

Automatic calibration of a farm irrigation model: a multi-modal optimization approach

Amaury Dubois, Fabien Teytaud, Eric Ramat, Sébastien Verel

▶ To cite this version:

Amaury Dubois, Fabien Teytaud, Eric Ramat, Sébastien Verel. Automatic calibration of a farm irrigation model: a multi-modal optimization approach. EA2019 Artificial Evolution, Oct 2019, Mulhouse, France. hal-02406612

HAL Id: hal-02406612

https://hal.science/hal-02406612

Submitted on 12 Dec 2019

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

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

Automatic calibration of a farm irrigation model: a multi-modal optimization approach

Amaury Dubois^{1,2}, Fabien Teytaud¹, Eric Ramat¹, and Sébastien Verel¹

Université du Littoral Côte d'Opale, LISIC,
 rue Ferdinand Buisson, 62228 Calais, France
 Weenat Technocampus Alimentation,
 impasse Therese Bertrand-Fontaine,44320 Nantes

Abstract. In agriculture, plant cultivation requires to take numerous decisions. One of the major problems is the irrigation: a proper irrigation decision has to be made accordingly to the hydric state of the plant and the soil, and the weather prediction. In precision agronomy, this leads to use hydric sensors combined with a numerical model of growth plant model. Such models can not often be tuned by experts. We proposed an automatic parameter calibration of the potato growth model based on data collected in several open fields. As these parameter calibration problem are ill-posed, the associated black-box optimization problem is supposed to be multi-modal. We then compare the performances of two state-of-the-art Evolution Strategies which use different restart mechanisms to automatically tune the set of parameters on different crops, and shows that multi-modal optimization methods may be recommended for such class of optimization problems.

1 Introduction

As others domains (industry, urban, etc.), precision agronomy benefits new sensors which are enhanced by numerical models and simulations. Therefore, the decision making process can be supported by the knowledge bringing by this new numerical environment. For plant cultivation, one major decision is the irrigation. The farmer have to decide to irrigate a field according to the hydric state of plant and the soil, the weather prediction, and the cost of an irrigation. In that case, new decision-making technic uses hydric sensors to measure the quantity of humidity of the soil, and, a growth plant model to able to estimate the hydric state of the plant, and the available quantity of water for the plant which depends on the root size, and the characteristics of the soil. Although such numerical approach can lead to an accurate prediction of the crop state, and beyond the sensor precision, one drawback is the setting of the numerous parameters of the plant growth model. Indeed, models combines different submodels based on differential equations, finite state transitions, etc. that require the settings of many numerical biological, or geological parameters. Even if experts can measure, estimate, or give bounds of some parameter values³, most of

³ Some parameters can also have no meaning from a biological point of view.

times precise value of parameters are not known for a field-scale crop as they depends on specific soil, and plant specie/variety. In this work, we show that it is possible to set precisely the model parameters of potato plant growth based on the data acquisition of hydric sensors, and a relevant optimization algorithm that minimizes the distance between a predicted values computed by the model, and real data.

In evolutionary computation, this black-box problem is known as a calibration problem [19,7]. The parameter settings of potato plan growth model shows specific difficulties. As the parameters depends on local specificity such as soil, potato variety, weather exposition, etc. the data are difficult to collect, and rare: a campaign of data acquisition with hydric sensors for a potato field lasts 4 months, and can be done on the same field every 3 years due to crop rotation. On the other hand, the number of model parameters is high: several dozens for representative models such as STICS [22], AquaCrop [14], or Weedric [15]. As a consequence, the calibration problem of potato growth model is ill-posed. Several parameters settings lead to the same input-output behavior, and the simulations are consistent with the measurements in the field. Thus, the optimization problem related to model calibration is not only a problem with many local optima, but a multi-modal problem for which the quality of several local optima is very close to those of the global optima.

In this work, we formulate the calibration optimization problem from the farming irrigation model Weedric (defined in section 2.1), and available hydric data. More precisely, we have a model with many continuous parameters for which we have no a priori knowledge about their implications in equations, simulations or the interactions between the different model parts. Experts can only define the bounds of each parameter values. So, we consider it as a continuous black-box function from the search space of dimension d > 1: $[0,1]^d$. As stated above, we assume that this problem is highly multi-modal, with one global optimum but has many local optima close to the global one. It is therefore necessary to check the maximum number of optima in order to determine the global one.

This kind of problems is known as Multi-Modal Optimization (MMO). Numerous algorithms have been proposed, many of them use the derivative of the gradient but in black box context, these algorithms are not directly applicable. Gradient free methods are generally based on Evolution Strategy (ES) which have shown their robustness and their efficiency [6,16,17]. ES consists in generating better solutions iteratively from a starting point. In MMO context, ES are generally combined with either a niching technique [12,13] or a restart strategy [1,3,10], in order to find all the optima. In this paper, we propose to compare different state-of-the-art restart strategy algorithms: QRDS [20,18] and CMAES-IPOP [3] to automatically tune the parameters model.

The rest of this paper is organized as follows: Section 2.1 describes the *Weedric* simulator. Then, we present the Quasi-Random Restart Strategy (QRDS) and Covariance Matrix Adaptation Evolution Strategy with Increasing POPulation (CMAES-IPOP) in section 3. Next, we compare their performances in section 4. Finally we conclude in section 5.

2 Calibration problem of an irrigation model

2.1 Farming irrigation model

Many models have been proposed to deal with this plant growth. For example, STICS [22] is a deterministic generic model for the simulation of crops and their water and nitrogen balance developed at INRA institute (France) since 1996. It calculates both agricultural variables (yield, input consumption) and environmental variables (water and nitrogen losses). AquaCrop [14] is also a deterministic generic model. It provides an estimation of crop productivity in relation to water supply and agronomic management in a framework based on current plant physiological and soil water budgeting concepts. Unfortunately, these models requires a large number of parameters such as they are generic, several type of plants can be modeled, and involve other biological mechanisms in addition to the irrigation issue.

Weedric is an agricultural irrigation model for the culture of potatoes developed by the Weenat company ⁴. This model has emerged from a collaborative project between computer, and agriculture researchers [15]. It is intended to farmers in order to help them with a decision support. The Weedric model consists in the combination of several deterministic existing biological models [4,2,11,5] to provide a specific model for this kind of culture. Initially, theses models are independent, and the interest of the Weedric model (see Fig. 1) is to be able to connect them in order to propose two high-level models:

- Soil model: this model considers the soil as a succession of horizontal layers and each layer has a quantity of water varying over time according to the different exchanges between the layers (percolation, upwelling), the weather (temperature, rain, wind,...) and the interactions with the Plant model.
- Plant model: it simulates the development and the behavior of a potato plant from planting to harvesting, based on current water quantity and weather forecasts.

Using the planting date and the weather forecasts, the multi-model Weedric can predict the water stress of the potato plant, and the Soil Water Content (SWC). The SWC is the available quantity of water that the potato plant can use, and extract from the soil. Unfortunately, to be fully effective, the d=38 real parameter values must be tune for a particular potato variety and soil type.

2.2 Black-box calibration problem

The goal is to calibrate the Weedric model using the Soil Water Content (SWC), and the hydric sensors. Sensors are put in the field, and they regularly send data. This makes it possible to obtain an approximation of the pressure of the water in the ground which can be converted into a quantity of water using the well know Van Gernuchen equation [21]. The black curve "sensors" of the Fig. 2 shows the SWC over a season of n=73 days.

⁴ https://www.weenat.com/

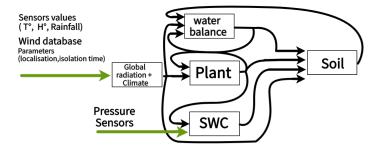


Fig. 1: Simplified diagram of how the multi-model Weedric growth crop model works. This one is divided into several sub-models in which the d=38 variables are assigned. The green arrows indicate the entry points of the model by which the different inputs (pressure, temperature, wind, rainfall, ...) can be filled in. The black arrows indicate the interaction and information sharing of the models between them. SWC is the Soil Water Content under interest in this work.

Following the expert knowledge, a set of default parameters value is defined. The green curve "default" on the Fig. 2 shows the predicted SWC by the Weedric using those default values. From the first day to approximatively the 30th day, the predicted value follows the measured values by sensors. During this period, the SWC increasing is mainly due to the increasing of the roots. During a dry period after the 30th day, the model with default parameter values seems to over-estimate the dryness of the soil which could be due to misleading values of the soil model, or plant model, or a combinaison of the two. The interaction of different components of the model are not linear.

The fitness function of the calibration problem is defined as the root mean square error over the crop period of the SWC value predicted by the model. More formally, for every settings $x \in [0,1]^d$, with d=38, of the Weedric model, the fitness function f is defined by:

$$f(x) = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (SWC_t - \widehat{SWC}_t)^2}$$
 (1)

where n is the number of days of culture period, SWC_t is the SWC at the day t measured by the sensors, and $\widehat{SWC_t}$ is the predicted SWC by the model with the parameters settings x. The fitness function is to be minimized in order to reduce the distance between predicted, and real SWC values.

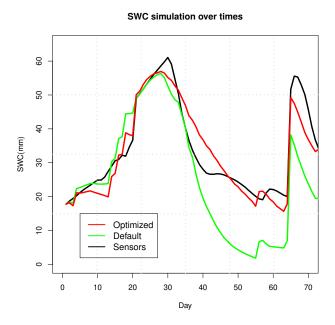


Fig. 2: Evolution of the Soil Water Quantity (SWC) over time (n=73 days). The black curve represents the sensor values, the green represents the prediction of the model with the basic values and the red curve, the prediction of the model with the optimized values. The closer the curves are to the black curve, the better the prediction. At the beginning of the simulation the two parameter sets are quite similar but over time, the model with the default parameters is no longer accurate.

3 Multi-Modal Optimization algorithms

In this section we present the Quasi-Random restarts with Decreasing Step-size algorithm (QRDS) and the Covariance Matrix Adaptation Evolution Strategy with Increasing POPulation (CMA-ES IPOP), one of its variants for multi-modal problems.

3.1 Quasi-Random restarts with Decreasing Step-size

Random restarts with Decreasing Step-size and its improvement (Quasi-Random restarts with Decreasing Step-size[18]) are an Evolution-Strategy-based Multi-Modal Optimization algorithms which uses the restarting technique.

It is composed by a simple local search algorithm combined with a restart strategy following a random (or quasi-random) sequence. Local search is a simple (1+1)-ES using the 1/5 adaptation rule (see Algorithm 1).

At the beginning, a point is selected. Then, Iteratively, the algorithm generates a candidate by mutating the current best point according to a normal distribution with a standard deviation (step-size) σ . The best of both points is kept. The update of the step-size σ is really simple: if the candidate solution, i.e. the new generated point is better, the step-size σ is increased, otherwise σ is decreased as we may need to focus on smaller neighborhood.

We use this step-size value as the stopping criterion of the local search. If it is too small, we consider that the local search has converged to an optimum (global or local). The solution is saved and the local search is restarted until the evaluation budget is reached.

A feature of QRDS is its "murder operator". In order to avoid converging to an already known solution, the algorithm checks at each evaluation, if the current solution is greater than a distance $\delta_{threshold}$ of all the optima already discovered. If it is true, the search is aborted without saving the solution (we don't want to spend time for an already found optimum).

For the restart strategy (see algorithm 2), each time the algorithm is restarted, the initial position is sampled according a quasi-random sequence.

3.2 CMA-ES IPOP

The Covariance Matrix Adaptation Evolution-Strategy is an Evolution Strategy that adapts the full covariance matrix of a normal search distribution [9]. This algorithm is presented in Algorithm 3. An important property of this algorithm is its invariance against linear transformations of the research space. CMA-ES is effective in minimizing unimodal function, and is superior when the problem is ill-conditioned and non-separable. In multi-modal context, [8] shows that increasing the size of the population can improve performances of the CMA-ES. [3] proposes a version of CMA-ES using a restart strategy: at each restart (whenever the stopping criterion is met), the size of the population is doubled see Algorithm 4. By increasing the population size, the local search becomes more global after each restart. The results given in [3] show that this improvement provides good performances on multi-modal black-box context.

4 Experimental analysis

We compare the multi-modal algorithms, on the Weedric model calibration problem defined in the previous section 2. The data from 5 different crops are used with different soil types, and potato varieties. The two algorithms are also compared with default parameters values given by the experts. For each crop, the number of independent runs of each algorithms is 100. The maximum number of evaluations for each algorithm is 10^5 . Notice that the simulation time of Weedric model is enough short for such number of evaluation within minutes.

All the results are reported in Tab. 1, and correspond to the best value found over the 100 runs, mean with confidence interval, and the median value. Both

Algorithm 1: SearchDS

```
{Search an optimum using a Decreasing Step-size }
    Input:
       f: function to optimize
       \sigma_0: initial step-size
       \epsilon_{\sigma}: threshold value of the step-size
       y^*: maximum fitness of the function
       \epsilon_y: threshold value of the fitness
       x: initial position for the search
       \epsilon_x: threshold value of the position
       \hat{\mathbf{X}}: set of previously found optima
     Output:
       \hat{\mathbf{X}}: updated set of optima
 1 begin
 2
          y \leftarrow f(\mathbf{x})
          \sigma \leftarrow \sigma_0
 3
          repeat
 4
                {mutation}
                \mathbf{x}' \leftarrow \mathcal{N}(\mathbf{x}, \sigma)
 5
                y' \leftarrow f(\mathbf{x}')
 6
                {selection with 1/5th adaptation}
                if y' > y then
 7
                     \mathbf{x} \leftarrow \mathbf{x}'
 8
                     \sigma \leftarrow 2\sigma
 9
                else
10
                 \sigma \leftarrow 2^{-1/4}\sigma
11
                {discard search if optimum already known}
                if \exists \hat{\mathbf{x}} \in \hat{\mathbf{X}}, \|\mathbf{x} - \hat{\mathbf{x}}\| < \epsilon_x then
12
                 break
13
                {store found optimum}
                if ||y-y^*|| < \epsilon_y then
14
                     \hat{\mathbf{X}} \leftarrow \hat{\mathbf{X}} \cup \{\mathbf{x}\}
15
                      break
16
          until \sigma < \epsilon_{\sigma} or max nb of function evaluations reached
17
```

algorithms substantially improved the default settings of the experts. These results show the relevance of using data-oriented calibration with an ES algorithm on this kind of real world application since it is able to find a set of parameters that allows the model to correctly predict sensor values. Moreover, according to the Mann-Whitney test at confidence level 0.01, the QRDS outperform the CMAES-IPOP on all crops. Figs. 3 shows the dispersion of the values found over all runs (smaller is better). The restart strategy of QRDS seems to be more effective on such multi-model problem. Indeed, the exploration behavior of the QRDS allows to find more interesting search space area. As an example of the

Algorithm 2: RDS

```
{Random restarts with Decreasing Step-size}
   Input: f, \sigma_0, \epsilon_\sigma, y^*, \epsilon_y, \epsilon_x
   Output: X
1 begin
         \mathbf{\hat{X}} \leftarrow \varnothing
\mathbf{2}
          repeat
3
                 \mathbf{x} \leftarrow \text{sample search-space}
4
                 \hat{\mathbf{X}} \leftarrow \text{SearchDS}(f, \sigma_0, \epsilon_\sigma, y^*, \epsilon_y, \mathbf{x}, \epsilon_x, \hat{\mathbf{X}})
5
         until all optima found or max nb of function evaluations reached
```

Algorithm 3: CMA-ES

```
{Covariance Matrix Adaptation Evolution Strategy}
   Input:
    \bar{f}: function to optimize
    \lambda : number of sample per iteration
    x: set of population
    s: set of fitness
   Output: x_1: best optima find so far
1 begin
       Initialization while stop criterion not met do
\mathbf{2}
           for i in 1...\lambda do
3
               x_i = \text{sample}
 4
               s_i = f(i)
 5
               Sort(x_{1...\lambda})
6
               Update_mean_to_better_solution()
               Update_isotropic_evolution_path()
8
               Update_anisotropic_evolution_path()
9
               Update_covariance_matrix()
10
               Update_step-size()
```

Algorithm 4: CMA-ES-IPOP

```
{Covariance Matrix Adaptation Evolution Strategy with Increasing
            POPulation } Input:
    f: function to optimize
    \lambda: number of sample per iteration
    \lambda_0: size of the initial population
    n: number of restart
    X: set of solution
  Output: X or X_1: set of optima or the best one
1 begin
       n \leftarrow 0 while number of evaluations not reach do
2
           n \leftarrow n + 1
3
           \lambda \leftarrow \lambda_0 + 2 \times n
4
           X \leftarrow \operatorname{cmaes}(\lambda, f)
```

result, the Fig. 2 shows the predicted SWC by the best parameter settings of the Weedric model. In particular, the parameters setting improves the prediction for the dry period after the 30 day.

The second experiment consists in testing the robustness of an optimized parameter set. To do this, we select the parameter set that has obtained the best (smallest) fitness (one solution of the crop 4), then we use this set (from crop4) on the other crops. Tab. 2 presents the results of the corresponding fitness with the fitness of the best optimized solutions and default parameters. We can see that, indeed, results are not as good as a specific optimization, but the solution is robust enough as it is by far better than the specific expert parameters. Fig. 4 presents the results of the corresponding fitness with the fitness of the best optimized solutions and basic parameters. Without a priori knowledge, statistically QRDS always finds better solutions than CMAES-IPOP.

5 Conclusion

In this paper we propose to describe a farming decision-making model for irrigation into a black box optimization problem and we experiment on two state-of-the-art algorithms QRDS and CMA-ES IPOP. Results show that for this kind of problem using an ES algorithms is very efficient, since independently of the crops, the set parameters calibrated by the algorithms are always significantly better than those by default (which have been designed by experts or available in the literature). Moreover, the multi-modal QRDS seems to be very effective on such problems.

The proposed method is "offline" which means that the optimization of the parameter sets can only be done once the growing season is completed. Moreover, in agriculture, it is impossible to replant the same plant before 3 to 5 years on the same crop (and obviously the weather changes from year to year). This is for these reasons, that it is interesting to be able to find a robust solution, which may

Table 1: Best, median and mean fitness (with the confidence interval at 95%) found by each algorithm over 100 runs with a budget of 10^5 evaluations (smaller is better). The bolded median values are significatively better according to the Mann-Whitney at confidence level of 0.01.

Crop	Algorithm	Best	Mean	Median
Crop 1	Default QRDS CMA-ES IPOP	15.8	$/$ 16.3 ± 0.03 18.5 ± 0.12	
Crop 2	Default QRDS CMA-ES IPOP	16	16.7 ± 0.03	
Crop 3	Default QRDS CMA-ES IPOP	17.5	18 ± 0.03	
Crop 4	Default QRDS CMA-ES IPOP	14.2	15.8 ± 0.06	
Crop 5	Default QRDS CMA-ES IPOP	15.1	15.3 ± 0.02	

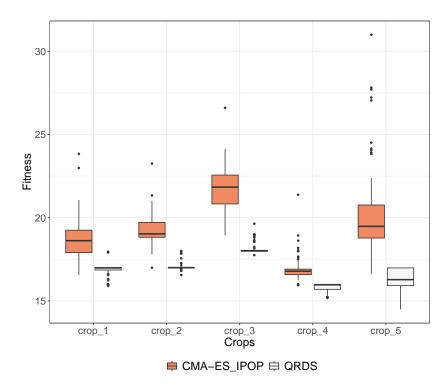


Fig. 3: Spread out of the best solutions found by QRDS and CMA-ES IPOP over 100 runs (smaller is better) on the 5 crops.

Table 2: Evaluation of the robustness of the best set of parameters found on crop 4 on the other crops compared to the fitness of their best optimized solution as well as the basic one (smaller is better). The best set of parameters ever found is not as good as the specific optimization of the problem but greatly improves fitness compared to the basic values.

Parameters	Crop1	Crop2	Crop3	Crop5
Default	80.83	57.81	47.85	50.74
Best Optim	15.8	16	17.5	15.1
Crop 4	38.27	25.78	31.8	17.71

still be good for the next year. In order to have a solution as robust as possible, a possible future work could be to optimize different crops at the same time. Finally, we could also develop an "online" method able to predict the outputs of the model according to the already collected data and parameter sets.

Acknowledgement. The authors would like to thank the **WEENAT** company in particular for the financing of the CIFRE thesis and for their material support. Experiments presented in this paper were carried out using the CALCULCO computing platform, supported by SCOSI/ULCO (Service COmmun du Système d'Information de l'Université du Littoral Côte d'Opale).

References

- 1. Ahrari, A., Deb, K., Preuss, M.: Multimodal optimization by covariance matrix self-adaptation evolution strategy with repelling subpopulations. Evolutionary computation **25**(3), 439–471 (2017)
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., et al.: Crop evapotranspirationguidelines for computing crop water requirements-fao irrigation and drainage paper 56. Fao, Rome 300(9), D05109 (1998)
- 3. Auger, A., Hansen, N.: A restart cma evolution strategy with increasing population size. In: 2005 IEEE congress on evolutionary computation. vol. 2, pp. 1769–1776. IEEE (2005)
- 4. Beaujouan, V.: Modélisation des transferts d'eau et d'azote dans les sols et les Nappes. Développement d'un modèle conceptuel distribué. Application à de petits bassins versants. Ph.D. thesis, Ecole Nationale Supérieure Agronomique de Rennes (2001)
- 5. Beaujouan, V., Durand, P., Ruiz, L.: Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments. Ecological modelling 137(1), 93–105 (2001)
- 6. Beyer, H.G.: The Theory of Evolution Strategies. Springer Science & Business Media (2001)
- 7. Gupta, H.V., Beven, K.J., Wagener, T.: Model calibration and uncertainty estimation. Encyclopedia of hydrological sciences (2006)

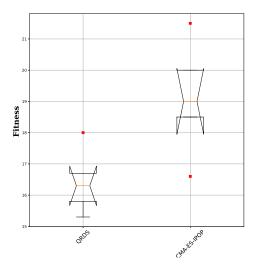


Fig. 4: Spread out of the mean fitness for QRDS (left) and CMA-ES IPOP (right) on the various crops. the average fitness of QRDS is statistically better (smaller) than that of CMA-ES IPOP .

- 8. Hansen, N., Kern, S.: Evaluating the cma evolution strategy on multimodal test functions. In: International Conference on Parallel Problem Solving from Nature. pp. 282–291. Springer (2004)
- 9. Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). Evolutionary computation 11(1), 1–18 (2003)
- Kadioglu, S., Sellmann, M., Wagner, M.: Learning a reactive restart strategy to improve stochastic search. In: International Conference on Learning and Intelligent Optimization. pp. 109–123 (2017)
- 11. KB Johnson, SB Johnson, P.T.: Development of a simple potato growth model for use in crop-pest management. Agricultural systems 19(3), 189–209 (1986)
- 12. Li, X.: Multimodal Optimization using Niching Methods, pp. 1–8. American Cancer Society (2016)
- 13. Preuss, M.: Multimodal Optimization by Means of Evolutionary Algorithms. Springer Publishing Company, Incorporated, 1st edn. (2015)
- 14. Raes, D., Steduto, P., Hsiao, T.C., Fereres, E.: Aquacrop—the fao crop model to simulate yield response to water: Ii. main algorithms and software description. Agronomy Journal **101**(3), 438–447 (2009)
- 15. Ramat, E., Vandoorne, B.: Plant growth model for decision making support. Tech. rep., Université du Littoral Côte d'Opale, and ISA Lille (2002)
- 16. Rapin, J., Teytaud, O.: Nevergrad A gradient-free optimization platform. https://GitHub.com/FacebookResearch/Nevergrad (2018)
- 17. Rechenberg, I.: Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. No. 15 in Problemata, Frommann-Holzboog (1973)

- 18. Schoenauer, M., Teytaud, F., Teytaud, O.: A Rigorous Runtime Analysis for Quasi-Random Restarts and Decreasing Stepsize. In: Artificial Evolution. Angers, France (2011)
- 19. Tang, Y., Reed, P., Wagener, T.: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrology and Earth System Sciences Discussions 2(6), 2465–2520 (2005)
- 20. Teytaud, F., Teytaud, O.: Qr mutations improve many evolution strategies: A lot on highly multimodal problems. In: Proc. of the 2016 GECCO conference. pp. $35-36\ (2016)$
- 21. Van Genuchten, M.T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. Soil science society of America journal 44(5), 892–898 (1980)
- 22. Whisler, F., Acock, B., Baker, D., Fye, R., Hodges, H., Lambert, J., Lemmon, H., McKinion, J., Reddy, V.: Crop simulation models in agronomic systems. In: Advances in agronomy, vol. 40, pp. 141–208. Elsevier (1986)