

A Global Optimization Approach to Assimilate the Initial Conditions into a SVAT Model

W. Bouarifi and N. Alaa

Laboratoire LAMAI, FST Marrakech, Cadi Ayyad University, Morocco
e-mail:walid.bouarifi@gmail.com
e-mail:alaa@fstg-marrakech.ac.ma

Abstract

Data assimilation is a novel, versatile methodology for estimating hydrologic variables; it is a statistical technique to combine measurements of variables such as the atmospheric temperature, water content or precipitation, with models that describe the time evolution of these variables. The melding of data and dynamics is a powerful methodology, which makes possible efficient, accurate and realistic estimations which might not otherwise be feasible. It is providing rapid advances in important aspects of both basic sciences and applied technologies and operations. Our approach is based in variational data assimilation approach based on optimal control. All control theory or variational assimilation approaches perform a global time-space adjustment of the model solution to all observations and thus solve a smoothing problem. The optimisation problem is formulated in the framework of control optimal theory, followed by a brief discussion of genetic algorithms used this study as a global optimization algorithm.

Keywords: *Data assimilation; Optimal control; Inverse modeling; Hydrology; Genetic Algorithms.*

1 Introduction

Within the arid and semi-arid regions, water availability is a major limitation for crop production. The Haouz plain that surrounds the city of Marrakech (Center of Morocco) is classified among regions with scarce water resources.

In this region, wheat yield depends on the water supply. The main objective of this study is to develop a mathematical methodologies allowing the integration of ground data, simulation models, and remote sensing data to evaluate the hydro-ecological resources of semi-arid regions and to understand and predict changes in water resources of these heterogeneous regions in terms of sustainable management.

Environment and resources management requires characterizing vegetation status. Vegetation canopy functioning can be studied combining both Soil-Vegetation- Atmosphere Transfer (SVAT) models and remote sensing data. These models describe energy and mass transfers in the soil-plant-atmosphere continuum. Remote sensing provides useful information for driving such models. Global circulation models are now routinely used in weather and climate predictions, but they usually contain only inadequate representations of the physical processes at the land-atmosphere interface. A better understanding of soil moisture and temperature dynamics will therefore help us with the assessment and prediction of global change and improve our ability to produce reliable short-term weather forecasts [10].

The advent of remote sensing data has now made it possible to study land-atmosphere processes on large spatial scales. Ideally, the satellite data are used in conjunction with the existing land-surface models to extract the valuable information contained in both the data and the models. Such optimal merging of data and models is generally termed data assimilation [9]. Therefore, an efficient data assimilation method is needed to combine information from observations and models in order to get the best possible estimate of the state of that system. In a variational assimilation scheme, the estimates are determined by minimizing a measure of fit between the land-surface states and both prior information and new data (cost function). The measure of fit is formulated using weights that depend on the corresponding uncertainties. The main objective of this work was to assimilate the surface temperature into a SVAT model in order to improve estimates of surface convective fluxes for evaluating evapotranspiration (latent heat flux) to manage irrigation water. A better understanding of the water balance is essential for exploring water-saving techniques. One of the most important concepts regarding water balance in semi-arid areas is crop evapotranspiration (ET) which is a key factor for determining proper irrigation scheduling and for improving water use efficiency in irrigated agriculture. Accurate estimation of evapotranspiration constitutes a very important part of irrigation system planning and designing, and accurate spatial determination is crucial to achieving sustainable agriculture. A better agreement was observed between measured and simulated fluxes.

This paper is organized as follows. In section 2, we provide a brief description of study sites and data collected during 2002-2003 season. In section 3,

we provide a theoretical background of the ICARE Model used in this study. Section 4 presents the approach based on variational data assimilation used to assimilate the initial condition in ICARE Model. Section 5 describes the global algorithm developed for the cost function minimization. Section 6 presents the results obtained by implementation of our algorithm, and concludes by discussing the advantage and the limitation associated with each method.

2 Experimental data

The experimental site is located in the Haouz plain semi-arid region in the center of Morocco, 40 km East of Marrakech city. It is an irrigated area managed by the ORMVAH (Office Régional de Mise en Valeur Agricole du Haouz) since the year 1999. The area covers 2800 ha and is almost flat. The dominant crop is wheat. The climate in the area is typically Mediterranean semiarid, with around 250mm of average annual rainfall, concentrated mainly from autumn to spring. The soil is very homogeneous; it is poor in organic matter (<2%), with a fine texture (clay to loamy). In this irrigated area, ORMVAH manages the distribution of water starting from December through May. The frequency and the amount of water for each irrigation are predetermined depending on the dam water level at the beginning of the cropping season without any consideration of the actual soil moisture status. Additionally flood irrigation is the most widely used method in this district [10].

2.1 Field experiments

Two wheat parcels (denoted R3-B123 and R3-B130) were fully equipped with instrument measurements described in (see [2] for full details). All fields were cropped with durum variety wheat, which has relatively short life cycle. This variety is suitable for semi-arid conditions and commonly used in the Marrakech-Haouz plain (ORMVAH technical documentation, Karrou, 2003). The studied fields were irrigated by using concrete canals that carry water from the main canal to the irrigated units.

2.2 Data description

From December 2002 to May 2003, the experiment was carried out on wheat crops to monitor the variables of the surface energy and water balance as well as soil and vegetation data during the entire growing cycle. More detailed informations about instrumentation are given in [2].

3 ICARE Model

The Interface-Canopy-Radiation-Exchange scheme ICARE rely on the detailed description of the canopy energy balance of the ecosystem under consideration, it's designed to describe the basic evaporation processes at the surface together with the water partitioning between vegetation transpiration, surface runoff and soil moisture change. Soil vegetation atmosphere transfer model ICARE is used to simulate energy, moisture, and fluxes, and include biosphere atmosphere transfer scheme. A functional requirement of these models is that they maintain estimates of the stored water field, i.e., estimates of the spatial distribution of water stored in soil, snow, and vegetation that is available to the atmosphere through evaporation and transpiration. ICARE uses the common dual-source formulation by Shuttleworth and Wallace [14] and its network of resistances. The state variables of ICARE Model are the temperature T_2 and T_s , and the moistures content θ_s and θ_2 .

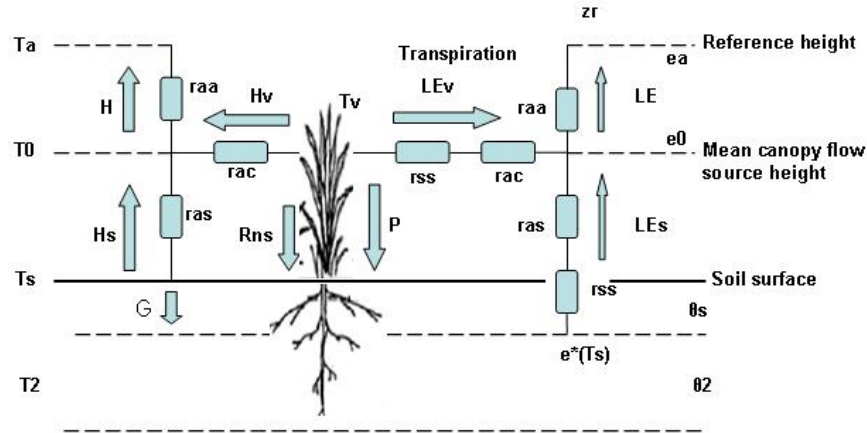


Figure 1: ICARE Schematization

The ICARE model is described in [2]. All the aerodynamic resistances are expressed as given by Choudry and Monteith [8], after computation of the stability of the aerodynamic effect. The aerodynamic resistance r_a (for heat and water vapor) is calculated as in Brutsaert [15]. The stomatal resistance describes the closure of the plant's stomata due to environmental impacts [11]. The most important factors determining the transpiration under unstressed conditions (no water limitation) are shortwave (or photosynthetically active) radiation, temperature, and vapor deficit, although some authors believe that the dependence on vapor deficit is an artificial effect. The prognostic equations of the state variables (water content and temperature of the tanks surface and deep) are following form

$$\left\{ \begin{array}{l} \frac{\partial \theta_s}{\partial t} = \frac{C_1}{\rho_w \cdot d_1} (P - E_s) - \frac{C_2}{\tau} (\theta_s - \theta_{eq}) \quad 0 \leq \theta_s \leq \theta_{eq} \\ \frac{\partial \theta_2}{\partial t} = \frac{1}{\rho_w \cdot d_2} (P - E_s - E_{tr}) - \frac{C_3}{d_2 \cdot \tau} \max[0, (\theta_s - \theta_{fc})] \quad 0 \leq \theta_2 \leq \theta_{sat} \\ \frac{\partial T_s}{\partial t} = C_G \times G - \frac{2\pi}{\tau} (T_s - T_2) \\ \frac{\partial T_2}{\partial t} = \frac{G}{(356C(\theta)\lambda_2\tau)^{\frac{1}{2}}} \end{array} \right. \quad (1)$$

4 Variational data assimilation

The goal of data assimilation is to reconstruct the model state with the help of different sources, such as that merge data and model in data assimilation follow mainly two approaches: control theory and estimation theory. The latter is the sequential approach and is based on Kalman filter and its variations, that only consider observation made in the past until the time of analysis, which is the case of real-time assimilation systems; the former relies on variational methods that take advantage of the full power of optimal control techniques where observation from the future can be used, for instance in a reanalysis exercise. The development of data assimilation in hydrology is a consistent inter-evolution of both data and model from coarseness to refinement. In the beginning, poor models and few data made it hard to insert data into model by "nudging" methods. Gradually plentiful data sets and truly fine models empower the efficient data assimilation schemes.

An assimilation algorithm is developed in this work, the approach is a variational algorithm based on the minimization of an objective function. This function consists of the weighted sum of squared differences between the observed and the estimated temperature of the soil. The objective function is minimized with an evolutionary algorithm method. This optimal scheme takes the modeled nonlinear relationship between the low-level parameters and soil moisture fully into account and is therefore computationally expensive.

The system under consideration can be modeled as:

$$\left\{ \begin{array}{l} \frac{dX}{dt} = M(U, t), \quad t \in]0, T[\\ X(0) = U \end{array} \right.$$

Where U is the initial conditions to be controlled, here $U \in U_{ad}$, where U_{ad} denotes the admissible set of control. X the model state vector in to a Hilbert space, $X = (T_s, T_2, \theta_s, \theta_2)$, and $M = (f_1, f_2, f_3, f_4)$ is the differential operator where :

$$\begin{aligned}
f_1(t, U) &= \frac{1}{\rho_w \cdot d_1} (\alpha T_s + \beta) \exp \left(\frac{(\theta_s - K_1 T_s - K_2) \ln \left(\frac{0.01}{\alpha T_s + \beta} \right)}{(K_1 T_s + K_2)^2} \right) \left(P - \frac{e_{sat}(T_s) - e_0}{r_{as} + \exp \left(A - B \frac{\theta_s}{\theta_{fc}} \right)} \right) \\
&\quad - \frac{1}{\tau} C_{2ref} \left(\frac{\theta_2}{\theta_{sat} - \theta_s - \theta_1} \right) (\theta_s - \theta_{eq}) \\
f_2(t, U) &= \frac{1}{\rho_w \cdot d_2} \left(P - \frac{e_{sat}(T_s) - e_0}{r_{as} + \exp \left(A - B \frac{\theta_s}{\theta_{fc}} \right)} - \rho(1 - f) \left(\frac{e_{sat}(T_v) - e_0}{r_{as} + \exp \left(A - B \frac{\theta_s}{\theta_{fc}} \right)} \right) \right) \\
&\quad - \frac{C_3}{d_2 \cdot \tau} \max[0, (\theta_s - \theta_{fc})] \\
f_3(t, U) &= C_G \left[Rn_s - \frac{\rho C_P}{r_{as}} (T_s - T_0) - \frac{\rho C_P}{\gamma} \left[\frac{e_{sat}(T_s) - e_0}{r_{as} + \exp \left(A - B \frac{\theta_s}{\theta_{fc}} \right)} \right] \right] - \frac{2\pi}{\tau} (T_s - T_2) \\
f_4(t, U) &= \frac{Rn_s - \frac{\rho C_P}{r_{as}} (T_s - T_0) - \frac{\rho C_P}{\gamma} \left[\frac{e_{sat}(T_s) - e_0}{r_{as} + \exp \left(A - B \frac{\theta_s}{\theta_{fc}} \right)} \right]}{(356C(\theta)\lambda_2\tau)^{\frac{1}{2}}}
\end{aligned}$$

We use this approach to minimize the cost function:

$$J(U) = \frac{1}{2} \int_0^T (T_s(U) - T_{s_{obs}})^2 dt \quad (2)$$

Where T_s is the temperature of soil of SVAT-ICARE Model to be assimilate, and $T_{s_{obs}}$ is the vector values of T_s observe. In this work, our approach is based in variational data assimilation where we try to find U_{opt} , the variational method is algorithm to solve the optimization problem

$$J(U_{opt}) = \inf J(U) \quad (3)$$

The condition to have the minimum is the Euler-Lagrange condition

$$\nabla J(U_{opt}) = 0 \quad (4)$$

The determination of ∇J can be performed by implementation of local optimization methods like descent, so the minimization is effectuate using a descent algorithm, for example, conjugate gradient, which requires the evaluation of the quantity ∇J each iteration of minimization. Given the size of the area of control, it would be unrealistic to directly estimate this gradient. In our previous work, it was therefore thought to use a resolution by adjoint method, based on the theory of optimal control [6] and [9], however, the difficulty in this study is writing adjoint code and implementation algorithm to minimize the cost function. In view of recent progress advances in global methods for minimization, it seemed natural to apply a genetic algorithm to estimate initial conditions model ICARE with this new approach.

5 Optimization

In our first work on the control of initial conditions, minimizing the cost function was performed using a standard method of descent. In general, the classical optimization techniques have difficulties in dealing with global optimization problems. One of the main reasons of their failure is that they can easily be entrapped in local minima. Moreover, these techniques cannot generate or even use the global information needed to find the global minimum for a function with multiple local minima. Global optimization refers to finding the extreme value of a given non convex function in a certain feasible region. Solving global optimization problems has made great gain from the interest in the interface between computer science and operations research.

The interaction between computer science and optimization has yielded new practical solvers for global optimization problems, genetic algorithms. Genetic algorithms (GAs) are a family of computational models inspired by evolution. By imitating basic principles of nature they created optimization algorithms which have successfully been applied to a wide variety of problems. The basic principles of GAs are derived from the principles of life which were first described by Darwin (1859). These algorithms encode a potential solution to a specific problem on a simple chromosome (initial conditions in this work). Genetic algorithms are often viewed as function optimizers, although the range of problems to which genetic algorithms have been applied is quite broad.

An implementation of a genetic algorithm begins with a population of (typical random) chromosomes. One then evaluates these structures and allocates reproductive opportunities in such a way that those chromosomes which represent a better solution to the target problem are given more chances to "reproduce" than those chromosomes which are poorer solutions. The "goodness" of a solution is typically defined with respect to the current population.

6 Main results

From simulated data, we will test in this part of the system's ability to assimilate the optimal state vector. The results presented in this section are obtained by running the algorithm developed in this paper, using data (observations) collected in SudMed project during the period 2002-2003 for a wheat parcel. The data set used in this study is to consider the measures and results of the day between 8h: 30 and 17h: 30 for the principal fluxes in this study: sensible heat flux H and latent heat flux LE , and for the surface temperature Tr_0 . The choice of this period corresponds to unstable conditions (length of Monin-Oubukhov < 0) where the fluxes are very important.

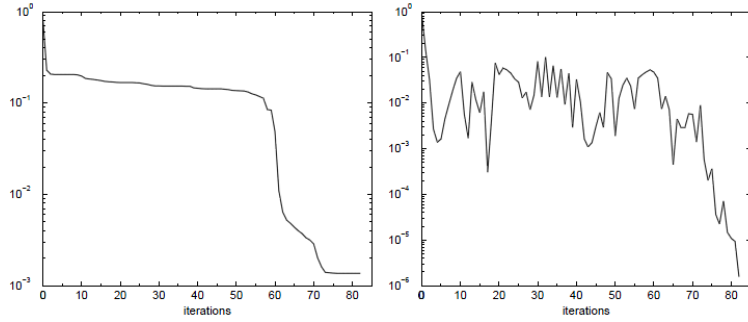


Figure 2: Variations of the cost function J (on the left), and the norm of its gradient ∇J_U (on the right) during the 80 iterations.

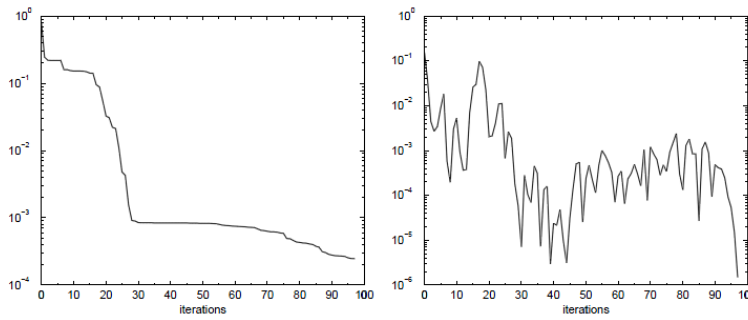


Figure 3: Variations of the cost function J (on the left), and the norm of its gradient ∇J_U (on the right) during the 100 iterations.

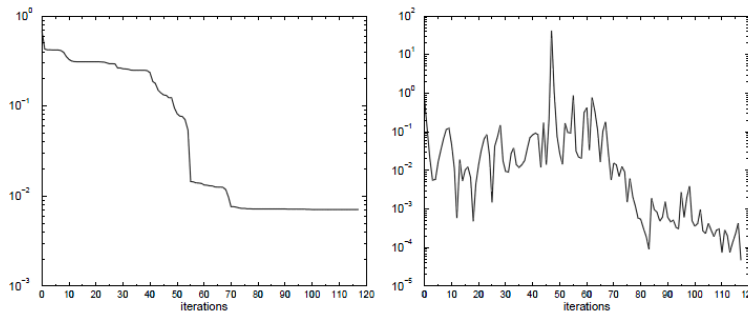


Figure 4: Variations of the cost function J (on the left), and the norm of its gradient ∇J_U (on the right) during the 120 iterations (convergence).

On Figures 2, 3, 4 there is a net decrease of the relative values of the cost functional and the norm of its gradient. The relatively low decay the cost function and the norm of its gradient can be justified by the strong nonlinearity: Those of the model (the equations), the dependence of the cost function compared to initial conditions. The functional to minimize is obviously not quadratic, and therefore we can only find local minima, which in turn depend on the starting point for minimization.

The results obtained on all simulated variables were assessed graphically and statistically. It uses several types of calculations on the state variables or parameters estimated. The diagnoses used in this work are to quantify the difference between a real vector x estimated / simulated and a real vector reference y . For $(x, y) \in IR^N \times IR^N$, the *RMSE* and the R^2 efficiency have been particularly used.

The RMSE (Root mean square error)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_i - X_i)^2}, \quad N \text{ is the number of observations.} \quad (5)$$

The coefficient of determination R^2 , which is interpreted as a measure of the quality of an estimate (regression type, inversion...). It represents the proportion of variance explained by the model to estimate the total variance. More R^2 is close to 1, more fluctuations in the series are explained by the model.

$$R^2(x, y) = 1 - \frac{\sigma^2(x, y)}{\sigma^2(x)} \quad (6)$$

Where σ is the deviation and σ^2 is the variance

$$\sigma(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (7)$$

Where \bar{x} is the arithmetic mean of the vector x

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (8)$$

The RMSE and regression coefficient R^2 will allow us to assess the quality of estimation of initial conditions by assimilation, by comparing estimated parameters to those used to simulate the data before assimilation. The Figure 5 shows the comparison between the fluxes and surface temperature simulated by the ICARE model and measures for the wheat parcel during 2002-2003. This comparison is based on the consideration of results before data assimilation.

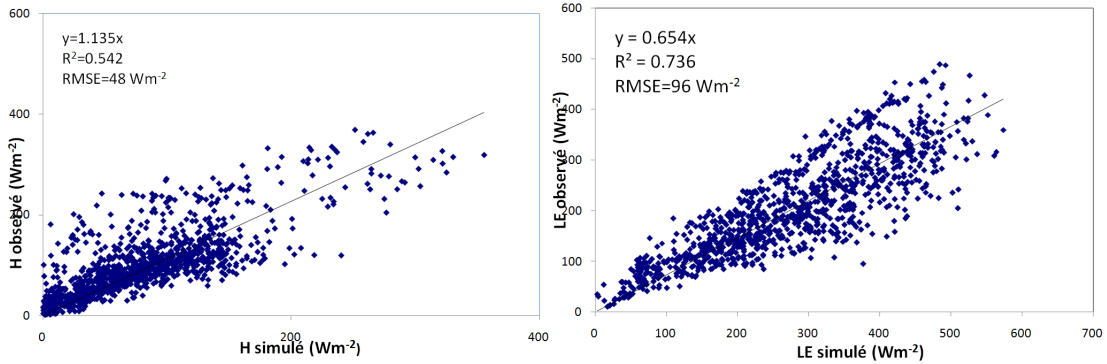


Figure 5: Comparison between measured and simulated fluxes by ICARE model before data assimilation, (on the left the heat sensible flux H) ($\text{RMSE}=48 \text{ Wm}^{-2}$), (on the right the heat latent flux LE) ($\text{RMSE}=96 \text{ Wm}^{-2}$).

The figure 6 shows the comparison between the surface temperature measured and simulated before assimilation. We will see that adjusting the surface temperature decreases significantly the difference between latent heat flux simulated and observed, which is encouraging. Other prospects for assimilation.

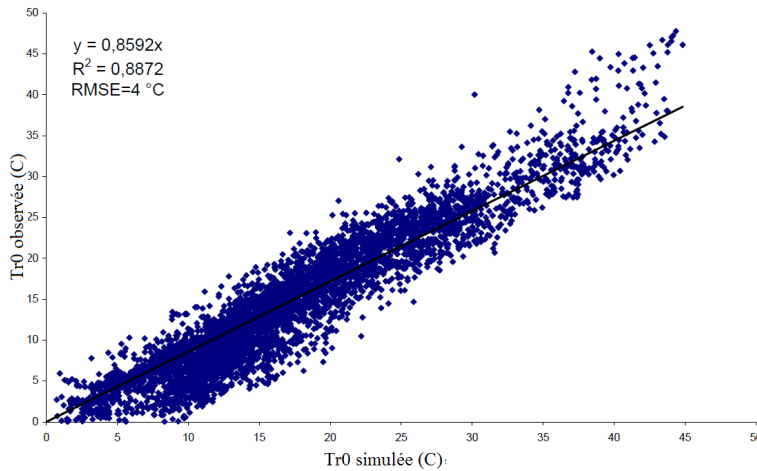


Figure 6: Comparison between soil temperature measured and simulated before assimilation ($\text{RMSE}=4^{\circ}\text{C}$).

6.1 Results after variational assimilation

The assimilation is performed on the soil temperature T_s potentially accessible by Remote Sensing (observations), these observations are well distributed in the cycle: early, maximum growth, senescence and finally phase. The figures 7,

8 and 9 give the results (after assimilation) of the comparison between fluxes H and LE measured and simulated. In general, we can estimate that the results are relatively satisfactory for the flux H , since for the 141 days, the RMSE was reduced, and to a lesser extent for the flux LE with a slight improvement because its measurements are however more strongly marred to uncertainties and concern essentially the second half of simulation, including the stage of senescence cover: the RMSE between the values of H and LE flux simulated and observed throughout the assimilation window is 29 and 68 respectively. In the figure 10, the soil temperature was also improved after assimilation (transition from a RMSE : $4^{\circ}C$ to $3.5^{\circ}C$).

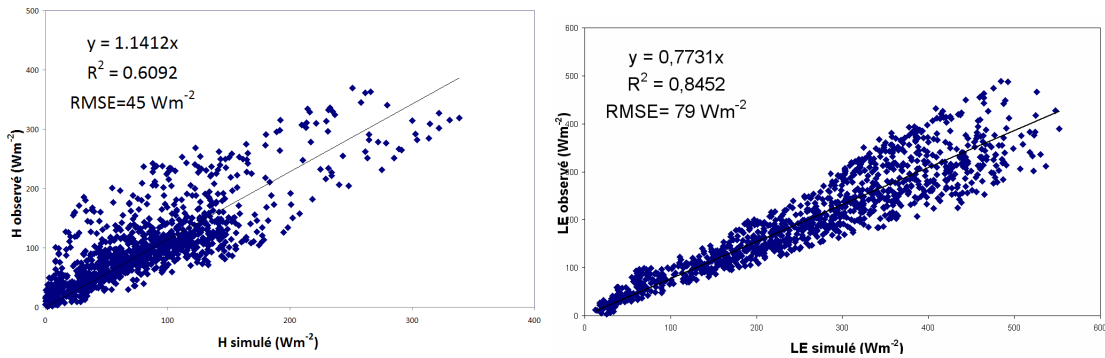


Figure 7: Comparison between measured and simulated fluxes by ICARE model after data assimilation (80 iterations), (on the left the heat sensible flux H)(RMSE= 45 Wm^{-2}), (in right the heat latent flux LE)(RMSE= 79 Wm^{-2}).

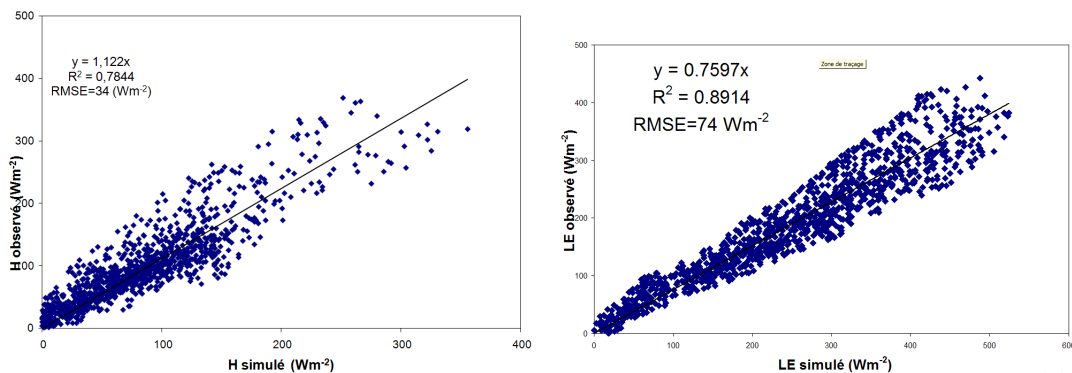


Figure 8: Comparison between measured and simulated fluxes by ICARE model after data assimilation (100 iterations), (on the left the heat sensible flux H)(RMSE= 34 Wm^{-2}), (on the right the heat latent flux LE)(RMSE= 74 Wm^{-2}).

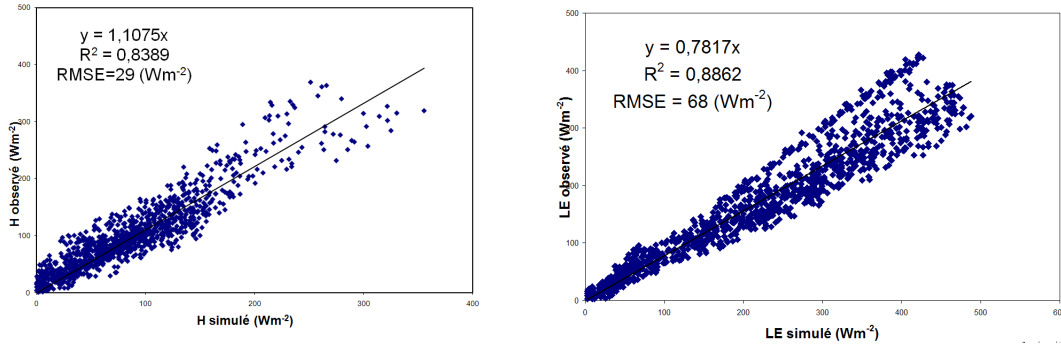


Figure 9: Comparison between measured and simulated fluxes by ICARE model after data assimilation (convergence), (on the left the heat sensible flux H) ($RMSE=29 \text{ Wm}^{-2}$), (on the right the heat latent flux LE) ($RMSE=68 \text{ Wm}^{-2}$).

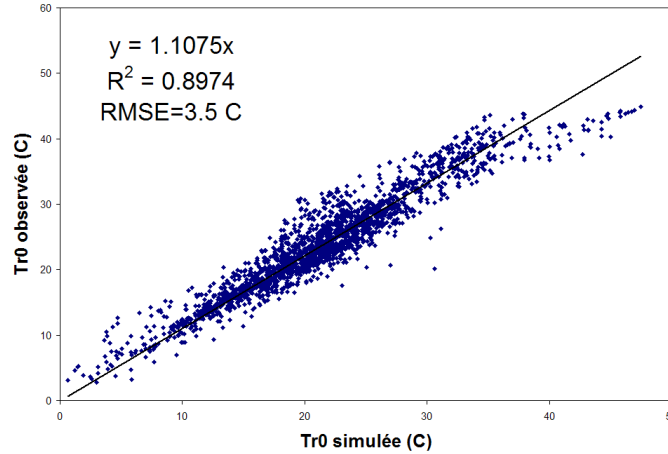


Figure 10: Comparison between soil temperature measured and simulated after assimilation ($RMSE=3.5^\circ\text{C}$).

7 Conclusions and Further works

The aim of this work was to demonstrate the usefulness of mathematics and optimization methods to environmental problems. The method used in this work has been motivated in response to several difficulties encountered in the implementation of the algorithm of assimilation and the minimization method used in our previous paper on the control of initial conditions. Among these difficulties, for examples the large number of parameters in ICARE model, dependencies between these parameters that make complicated the model derivations, minimizing the cost function J which requires the calculation of its gradient subsequently used in the minimization of J , include also the difficulties posed

by the use of a descent method (First-Guess) to optimize the functional J , without forgetting the problems of storage arrays and the computation time and their implementation difficult and laborious.

Global algorithms aim to find the global optima, whereas local algorithms are satisfied with local optima. Local algorithms usually employ "downhill" searching strategy, and the search stops whenever an optimum is found. By contrast, the searching process of global algorithms escapes from the local optima according to either deterministic or stochastic searching methods. The backtracking mechanism, i.e. branch and bound, is being employed extensively in deterministic methods. For stochastic methods, the search jumps away local minima based on probabilistic decisions for instance genetic algorithm developed for this work to improving our understanding of soil moisture and temperature conditions that will help us in many ways. Furthermore, we must develop an evolutionary approach suited to this optimization problem to make efficient and better quality of the changes in populations over time. This method should be sufficiently powerful to manage individuals with high complexity in a reasonable time.

The general objective of this study is to solve the inverse problem, i.e. be able to find the conditions that allow a given output from a given input. The difficulty of this problem stems from the complexity of the model, of course, but also the large number of parameters. Our further work is based on a sensitivity analysis to observe the influence of parameters on the model output. Those who are most sensitive we will be useful to quickly change the output to recover the observed signal. Our current work is to identify sensitive parameters in the ICARE model.

ACKNOWLEDGEMENTS. This work was partially supported by French-Morocco scientific cooperation project "Action Hubert Currien MA/148/06 and MA/164/07", and sponsored by collaboration with the "Institut de Recherches pour le Développement (IRD)".

References

- [1] N. Alaa, W. Bouarifi, G. Chehbouni, G. Boulet, R. Khiri and L. Hanich, "Variational Assimilation: application to the control of the initial data in ICARE model", *IJEES*, Vol.9, No.07, (2007), pp.13-26.
- [2] N. Alaa, W. Bouarifi, G. Chehbouni, R. Khiri, L. Hanich and J. R. Roche, "Assimilation of the soil resistance to evaporation in ICARE model", *IJMS*, Vol.04, No.S09, (2009), pp.38-56.

- [3] J. W. Deardorff, "A parameterization of ground temperature and moisture content with inclusion of layer of vegetation", *J. Geophys.*, Vol.83, (1978), pp.1889 - 1903.
- [4] Lawrence Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, (1991).
- [5] David E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", *Addison-Wesley*, (1989).
- [6] J. L. Lions, Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles, *Edition Dunod*,(1968).
- [7] P.G. Jarvis, The interpretation of the variations in the leaf water potential and stomatal conductance found in canopies in the field, *Phil. Trans. Roy. Soc. Lond.*, **273** (1976), 593 - 610.
- [8] B. J. Choudhury and J. L. Monteith, A four-layer model for the heat budget of homogeneous land surface, *Q. J. R. Meteorol. Soc.*: 373398, **114** (1988).
- [9] F. X. LeDimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations : Theoretical aspects, *Tellus*, **4** (1986), 97 - 110.
- [10] S. Er-Raki, A. Chehbouni, J. Hoedjes, J. Ezzahar, B. Duchemin, F. Jacob, Improvement of FAO-56 method for olive orchards through sequential assimilation of thermal infrared-based estimates of ET, *Agriculture water management*, 2008, vol. 95, no3, pp. 309-321.
- [11] J.P. Lhomme, B.A Monteny and M. Amadou, Estimating sensible heat flux from radiometric temperature over sparse millet, *Agric. For. Meteorol.*, **68** (1994), 93 - 105.
- [12] Lawrence Davis, editor, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, *New York*,(1991).
- [13] E. Falkenauer, A new representation and operators for genetic algorithms applied to grouping problems,*Evolutionary Computation*, **2** (1994), 123 - 144.
- [14] J. Noilhan and S. Planton, simple parameterization of land surface processes for meteorological models, *Weather Rev.*, **117** (1989), 536 - 549.
- [15] W. Brutseart, On a derivable formula for longwave radiation from clear skies, *Water Resour. Res.*, **11** (1975), 742-744.