of Environmental Protection (DPRE), CE Cadarache, 13108 Saint-Paul-lez-Durance cedex, France

Abstract. The "Data Assimilation for the Management of Environmental and Radiological Data in Case of Nuclear Accidents" project managed by the IPSN/DPRE started in 2001 for at least three years. It is based on the IPSN skills in the field of data analysis and the modeling of radionuclide transfers in the environment, more precisely along the food chain pathways. The main part of the project consists in defining methods to fit radioecological models with incoming measured values when predictions do not correspond to such measured values. In such a case, the radioecological parameters of the model or some poorly assumed initial conditions are supposed to be responsible and have to be modified. In order to update the modeling, the "parameters" of the model have to be adjusted considering the measured values. This problem of data assimilation can be represented by a nonlinear program (or optimization problem), where the relations between the parameters and their definition domains constitute the set of constraints; then, the cost function to minimize is the difference between predictions and measured values. The experts can control the process of optimization by adding some well chosen constraints expressing their understanding of the phenomena considered in the studied case.

Problems of nonlinear programs are known to be very difficult to solve and the calculation time is rarely reasonable with deterministic methods, which scan all the research space. In this context, it is proposed to treat this problem by using a combination of a deterministic method such as Constraints Satisfaction Problems (CSP) and a stochastic method with a genetic algorithm to reach best performances.

1 INTRODUCTION

The *Data Assimilation for the Management of Environmental and Radiological Data in Case of Nuclear Accidents* project managed by the IPSN/DPRE has started this year and will continue for three years. It is based on IPSN competences in the field of Data Analysis and modeling transfers in the environment, more precisely the food chains. We have established collaborations with the NRPB (National Radiological Protection Board of United Kingdom) who will bring their competences in Data Analysis and will also provide values measured after the Windscale accident (1957 in UK).

The topic of this paper is to present what Data Assimilation is in the context of a nuclear crisis and how to consider its related constraints (processing speed, global and general information available on radioecological parameters) and what is our approach to cope with them.

The paper begins by a description of the problem of Data Assimilation as an optimization problem and after a brief summary on the techniques of CSP (Constraint Satisfaction Problems) and Genetic Algorithms, it presents our method based on the hybridization of these methods and its application on a case study.

2 DATA ASSIMILATION

2.1 Description of the Problem

Data assimilation [1] is often used in meteorology and in all the fields with a flux of observations requiring the adjustment of a model, which presents some variable elements. These elements can be initial conditions, which are not always known initially, transfer factors, from the experiments or measurements made in the environment, which are linked to validity conditions rarely verified in the case studied.

The problem is thus as follows : some observations in certain points of the study space considered (most often 2 dimensions : a geographical surface) are available. Moreover, a model has been established and allows the calculation of predicted values everywhere in this space. The data assimilation consists in fitting the results of a model with the observations by adjusting the parameters and the initial conditions of the model.

However, the adjustment of a model must not be confused with the assimilation of data. In the first case, the model describes phenomena for which the characteristics, such as parameters, initial conditions, boundary conditions, are well defined beforehand, and which are adjusted according to some examples. In the second case, the model used describes phenomena for which the parameters, initial conditions and/or boundary conditions, depend each time on the case considered.

There exist two large classes of data assimilation methods : the variational methods based on the theories of optimal control and the sequential statistical methods based on the optimal statistical estimation.

The most used method from the first class of methods is the adjoint method [2], and the most used method from the class of statistical methods is the Kalman filter [3] rather used in relatively linear cases and using the error covariance matrices of the observations and the predicted values of the model.

All the methods of optimal assimilation (variational, statistical) collide difficulties of implementation mainly of computational type (space memory, calculation cost).

2.2   Data Assimilation and Optimization

Variational data assimilation is an optimization problem consisting in minimizing the difference between observations and predicted values produced by a model. This problem can be expressed as follow :

$$\underset{D}{Min} \left\| C.\phi(V) - \tilde{\phi}_D \right\|$$

with $\tilde{\phi}_D$ the observations on the domain D, ϕ the model, V the variable elements of ϕ and C a linear operator linking the space of calculations to the space of observations (spatial interpolation, projection).

In the simple case of an analytical one dimensional model, the following formulation is obtained :

$$(S_1) \begin{cases} P_j \in I_j \ for \ j = 1,\ldots,m \\ \underset{P_1,P_2,\cdots,P_m}{Min} \sum_{i=1}^{n} \left| \phi(X_i,t_{obs},P_1,\cdots,P_m) - \tilde{\phi}(X_i) \right| \end{cases}$$

with P_j the parameters to be fitted, I_j the interval or validity domain of the parameter P_j, and t_{obs} the time at which the observations were made.

2.3   Data Assimilation in the context of nuclear crisis

The constraints related to a situation of nuclear crisis are a lack of information on some environmental or radioecological parameters and a lack of time for the calculations ; therefore the models used for crisis management are often analytical models using a moderate number of parameters and sometimes aggregated parameters. Consequently, adjusting these parameters will consist in solving a data assimilation problem like (S_1).

3   DATA ASSIMILATION AND GENETIC ALGORITHM

In most cases, the problem (S_1) to be solved is an optimization problem under constraints. Depending on the form of the constraints, a linear or non-linear program will have to be solved. In both cases, the problem is NP-Complete, in other words, there is no algorithm allowing a solution to be obtained with calculation time which is a polynomial function of the number of parameters. Nevertheless, there are several non-deterministic methods, such as genetic algorithms, which can calculate an optimal solution within a reasonable time, often the best solution.

In this article, a method based on the hybridization of a CSP (Constraint Satisfaction Problem) with a genetic algorithm is presented.

The techniques of constraints satisfaction are well adapted to the very constraining problems in which a thorough exploration of the research space is considered in order to provide acceptable solutions. However, it is difficult to apply them the most common cases, those with less constraints and/or a vast research space.

Moreover, the use of a genetic algorithm [4] is adapted to the rapid and global exploration of a large research space and is capable of providing several solutions. But it is not always easy to isolate the admissible solutions, namely those satisfying all the constraints.

Non-linear programming problems are complex problems and it seemed interesting to take advantage of the two methods by hybridizing them.

Here we consider a CSP [5] defined by a triplet (X,D,C) where X are a set of n variables, D their respective domains $(D(X_1),...,D(X_n))$ and C a set of relations between these variables. In the case of an optimization problem, one must add the function F_c to optimize (minimization or maximization) on the variables of X.

3.1   Constraint Satisfaction Problem (CSP)

Our CSP is expressed in a logical program with constraints and the ECLiPSe system [6] is used, as it implements all the classical, linear and other constraints. By doing this, the constraints can be checked to be satisfiable on a domain and an admissible solution in this domain can be calculated, which satisfy all the constraints.

3.2   Genetic algorithm

For a given optimization problem, an individual or a chromosome represents a point in the research space. The value of the criterion to be optimized is then associated to it, its *adaptation*. Then, iteratively, populations of individuals are generated, on which *selection, crossing* and *mutation* are applied.

In our case, a chromosome or individual is defined as being a set of genes with as many genes as variables. Each gene G_i, defines a sub-set of the domain of the X_i variable, and each chromosome has an adaptation. This adaptation is a valuation, which is obtained in two steps :

➢ Determination of an admissible solution S_a, of the CSP reduced to the sub-domains represented by the chromosome.

➢ If there is no admissible solution then the adaptation is nil otherwise the *adaptation* is equal to $F_c(S_a)$.

First, an initial population is randomly generated and to go from one generation to the next, *selection, crossing* and *mutation* operations are applied.

Figure 1 represents the *selection* operation on the initial population, P_s. It is obtained through the selection of the $N/2$ best adapted individuals among the N chromosomes of the global population to which are added the individuals from a sampling with the probability $_s$ in the other remaining half population. The next generation is made up of P_c coming from the *crossing* operation and from non-selected individuals. The *crossing* operation on two chromosomes consists in randomly sampling a gene number from which the genes of the two chromosomes to be crossed will be exchanged, thus giving two new chromosomes.
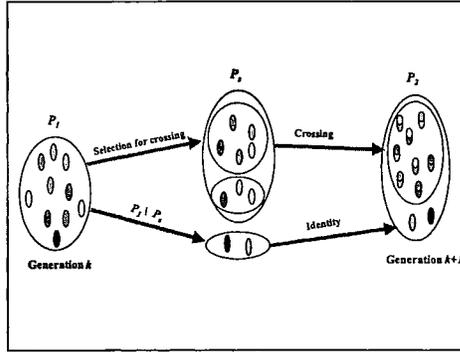
Figure 1 : Process to produce a new generation

Figure 2 presents our *mutation* operation, which consists in applying a mutation to the individuals of the population of the current generation with a probability $_m$. This mutation consists, for a chromosome, in randomly selecting a gene and replacing its domain with another.
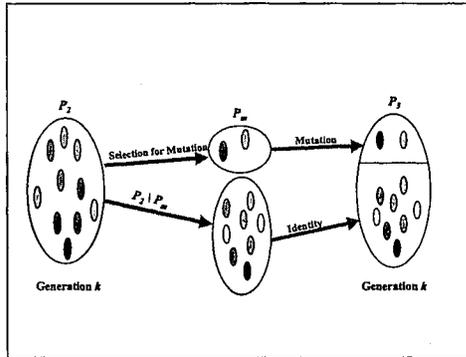

Figure 2 : Operation of mutation on a population

4   GENERAL ALGORITHM AND RESULTS

It is to be recalled that a non-linear programming problem is represented by a $CSP(X,D,C)$ and a cost function F_c on the variables of X to be optimized (maximization or minimization). The term « non-linear » is used when the CSP constraints and/or the cost function are non-linear.

Below, are presented the details of the application of this hybrid algorithm to part of a model developed by one of our teams. Figures 4 and 5. illustrate this case study.

4.1   Algorithm

The algorithm below shows the introduction of the constraint satisfaction techniques in the different stages of the genetic algorithm.

The system of constraint management ECLiPSe is used to evaluate the *adaptation* of each individual or chromosome whereas the genetic algorithm allows a global exploration of the whole research space.

```
Construction of Initial Population :
    N Chromosomes  C = (G₁,...,Gₙ)
    Gᵢ ⊂ D(Xᵢ)
For Ngeneration=1 to Ngeneration_max :
    Selection
    Crossing
            Valuation
    Mutation
            Valuation
```

```
Valuation(C) :
    Calculation of a solution Sₐ on
    G₁ × G₂ × ... × Gₙ : ECLiPSe System
    If fail then Adaptation(C)=0
    Else Adaptation(C)= Fc(Sₐ)
```

At each *crossing* or *mutation* operation, the valuation consists in calculating the adaptation of the new chromosomes.

4.2 Description of the case study

ASTRAL [7] is a model developed at the IPSN whose aim is to estimate the impact of a nuclear accident on the environment and on the men bound to this environment. Starting with the value of the radionuclide deposits on soil and the plants, the calculations of these radionuclides concentrations in agricultural production and the level of internal and external irradiation to which man is exposed (dosimetric impact) are made.
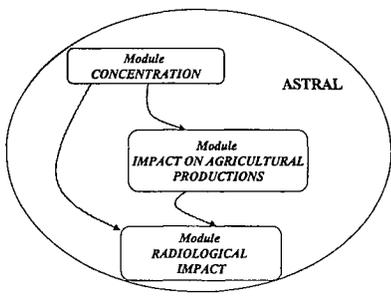

Figure 3 : The connection between the different modules of ASTRAL

ASTRAL includes three modules which are connected to each other and whose links are presented in Figure 3 :

1. The concentration module : calculation of the concentrations in all plant and animal production and in products derived from the latter.
2. The module of criterion of impact on productions : comparison with standards or intervention levels in order to rapidly assess the importance of the impact on the agricultural production from 1 to 3 years.
3. The module of radiological impact : calculation of the dose to man through ingestion and inhalation of particles in suspension (internal dose) and through external irradiation due to the deposits (external dose).

In order to test our method, the focus is on part of this model, namely the equation concerning the transfer of radionuclides in plants and in particular in market garden produce :

$$C_r(t, K, FTds, FTdh, Dils, Dilh) = D_r.(K.FTds.Dils + (1-K).FTdh.Dilh).e^{-(\lambda_{bw}+\lambda_r)t}$$

with C_r the concentration of the radionuclide r in the plant (in Becquerels per kilogram $Bq.kg^{-1}$), t the time in days from the day of the deposit, D_r the total deposit of the radionuclide r on the soil and the plants (in Becquerels per square meter $Bq.m^{-2}$), K the proportion of the dry deposit in the total deposit, λ_r the constant of radioactive decrease, λ_{bw} accounts for the weathering phenomena of leaves and the biological growth of plants. The parameters selected to be fitted considering the observations are the transfer factors for the dry deposit and wet deposit, $FTds$, $FTdh$ and the parameters allowing the integration of the protection of the greenhouses $Dils$, $Dilh$.

The scenario studied here concerns an accident having occurred on the 30th of April. The deposit was made on the 1st of May. The plants studied are vegetables and more exactly leaf vegetables and the radionuclide selected is ^{137}Cs.

Simulation using the parameters obtained from the experiments provided the following concentrations ($C(D_i)$), for different values of the deposit (D_i) corresponding to different localizations :

$t = 7$ days , $r = {}^{137}$Cs, v = leafy vegetables	
Deposit (Bq/m²)	Concentration (Bq/kg)
10^3	1365
$1.2\ 10^3$	1640
$5\ 10^2$	1500
$2.5\ 10^2$	700
$5\ 10^1$	150
$2.5\ 10^1$	40

4.3   Results

The scenario described in the previous section and a knowledge of the orders of magnitude allow the following mathematical problem to be set :

$$(P) \begin{cases} K \in [0;1] \\ Dils \in [0;1] \\ Dilh \in [0;1] \\ FTds \in [0;100] \\ FTdh \in [0;100] \\ FTds \geq FTdh \\ Min \sum_{i=1}^{6} |C_{Astral}(D_i, K, Dils, Dilh, FTds, FTdh) - C_s(D_i)| \end{cases}$$

The algorithm presented in section 4.1 is applied with crossing and mutation probabilities of 0.6 and 0.2 respectively over a population of 20 individuals. The calculation made on 20 generations, produces the results presented in Figures 4 and 5. The first curve shows the evolution of the cost function and the second curve presents the evolution of the two main parameters $FTds$ and $FTdh$. In this example, the stabilization of the solution is rapid with very short calculation times, less than one second.
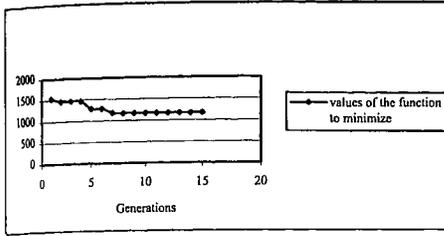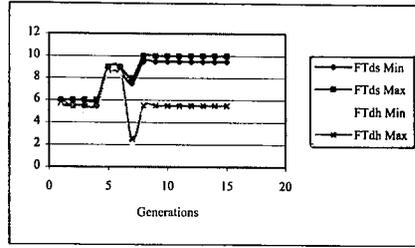
Figure 4 : Evolution of the function to minimize

Figure 5 : Evolution of the parameters FTds and FTdh

5 CONCLUSIONS AND PERSPECTIVES

As explained at the beginning of this paper, the context of a nuclear crisis is characterized by the need of processing speed, the lack of accurate information on environmental and radioecological parameters and the use of explicit models.

These considerations have led us to divert the classical methods of data assimilation and to turn ourselves to other types of methods like the CSP and the genetic algorithms and to combine them to get a good performance.

The first case study presented here in, although limited, is encouraging and other tests are foreseen to confirm the interest of such an approach.

References

[1] Pérot N., *Analyse de données et Assimilation de données (Data Analysis and Data Assimilation)*, IPSN Report, IPSN/DPRE/SERLAB/2000/07.

[2] Penenko V. V. and N. N. Obraztsov, *A variational initialisation method for the fields of the meteorological element*, Soviet Meteorol. Hydrol., 11, 1-11, 1976.

[3] Ghil M. and P. Manalotte-Rizzoli, *Data assimilation in meteorology and oceanography*, Advances in Geophysics, 23, 141-265, 1991.

[4] Goldberg D., *Genetic Algorithms*, Addison Wesley, 1989.

[5] Van Hentenryck P., *Constraint Satisfaction in Logic Programming*, MIT Press, Cambridge, MA, 1989.

[6] ECLiPSe User Manual (ECRC Common Logic Programming System), 1999.

[7] Mourlon Ch., *Description des fonctions et des paramètres du module de calcul ASTRAL Version 1.1*, IPSN Report, IPSN/DPRE/SERE/98/033