CIÊNCIAᵉNATURA 35

Multi-objective calibration of IBIS model by genetic algorithm with parametric sensitivity analysis

Amarísio S. Araújo, Haroldo F. de Campos Velho[2], Lu Minjiao[3]

[1] Universidade Federal de Viçosa (UFV), Viçosa (MG), Brazil
amarisio@ufv.br
[2] Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos (SP), Brazil
haroldo@lac.inpe.br
[3] Nagaoka University of Technology, Nagaoka (Niigata), Japan
u@nagaokaut.ac.jp

Abstract

Atmospheric circulation models combine different modules for a good description of the atmospheric dynamics. One of these modules is the representation of surface coverage, since the dynamics depends on the interaction between the atmosphere and the surface of the planet. However, these modules depend on a number of parameters that need to be adjusted. The parameter adjustment process is called model calibration. In this study, the IBIS (Integrated Biosphere Simulator) model is calibrated following a multi-objective strategy. The Pareto set, which embraces the non-dominated solutions in the search space of objective functions, is determined by a version of multi-objective genetic algorithm (NSGA-II). The model sensitivity to the parameters is evaluated by the Morris' method. Synthetic data for calibration were obtained from the Tapajós National Forest (FloNa Tapajós), located near to the 67 km from Santarém-Cuiabá highway (2,51S, 54,58W).

Keywords: Multi-objective calibration, IBIS model, Morris' method, NSGA-II, FloNa Tapajós.

1 Introduction

Mathematical equations are largely employed for modeling complex phenomena, including the geophysical fluid dynamics. For representing the dynamics appropriately is necessary to attach several modules to simulate the physical processes, most of them are not yet fully understood. One of these processes is the interaction between the surface and atmosphere.

The transfer of heat, moisture and momentum is strongly related to the type of surface coverage. Therefore, surface models are modules to represent the interaction between the surface and the atmosphere. Unfortanately, such modules has many parameters to be identified. The process to calculate the model parameters is called *calibration*. One strategy to determine the model parameters is to compute a set of parameters for providing the best agreement between the model and the observation data.

The IBIS (Integrated BIosphere Simulator) model is a surface model representing diferent processes between the surface and the atmosphere (Foley et al., 1996). However, such model is unable to carry out a correct representation without a good estimation of the model parameters. The procedure to determine the cited parameters is called *calibration*. Calibration is a type of inverse problem: *parameter estimation* (Beck et al., 1985). The calibration process can be subjetive or objective. Here, a multi-objective calibration will be applied to the IBIS model. The inverse problem is formulated as an optimization problem.

There are an extensive literature to deal with multi-objective optimization, for instance: Hwan and Masud (1979), Miettinen (1999), Coello et al. (2007), Deb (2009). Our calibration is based on the NSGA-II (Non-dominated Sorting Genetic Algorithm) – see Deb et al. (2000).

Different parameters have distinctive impacts on the models. In addition, predictive models can be applied to different time scale to forecast. Therefore, nowcasting, forecasting, and climate models the same parameter could have different influence on the model. The Morris' method can be employed to select the parameters in different classes, according of their impact on the model.

This methodology is based on the computation of the sensitivity analysis, evaluating the model response against the parameter changes (Morris, 1991; Minjiao and Xiao, 2014).

A brief description for the IBIS model is addressed in Session 2. The calibration scheme is explained in Session 3, where the evaluation funcionals are introduced (objective functions), the Morris' approach is discussed, and the method NSGA-II to estimate the Pareto set is presented. The last session is dedicated to the final remarks.

2 The IBIS model

The IBIS model is designed to explicitly connect Earth's surface and hydrological processes, terrestrial biogeochemical cycles, and dynamics of vegetation in a single structure modeling aimed at a better understanding of the consequences of human activity in global biophysical processes (Foley et al. 1996; Kucharik et al., 2000).

IBIS was developed by the Center for Sustainability and the Global Environment (SAGE). The model is classified as a dynamic global vegetation one. In addition, to being able to simulate the dynamics of vegetation phenology, the model includes physical processes of the Earth's surface interactions between the land surface and the atmosphere, carbon and nitrogen cycles, and cycles for nutrient, also simulating the effects of changes in terrestrial carbon balance climate and atmospheric CO_2 concentration.

There are several processes emulated by the IBIS. Here, we point out only the variables (model ouputs: 9) used during our calibration process:

1. PAR_0: reflected photo-synthetically active radiation;
2. f_{APAR}: absorbed fraction of photo-synthetically active radiation;
3. R_n: radiation balance;
4. u_*: wind friction velocity;
5. HE: sensible heat flux;
6. LE: latent heat flux;
7. NEE: net exchange of the ecosystem;
8. NPP: net primary production;
9. LAI: leaf area index.

The output functions listed above will be denoted by $Y^{(j)}$, where $j = 1, ..., 9$. Model outputs will be compared with field measurements.

Figure 1 depicted the IBIS structure. The module "Atmophere" represent the data from the gaseous environment. The "Land Surface Module" is the module to simulate the canopy physics, soil physics, and plant physiology, with the relationship among these processes. "Vegetation Dynamics" module calculates the respiration and the net primary production (allocation, turnover, and competition among different plants). The last module is "Carbon and Nitrogen Cycling", where the soil-atmosphere gases exchange is performed. Typical time-scales for the processes is shown in the bottom of the figure.

All processes listed above are coupled, and they are defined according to mathematical equation used to model each process. The entire set of equations needs 43 parameters to be identified. Table 1 shows the parameter list to be found.

3 Calibration Procedure

There are 3 steps for our calibrarion scheme to determine the 43-parameters: (a) define the obejective(s) function(s), (b) compute the IBIS sensitivity for each parameter (Morris' method), (c) apply the optimizer.

3.1 The objective functions

Many real-world problems have a collection of goals to be considered. In many cases, the goals are in conflict with each other, that is, improvement of some(s) target(s) implies(y) in a worse performance to another one(s). Computer aquisition is an example of conflicting goals: fast processing vs price (better performance implies higher price). Under such situation, someone can identify a list os *solutions*, describing computer configurations and associated price. There is a lowest price, but with poor processing. There also is a computer with the best performance, but probably associated to the highest price. These extreme cases show two *non-dominate solutions*. There are more pairs computer_configuration-and-price non-dominated (they are not indicate here).
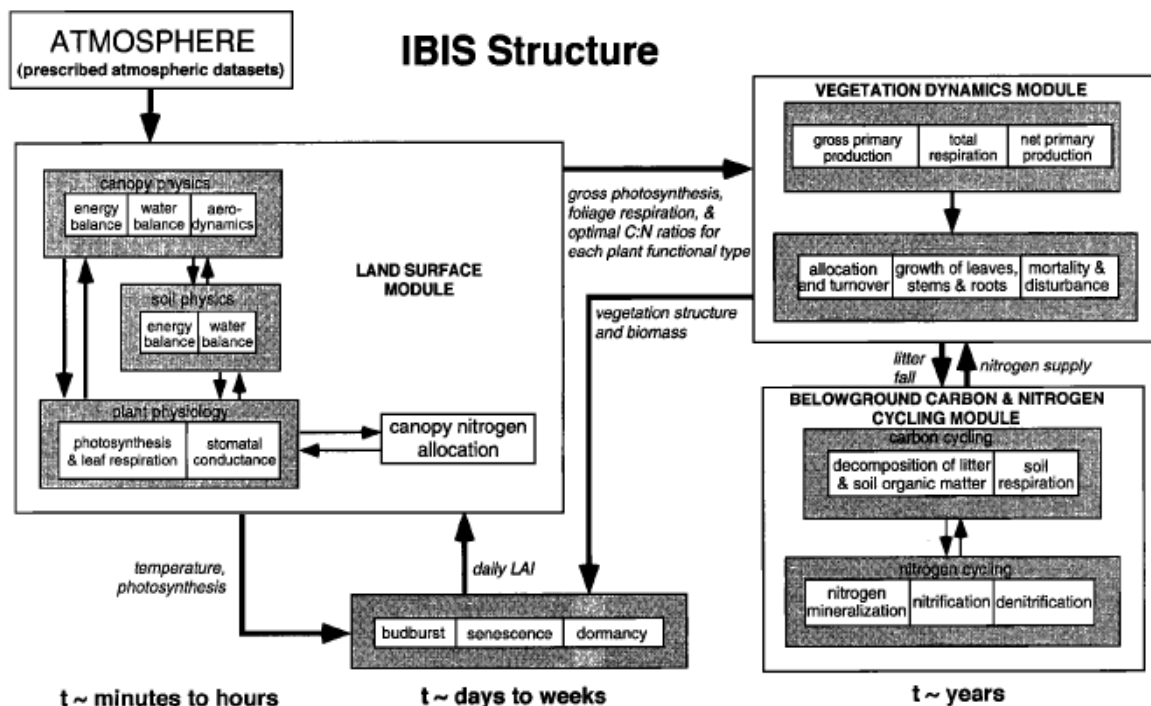

Figure 1 - Structure for the Dynamic Global Vegetation Model (DGVM) IBIS (following: Kucharik et al., 2000)

Table 1 - Description for the IBIS Parameters

No	Name	Description
1	*rhoveg_vis*	Leaf reflectance in the upper canopy - visible (dimensionless)
2	*rhoveg_NIR*	Leaf reflectance in the upper canopy - NIR (dimensionless)
3	*tauveg_vis*	Leaf Tramitância in the upper canopy - visible (dimensionless)
4	*tauveg_NIR*	Leaf Tramitância in the upper canopy - NIR (dimensionless)
5	*chifuz*	Sheet guiding factor in the upper canopy (-1: vertical, 0: random, 1: horizontal)
6	*vmax_pft*	Capacity maximum activity of the enzyme Rubisco (mol - CO_2 $m^{-2}s^{-1}$)
7	*coefmub*	Coefficient related to the stomatal conductance (dimensionless)
8	*chs*	Heat capacity of the upper canopy branches ($Jkg^{-1}m^{-2}$)
9	*chu*	Heat capacity of the upper canopy leaves ($Jkg^{-1}m^{-2}$)
10	*chl*	Thermal capacity of the leaves and the lower canopy branches ($Jkg^{-1}m^{-2}$)
11	*beta2*	Parameter related to the distribution of fine root (dimensionless)
12	*funca_coef*	Temperature coefficient function of the biomass of branches (dimensionless)
13	*funcb_coef*	Temperature coefficient function of the root biomass (dimensionless)
14	*rroot_coef*	Maintenance respiration coefficient of the roots (s^{-1})
15	*rwood_coef*	Coeficiente breathing maintenance of logs (s^{-1})
16	*Rgrowth_coef*	Growth respiration coefficient (dimensionless)
17	*Tempvm_coef*	Parameter thermal stress Vmax (dimensionless)
18	*stressf_coef*	Coefficient related to water stress, soil (dimensionless)
19	*clitll_coef*	Initial carbon in the litter container sheet - lignin (kg - Cm-2)
20	*clitlm_coef*	Initial carbon in the litter container sheet - metabolic (kg - Cm-2)
21	*clitls_coef*	Initial carbon in the litter container sheet - structural (kg - Cm-2)
22	*clitrl_coef*	Initial carbon in the litter reservoir of fine roots - lignin (kg - Cm-2)
23	*clitrm_coef*	Initial carbon in the litter reservoir of fine roots - metabolic (kg - Cm-2)
24	*clitrs_coef*	Initial carbon in the litter reservoir of fine roots - Structural (kg - Cm-2)
25	*clitwl_coef*	Initial carbon in wood litter reservoir - lignin (kg - Cm-2)
26	*clitwm_coef*	Initial carbon in wood litter reservoir - metabolic (kg - Cm-2)
27	*clitws_coef*	Initial carbon in wood litter tank - structural (kg - Cm-2)
28	*csoipas_coef*	Initial carbon in the soil - passive humus (kg - Cm-2)
29	*csoislon_coef*	Initial carbon in the soil - humus slow unprotected (kg - cm-2)
30	*csoislop_coef*	Initial carbon in the soil - slow protected humus (kg - cm-2)
31	*wsoi_coef*	The initial soil moisture (dimensionless)
32	*kfactor*	Multiplication factor of the decline of C contained in reservoirs (dimensionless)
33	*tauleaf*	Return time constant of leaf biomass (years)
34	*tauroot*	Return time constant of fine root biomass (years)
35	*specla*	Return time constant of wood biomass (years)
36	*aleaf*	Specific leaf area (m^2kg^{-1})
37	*aroot*	Carbon fraction in the allocation sheets (dimensionless)
38	*awood*	Carbon allocation fraction in fine root (dimensionless)
39	*dispu_coef*	Carbon allocation fraction in the wood (dimensionless)
40	*dispu_coef*	Zero plane displacement height for the upper canopy (m)
41	*alogl_cef*	Coefficient for calculation of the lower canopy roughness (dimensionless)
42	*alogu_coef*	Coefficient for calculation of the upper canopy roughness (dimensionless)
43	*avmuir_coef*	Coefficient for calculating the canopy emissivity (dimensionless)

For our calibration procedure two objective are associated to the problem

$$F_1(\mathbf{w}) = \sum_{j=1}^{9} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y_i^{\text{Obs},(j)} - Y_i^{\text{Mod},(j)}(\mathbf{w}) \right)^2} \quad (1)$$

$$F_2(\mathbf{w}) = \sum_{j=1}^{9} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{Y_i^{\text{Obs},(j)}} - \frac{1}{Y_i^{\text{Mod},(j)}(\mathbf{w})} \right)^2} \quad (2)$$

The vector $\mathbf{w} = [w_1 \; w_2 \; ... \; w_P]^T$ is the unknown vector to be calculated, and the entries w_k ($k = 1, ..., P$) represent each parameter described in Table 1. For our study the number of unkowns w_k is $P=43$. These objective funtions were used to calibrate a hydrologic model (Araújo et al., 2013).

2.2 Morris' method

The sensitivity analysis (SA) is the core for Morris's analysis. SA is the study of the relationship between the impact of input changes on the model output. There are local and global SAs. The Morri's method for sensitivity analysis is a global screening method. It is suitable for dealing with models with a large number of parameters. The method proposes to identify which model parameters are: (a) negligible, (b) linear and additive, or (c) non-linear or involved in interactions with other parameters.

In this methodology, a small change is added to one parameter. Each parameter can assume a discrete number of values (levels), and the values are chosen within the range of parameter variation. Two sensitivity measures are proposed by Morris: the μ estimates the average effect of a parameter on the model output, and the σ measuring, through the standard deviation, the second-order and larger orders effects.

A box for experimentation Ω is defined on the search space of dimension P. Each direction on box Ω is discretized, and the sensitivy analysis for each parameter w_k will be performed inside of the box associated to the appropriated direction. The *elementar effect* to the k-th parameter is defined as:

$$d_k(\mathbf{w}) = \frac{Y(w_1,...,w_k + \Delta, w_{k+1},...,w_P) - Y(\mathbf{w})}{\Delta} \quad (3)$$

where Δ is a small free parameter. For each parameter, r realizations are computed. From the mentioned realizations, the average μ_k and the standard deviation σ_k are calculated. The μ_k measurement is used to detect whether the k-th parameter has any overall effect on the model output, and the measure σ_k is employed to detect whether the k-th parameter has nonlinear effects and/or is involved in interactions with other parameters.

Morris has developed a methodology to perform the r realizations inside the Ω box, where a random orientation matrix B is derided. The matrix B is used to define the realization to be executed (Araújo, 2014). Figure 2 illustrates trajectories on search spaces of dimensions 2 and 3.
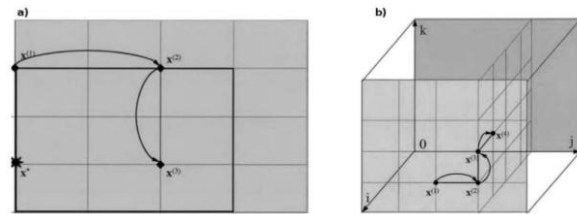

Figure 2: Two trajectories in the search spaces: (a) dimension=2, (b) dimension=3.

3.3 NSGA-II Algorithm

The NSGA-II (Elitist Non-dominated Sorted Genetic Algorithm II) is a multi-objective genetic algorithm proposed by Deb et al. (2000), and an improved version of NSGA (Non-dominated Sorted Genetic Algorithm) was dereived (Srinivas and Deb, 1994).

Genetic algorithms need a population for candidate solution. Each candidate solution is evaluated by using the fitness function (Eqs. (1) and (2), for the present paper).

In the NSGA-II, the population with M individuals is initialized of an usual way. The population is separated into clusters (fronts) based on non-domination into each front. The first front is fully non-dominant set, the second front is dominated by the the first front only, and layeres of fronts are strutured. Each individual in the each front is indexed with rank (fitness) values or based on front where they are. Individuals in first front are given a fitness value of 1 and individuals in second are assigned fitness value as 2, and so on.

A homegeneous distribution among the candidate solutions is searching with application of a second rank-parameter named crowding distance. This second parameter is calculated for each individual. The crowding distance is a measure of how close an individual is to its neighbors. Large average crowding distance implys in a better population diversity.

The two parameters, rank and crowding distance, are employed to select the parents in the population applying a tournament selection. The selected population generates offsprings from crossover and mutation operators. The current population and the offsprings are sorted one more time, looking for non-domination. The best *M* individuals are selected (elitism).

4 Numerical Results

Varejão et al. (2013) proposed a calibration based on a temporal hierarchy of simulated processes by IBIS, starting with faster processes moving to the slower ones. In Figure 1, typical time-scales for the processes are indicated. As mentioned in Session 2, there are 9 processes modeled. The time-scale and sensitivy analysis are used to estabilish the groups of processes and parameters associated.

With prior knowledge of the relationship between the model parameters and the simulated processes, and using the Morris' approach for the sensitivity analysis, the parameters can be grouped into 5 levels:, as following:

(a) L1: radiative flows: PAR_0 and f_{APAR} (3 parameters);

(b) L2: surface radiation balance: R_n (3 parameters);

(c) L3 - Turbulence: u_* (3 parameters);

(d) L4: turbulent flows: NEE, HE, LE (16 parameters);

(e) L5: Carbon Allocation: LAI (6 parameters).

Calibration starts with parameter estimation of parameters at level Lj (the parameter of other levels remain frozen). The NSGA-II is applied to determined the parameter. After that, the estimation of parameter of level Lj+1 is started, and so on.

Varejão et al. (2013) have used data obtinaed from the *Flona Tapajós*. This site is located near the 67 km of Santarém-Cuiabá highway (2.51S, 54.58W). The period for data aquisition was from January-2002 to December-2004. The parameter set estimated by Varejão et al. (2013) was used to generate a synthetic observations executing the IBIS model. A white Gaussian noise with 5% of level was added to the synthetic observation.

The synthetic observation were applied for testing the calibration methodology presented here. The Pareto set for the objetive functions (1) and (2) was determined employing the NSGA-II. As mentioned before, the hierarchical strategy with 5 levels was adopted here. Only the final result is presented, and the estimated values are shown in Table 3.

Table 3: Example table 1

Parameter	Exact	Estimated
rhoveg−vis	0.0872	0.0778
tauveg−vis	0.0498	0.0498
chifuz	−0.2249	−0.1643
rhoveg−NIR	0.2966	0.3468
tauveg−NIR	0.2038	0.2235
avmuir−coef	370.9740	366.4872
dispu−coef	0.9779	0.9895
alogl−coef	3.9289	4.4682
alogu−coef	7.2151	6.4447
vmax−pft	0.0001	0.0001
coefmub	7.5494	7.6680
chs	33448.5189	22679.8452
beta2	0.7838	0.8081
funca−coef	5477.4147	6678.2708
funcbcoef	5900.2127	5306.0478
root−coef	0.8429	1.0703
rwood−coef	0.1563	0.0857
tempvm−coef	3961.8800	3502.3284
stressf−coef	−5.3121	−5.1114
clitls−coef	1.4686	2.1727
clitrs−coef	4.5124	4.9226
clitws−coef	1.0452	1.0452
csoislon−coef	0.2502	0.3429
csoislop−coef	6.3736	6.6974
kfactor	1.4744	1.3810
rgrowth−coef	0.2965	0.2738
tauleaf	0.5808	0.7208
specla	31.2950	32.9062
aleaf	0.2654	0.1856
aroot	0.3269	0.1870
awood	0.6280	0.4773

From the estimated parameter for the IBIS model, a comparison against model output and observation is carried out. IBIS was executed with the estimaed parameters in Table 3. Figure 3 shows one day of simulation for the reflected photosynthetically active radiation (PAR_0): a good agreement between simulation and observation is verified.
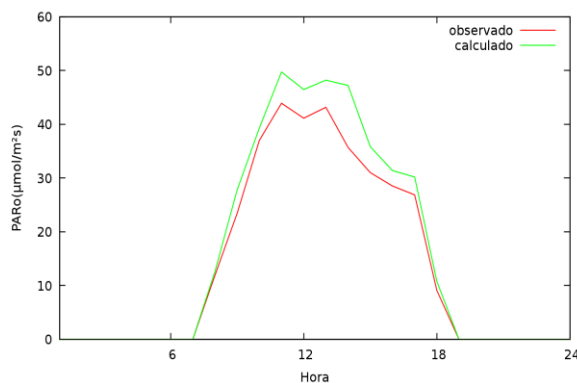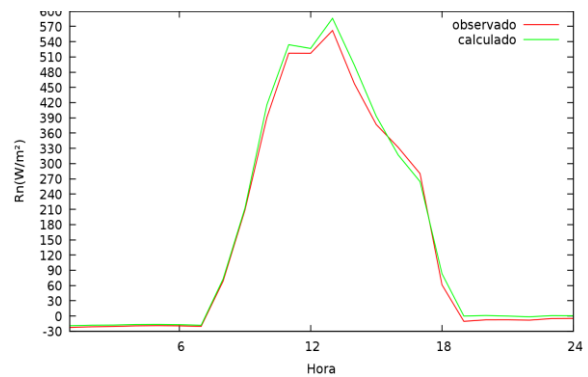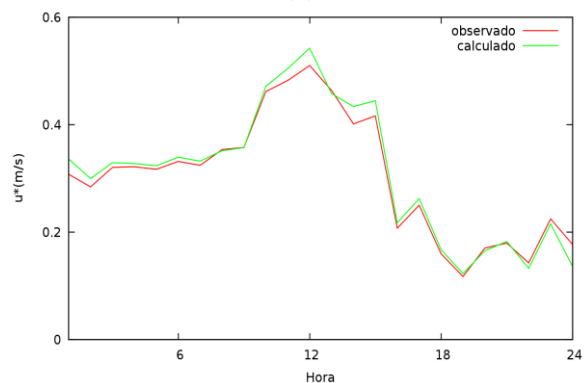

(a)

Figure 3: Hierarchical calibration: PAR_0 for one day simulation. Red: observation. Green: simulated.

(b)

Figure 4: Hierarchical calibration, simulation for one day: (a) R_n, (b) u_*. Red observation. Green simulated.

Figure 4 displays the comparison between observations and simulation with IBIS for the R_n, and u_*. A good agreement is noted again. Figure 5 provides graphs of calculated and observed for a one year to the Leaf Area Index (LAI). The simulation follows the dynamics shown in the observations. Other variables are not shown (*NEE*, *HE*, *LE*), but all variables presented good representation for the IBIS dynamics.
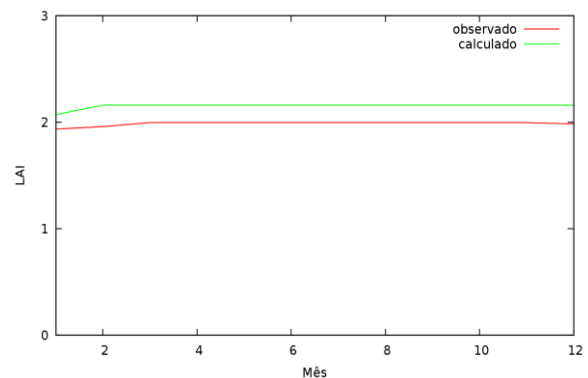

Figure 5: Hierarchical calibration, simulation for one year: *LAI*. Red observation. Green simulated.

4 Conclusion and Final Remarks

A multiobjective approach was employed to calibrate the Integrated Biosphere Simulator (IBIS) model. The objective functions differ from those used by Varejao et al. (2013). But, we also applied the NSGA-II as an optimization method to calculate the Pareto set.

The hierarchical strategy based on the time-scale of the processes of the ecosystem under

study and the The Morris' method were effective to select the five groups with different levels of impacto n the IBIS model.

Calibration is an importante class of inverse problems. However, no regularization was used.

Acknowledgement

Referências

ARAÚJO, A. S., CAMPOS VELHO, H. F., GOMES, V. C. F. (2013): Calibrating an hydrological model by an evolutionary strategy for multi-objective optimization. Inverse Problems in Science & Engineering, 21, 438-450.

ARAÚJO, A. S. (2014): Multi-objective calibration of hydrologic and atmospheric surface models. Ph.D. thesis on Applied Computing. São José dos Campos (SP), Brazil (in Portuguese).

Beck, J. V., Blackwell, B., Clair Jr., C. R. St. (1985): Inverse Heat Conduction: Ill-Posed Problems. Wiley-Interscience.

COELLO, C. A., LAMONT, G. B., VAN VELDHUISEN, D. A. (2007). Evolutionary Algorithms for Solving Multi-Objective Problems. Springer.

HWANG, C-L, MASUD, A. S. M. (1979): Multiple objective decision making, methods and applications: a state-of-the-art survey. Springer-Verlag.

DEB, K., PRATAP, A., AGARWAL, S., MEYARIVAN, T. (2000): A fast elitist multi-objective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6, 182–197.

DEB, K. (2009): Multi-objective optimization using evolutionary algorithms. John Wiley & Sons.

FOLEY, J. PRENTICE, I. C., RAMANKUTTY, N., LEVIS, S., POLLARD, D., SITCH, S., HAXELTINE, A. (1996): An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. Global Biogeochemical Cycles, 10(4), 603–628.

KUCHARIK, C. J., FOLEY, J. A., DELIRE, C., FISCHER, V. A., COE, M. T., LENTERS, J. D., YOUNG-MOLLING, C., RAMANKUTTA, N., NORMAN, J. M., GOWER, S. T. (2000): Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure. Global Biogeochem. Cycles, 14(3), 795–825.

MIETTINEN, K. (1999). Nonlinear Multiobjective Optimization. Springer.

MINJIAO, L., XIAO, L. (2014): Time scale dependent sensitivies of the XinAnJiang model parameters. Hydrological Research Letters, 8(1), 51–56.

MORRIS, M. D. (1991): Factorial sampling plans for preliminary computational experiments. Technometrics (American Society for Quality Control and American Statistical Association, 33(2), 161–174.

SRINIVAS, N., DEB, K. (1994): Multi-objective optimization using non-dominated sorting in genetic algorithms. Evolutionary Computation, 2(3), 221–248.

VAREJÃO, C. M., COSTA, M. H., CAMARGOS, C. C. S. (2013): A multi-objective hierarchical calibration procedure for land surface-ecosystem models. Inverse Problems in Science and Engineering, 21(3), 357–386.