

CARLA CRISTINA DE SOUZA CAMARGOS

**DESEMPENHO DE UM ALGORITMO DE OTIMIZAÇÃO HIERÁRQUICO  
MULTIOBJETIVO APLICADO A UM MODELO DE SUPERFÍCIE  
TERRESTRE E ECOSISTEMAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Meteorologia Agrícola, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

C172d  
2013

Camargos, Carla Cristina de Souza, 1985-

Desempenho de um algoritmo de otimização hierárquico multiobjetivo aplicado a um modelo de superfície terrestre e ecossistemas / Carla Cristina de Souza Camargos. – Viçosa, MG, 2013.

xvii, 48 f. : il. ; 29 cm.

Texto em inglês.

Orientador: Marcos Heil Costa.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 44 - 48.

1. Meteorologia - Modelos matemáticos. 2. Calibração.  
3. Otimização matemática. 4. Ecossistemas. 5. Florestas tropicais - Amazônia. I. Universidade Federal de Viçosa. Departamento de Engenharia Agrícola. Programa de Pós-Graduação em Meteorologia Agrícola. II. Título.

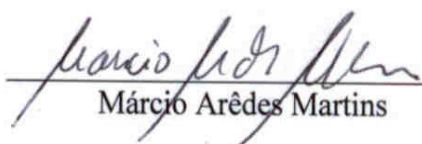
CDD 22. ed. 630.2515

CARLA CRISTINA DE SOUZA CAMARGOS


**DESEMPENHO DE UM ALGORITMO DE OTIMIZAÇÃO HIERÁRQUICO  
MULTIOBJETIVO APLICADO A UM MODELO DE SUPERFÍCIE  
TERRESTRE E ECOSISTEMAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Meteorologia Agrícola, para obtenção do título de *Magister Scientiae*.

APROVADA: 20 de março de 2013

  
Márcio Arêdes Martins

  
Gilberto Chohaku Sedyama

  
Marcos Heil Costa  
(Orientador)

*Ao meu anjo protetor e mãe: Angela*

*“Come to the edge.  
We might fall.  
Come to the edge.  
It's too high!  
COME TO THE EDGE!  
And they came,  
And he pushed,  
And they flew“*

*Christopher Logue*

*“As botas apertadas são uma das maiores venturas da terra, porque, fazendo doer os pés, dão azo ao prazer de as descalçar”*

*Machado de Assis*

## AGRADECIMENTOS

À Universidade Federal de Viçosa (UFV) e ao Departamento de Engenharia Agrícola (DEA), pela oportunidade de realizar o curso de pós-graduação em Meteorologia Agrícola.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao meu amigo Victor Hugo Benezoli, por abrir as portas, me apoiar e auxiliar.

Ao meu orientador, Prof. Marcos Heil Costa, pela oportunidade de desenvolver este projeto, pela paciência, pelo conhecimento compartilhado e pela grande amizade.

À minha família, que é a minha base para todas as decisões de minha vida. E à minha segunda família, as D.I.V.A.S, que são amigas e companheiras e tornam a distância física dos meus consanguíneos um fato mais suave.

Ao Claudeci Varejão-Jr., que me esclareceu vários pontos deste trabalho e sempre foi muito atencioso.

Ao Fabrício Murta, por me salvar diversas vezes quando me perdi no código do programa, sempre de forma prestativa e divertida.

Aos meus amigos e companheiros do DEA, em especial aos do Grupo de Pesquisas em Interação Biosfera-Atmosfera: Patrícia, Gabrielle, Lívia, Francisca, Matheus, Letícia, Luciana.

A todos os meus amigos de Viçosa, Carmo do Cajuru e Divinópolis que sempre tornarem meus dias prazerosos e divertidos.

## BIOGRAFIA

CARLA CRISTINA DE SOUZA CAMARGOS, filha de Angela Maria de Souza Camargos (*in memorian*) e José Carlos Camargos, nasceu em 04 de novembro de 1985 na cidade de Carmo do Cajuru.

Em março de 2004, ingressou no curso de Matemática pela Universidade Federal de Viçosa (UFV).

Em setembro de 2008, estagiou no departamento de Tecnomatemática da Universidade de Bremen (Alemanha) por 3 meses.

Em agosto de 2010 iniciou o curso de Mestrado no Programa de Pós-Graduação em Meteorologia Agrícola pela Universidade Federal de Viçosa (UFV).

## SUMÁRIO

LISTA DE FIGURAS .....	ix
LISTA DE TABELAS .....	xi
LISTA DE SÍMBOLOS .....	xii
RESUMO .....	xiv
ABSTRACT .....	xvii
1. INTRODUCTION .....	1
2. LITERATURE REVIEW .....	5
<b>2.1 Sensitivity analysis</b> .....	5
<b>2.2 Model evaluation</b> .....	6
<b>2.3 Multi-objective optimization</b> .....	7
<b>2.4 Genetic algorithms</b> .....	7
<b>2.5 Hierarchical optimization</b> .....	8
3. METHODOLOGY .....	9
<b>3.1 IBIS model</b> .....	9
<b>3.2 Site description and observed data</b> .....	10
<b>3.3 Data gap-filling and data filtering</b> .....	10
<b>3.4 OPTIS</b> .....	12
<b>3.5 IBIS' sensitivity analysis</b> .....	12
<b>3.6 Mono-objective calibration</b> .....	13
<b>3.7 Multi-objective hierarchical calibration</b> .....	14
<b>3.7.1 Comparison between simultaneous and hierarchical calibration</b> .....	15
<b>3.7.2 Hierarchical calibration</b> .....	16
4. RESULTS AND DISCUSSION .....	18
<b>4.1 IBIS' sensitivity analysis</b> .....	18
<b>4.2 Mono-objective calibration</b> .....	21



<b>4.3</b>	<b>Multi-objective calibration</b> .....	27
<b>4.3.1</b>	<b>Comparison between simultaneous and hierarchical calibration</b> .....	27
<b>4.3.2</b>	<b>Results for the hierarchical calibration</b> .....	30
<b>4.4</b>	<b>Discussion on the performance index</b> .....	37
<b>5.</b>	<b>CONCLUSIONS</b> .....	40
	<b>REFERENCES</b> .....	44

## LISTA DE FIGURAS

Figure 1: OPTIS structure.

Figure 2: PAR<sub>o</sub> results after the mono-objective calibration: a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 3: fAPAR results after the mono-objective calibration a) Dispersion plot and b) seasonal variation using all 43 parameters and c) Dispersion plot and d) seasonal variation using only sensitive parameters.

Figure 4: R<sub>n</sub> results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 5: u\* results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 6: H results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 7: LE results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 8: NEE results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 9: Index D for radiative flux output variables versus the number of optimized outputs (n), separated by the type of optimization (simultaneous or hierarchical).

Figure 10: Index D for turbulent flux output variables versus the number of optimized outputs (n), separated by the type of optimization (simultaneous or hierarchical).

Figure 11: PARo results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 12: fAPAR results after the multi-objective calibration a) Dispersion plot and b) seasonal variation.

Figure 13: Rn results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 14:  $u^*$  results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 15: H results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 16: LE results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 17: NEE results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 18: Index D for radiative and turbulent flux output variables versus the number of optimized outputs (n), separated by the type of optimization (simultaneous or hierarchical).

## LISTA DE TABELAS

Table 1: Potential parameters for the calibration of IBIS.

Table 2: List of optimization experiments considering the group of variables related to radiative fluxes.

Table 3: List of optimization experiments considering the group of variables related to turbulent fluxes.

Table 4: Degree of sensitivity ( $\mu^* + \sigma$ ) of all parameters for each output variables according to the Morris method.

Table 5: Parameters optimized in the calibration experiments related to radiative flux variables. Each pair of bracket represents one hierarchical level. Some parameters are not shown in the table because they are not used in any calibration below.

Table 6: MAE results for the mono-objective calibrations.  $f^{\text{mono,overfit}}$  is the simulation using all 43 available parameters and  $f^{\text{mono,SA}}$  is the simulation using only the parameters selected according to the sensitivity analysis validation

Table 7: Calibration results using only radiative flux output variables. The bracket is used to identify the hierarchical level,  $h$  is the number of hierarchical levels and  $n$  is the output quantity.

Table 8: Calibration results using only turbulent flux output variables. The bracket is used to identify the hierarchical level,  $h$  is the number of hierarchical levels and  $n$  is the output quantity.

Table 9: Calibration results using radiative and turbulent flux output variables. The bracket is used to identify the hierarchical level,  $h$  is the number of hierarchical levels and  $n$  is the output quantity.

## LISTA DE SÍMBOLOS

$a_L$	Carbon allocation fraction to leaves (dimensionless)
$a_R$	Carbon allocation fraction to fine roots (dimensionless)
$a_W$	Carbon allocation fraction to wood (dimensionless)
$\beta_2$	Parameter of fine roots distribution (Jackson rooting profile) (dimensionless)
$C_L$	Heat capacity of lower canopy leaves and stems ( $J kg^{-1}m^{-2}$ )
$c_{ll}^{lit}$	Carbon in leaf litter pool (lignin) ( $kg C m^{-2}$ )
$c_{lm}^{lit}$	Carbon in leaf litter pool (metabolic) ( $kg C m^{-2}$ )
$c_{ls}^{lit}$	Carbon in leaf litter pool (structural) ( $kg C m^{-2}$ )
$c_{rl}^{lit}$	Carbon in fine root litter pool (lignin) ( $kg C m^{-2}$ )
$c_{rm}^{lit}$	Carbon in fine root litter pool (metabolic) ( $kg C m^{-2}$ )
$c_{rs}^{lit}$	Carbon in fine root litter pool (structural) ( $kg C m^{-2}$ )
$c_{wl}^{lit}$	Carbon in woody litter pool (lignin) ( $kg C m^{-2}$ )
$c_{wm}^{lit}$	Carbon in woody litter pool (metabolic) ( $kg C m^{-2}$ )
$c_{ws}^{lit}$	Carbon in woody litter pool (structural) ( $kg C m^{-2}$ )
$C_S$	Heat capacity of stems on upper canopy ( $J kg^{-1}m^{-2}$ )
$c_{pas}^{soi}$	Carbon in soil (passive humus) ( $kg C m^{-2}$ )
$c_{slon}^{soi}$	Carbon in soil (slow unprotected humus) ( $kg C m^{-2}$ )
$c_{slop}^{soi}$	Carbon in soil (slow protected humus) ( $kg C m^{-2}$ )
$C_U$	Heat capacity of upper leaves ( $J kg^{-1}m^{-2}$ )
$d$	Zero-plane displacement height for upper canopy (m)
$D$	Relative performance index

EA	Evolutionary algorithm
fa	Temperature function for aboveground biomass (stems) (dimensionless)
fAPAR	Fraction of absorbed photosynthetically active radiation
fb	Temperature function for belowground biomass (roots) (dimensionless)
GA	Genetic algorithm
H	Sensible heat flux
IBIS	Integrated biosphere simulator
k	Multiplication factor of decay constant for carbon pools (dimensionless)
LE	Latent heat flux
log(Zol)	Logarithm of roughness length of lower canopy (dimensionless)
log(Zou)	Logarithm of roughness length of upper canopy (dimensionless)
LSEM	Land surface and ecosystem models
m	Coefficient for stomatal conductance ( dimensionless)
	Average diffuse optical depth (dimensionless)
$\mu$	Mean effect of the model output parameters
$\mu^*$	Absolute value of $\mu$
MAE	Mean absolute error
MBE	Mean bias error
ms	Factor of the soil moisture stress (dimensionless)
MSE	Mean square error
n	The number of optimized objective functions
NEE	Net ecosystem exchange
NSGA-II	Nondominated sorted genetic algorithm II

OAT	One-step-at-a-time method
$R_g$	Growth respiration coefficient (fraction) (dimensionless)
$R_{mr}$	Maintenance respiration coefficient for root ( $s^{-1}$ )
$R_{mw}$	Maintenance respiration coefficient for wood ( $s^{-1}$ )
RMSE	Root mean square error
$R_n$	Net radiation
PAR <sub>o</sub>	Outgoing photosynthetically active radiation
$\rho_{NIR}$	Leaf reflectance on upper canopy (NIR) (dimensionless)
$\rho_{VIS}$	Leaf reflectance on upper canopy (visible) (dimensionless)
AS	Sensitivity analysis
SLA	Specific leaf area ( $m^2 kg^{-1}$ )
$t_v$	V <sub>max</sub> thermal stress (dimensionless)
$\tau_L$	Foliage biomass turnover time constant (years)
$\tau_{NIR}$	Leaf transmittance on upper canopy (NIR) (dimensionless)
$\tau_R$	Fine root biomass turnover time constant (years)
$\tau_{VIS}$	Leaf transmittance on upper canopy (visible) (dimensionless)
$\tau_W$	Wood biomass turnover time constant (years)
$u^*$	Friction velocity ( $m s^{-1}$ )
$V_{max}$	Maximum capacity of enzyme Rubisco at 15 °C on upper canopy ( $mol[CO_2] m^{-2} s^{-1}$ )
$\chi_u$	Leaf orientation factor on upper canopy (-1 vertical 0 random 1 horizontal)
w <sub>soi</sub>	Initial soil moisture ( $m^3 m^{-3}$ )

CAMARGOS, Carla Cristina de Souza, M. Sc., Universidade Federal de Viçosa, março de 2013. **Desempenho de um algoritmo de otimização hierárquico multiobjetivo aplicado a um modelo de superfície terrestre e ecossistemas.** Orientador: Marcos Heil Costa.

O desempenho de um LSEM (Modelo de superfície terrestre e ecossistema) depende dos parâmetros das equações que representam os processos simulados. Contudo, a mensuração de alguns destes parâmetros pode ser impraticável ou até mesmo impossível; por isso, necessitam ser estimados ou, preferencialmente, otimizados para cada ecossistema. Quando os parâmetros são calibrados para uma única variável (problema mono-objetivo) eles podem não representar bem a realidade, dado a complexidade do modelo e sua dependência de diversas variáveis (problema multiobjetivo). Por isso, há a necessidade de uma otimização simultânea multiobjetiva. Porém, o desempenho da otimização diminui com o aumento do número de variáveis otimizadas simultaneamente e, além disso, o estudo da otimização simultânea de mais de três objetivos é uma área relativamente nova e não suficientemente estudada. Para a otimização simultânea de um grande número de variáveis, existe uma metodologia na qual se utiliza conceitos de teoria hierárquica de sistemas em que a otimização ocorre dos processos mais rápidos (fluxos radiativos) para os mais lentos (alocação de carbono). Este trabalho avalia o desempenho da otimização hierárquica do modelo, utilizando o índice  $D$  (a média das razões individuais entre as saídas das otimizações multiobjetiva e mono-objetiva). Entender como o índice de desempenho  $D$  do algoritmo de otimização hierárquico varia em relação ao número de funções objetivo otimizadas é de extrema importância para o desenvolvimento desta área de pesquisa. Para fazer atingir os objetivos, foram necessárias duas etapas. Primeiramente, foi feita uma análise de sensibilidade, a fim de conhecer a sensibilidade das variáveis de saída aos parâmetros do modelo. Depois, foram feitas simulações com todas as combinações possíveis entre as sete variáveis micrometeorológicas disponíveis ( $PAR_o$ ,  $fAPAR$ ,  $R_n$ ,  $u^*$ ,  $H$ ,  $LE$ ,  $NEE$ ) levando em consideração a hierarquia dos processos. Os resultados encontrados indicam que, para até três funções objetivo, a otimização multiobjetiva hierárquica pode gerar resultados melhores do que a otimização multiobjetiva tradicional (um único nível hierárquico), desde que a distribuição dos parâmetros entre as variáveis seja feita de forma coerente com a análise de sensibilidade. Outro resultado importante revela que para um mesmo número de saídas otimizadas,



quanto maior o número de níveis hierárquicos melhor o desempenho do modelo otimizado. Porém, o desempenho do modelo diminui rapidamente quando o número de funções objetivo aumenta, evidenciando que o poder da calibração hierárquica para o uso de um grande número de funções objetivo é altamente dependente de algumas restrições que o modelo possui e um alto desempenho do modelo para muitas funções objetivo será possível somente após a remoção delas.

## ABSTRACT

CAMARGOS, Carla Cristina de Souza, M. Sc., Universidade Federal de Viçosa, March, 2013. **Performance of a hierarchical multi-objective optimization algorithm applied to a land surface and ecosystem model.** Adviser: Marcos Heil Costa.

The performance of LSEMs (Land surface and ecosystem models) depends on the parameters of the equations representing the simulated process. However, the measurement of some parameters can be impractical or even impossible; therefore, they need to be estimated, or preferably optimized specifically for each ecosystem. When the parameters are calibrated to a single variable (mono-objective problem) they may not represent the reality, because the complexity of the model and its dependence on several variables (multi-objective problem). Thus, simultaneous multi-objective optimizations are indispensable. However, the optimization performance decreases as the number of variables to be optimized simultaneously increases. Furthermore, the study of simultaneous optimization using more than three objectives is a new area and not yet sufficiently studied. For simultaneous optimization of a large number of variables, there is a method that uses concepts of hierarchical systems theory in which the optimization occurs from the fastest (radiative fluxes) to the slowest process (carbon allocation). This study evaluates the performance of the hierarchical optimization using the index D (the average of the ratios between the individual outputs of multi-objective optimization and mono-objective). Understanding how the performance index D varies with respect to the number of objective functions optimized and to the number of hierarchical levels is important for the development of this research area. Two steps are necessary to achieve the study goals. First, a sensitivity analysis was performed to determine the output variables sensitivity to the model parameters. After, simulations were made using all possible combinations among the seven micrometeorological variables available (PAR<sub>o</sub>, fAPAR, R<sub>n</sub>, u \*, H, LE, NEE) taking into account the hierarchy of processes. The results indicate that for up to three objective functions, hierarchical multi-objective optimization generates better results than the simultaneous multi-objective optimization (one hierarchical level), provided that the parameters distribution among hierarchical levels is consistent with the sensitivity analysis.

Another important result shows that for the same number of outputs optimized, the greater the number of hierarchical levels the better the performance of the optimized model. However, the model performance falls quickly as the number of objective functions increases, evidencing that the power of hierarchy calibration that use a high number of objective functions is highly dependent on the removal of some constraints for model's performance.

## 1. INTRODUCTION

Land surface and ecosystem models (LSEM) mechanistically describe ecosystem function in terms of biogeophysical processes (such as fluxes of sensible and latent heat, momentum and radiation), biogeochemical processes (terrestrial carbon cycle and CO<sub>2</sub> fluxes) and ecological processes (phenology and competition among plant functional types) at the biosphere-atmosphere interface. Land surface processes influence the climate system through their control of energy, water and carbon balances. Thus integrated LSEMs are important tools not only for biosphere-atmosphere interactions at long-term scales such as paleoclimates and future climates, but also for short-term applications that depend on land surface fluxes, like climate prediction and weather forecast.

In the past fifteen years, a better understanding of terrestrial ecosystem functioning and computational advances enabled the incorporation of new processes to the LSEMs, increasing model complexity. Each modeled process requires parameters, which define the performance of the model [Groenendijk et al., 2011]; however, some parameters may be impractical or impossible to measure [Mackay et al., 2003], and must be estimated or, preferably, specifically calibrated for each ecosystem.

Parameter calibration may be carried out using an optimization process that minimizes differences between predicted and observed data. Hence the increase in the number of calibrated parameters usually leads to a more reliable model. In addition to improving its performance, the calibration also allows for a detailed study of each of its components, so that the models can be improved according to necessity and resources [Pitman, 2003].

The performance of models with multiple outputs such as LSEMs could be optimized through the use of all available observations, including data collected by multiple sensors in intensive field sites and regional observations by remote sensors (radar, aircraft, satellites, etc). However, this improvement can increase the complexity of the optimization problem.

Each output variable simulated by the model (that has a corresponding observation) represents one objective of the optimization problem. There are two kinds of optimization: mono-objective and multi-objective. A mono-objective optimization has a unique solution to the problem, whereas in multi-objective problems, conflicting objectives are common, i.e., the optimized set of parameters for one objective function is not suitable for another. Thus a multi-objective problem may have a set of solutions instead of a unique solution to the problem. Hence there is a performance drop in simultaneous optimization methods with the increase of optimized objective functions in the same simulation [Deb et al., 2002; Veldhuizen and Lamont, 2000; Vrugt et al., 2003].

The simultaneous optimization of two or three objectives has been widely investigated [Gupta and Sorooshian, 1998; Deb et al., 2002; Coello, 2006] resulting in a variety of very efficient algorithms. The simultaneous optimization of more than three objectives, however, is a relatively young field and has not been thoroughly studied yet [Praditwong and Yao, 2007; Schutze et al., 2011].

Varejão-Jr. et al. [2012] developed a calibration methodology that allows for the calibration of all processes simulated by a LSEM. This methodology proposes a hierarchical approach, in which the processes simulated by the model are divided in groups (i.e., radiative fluxes, turbulent fluxes or carbon allocation) and then hierarchically calibrated according to the organization of the groups, beginning with fast processes (radiative fluxes), then slow processes (mass fluxes) and finally the

slowest processes (carbon allocation). The concept is that a large number of objective functions can be optimized as long as (i) they are organized in hierarchical levels, which must contain no more than three objective functions of the same group to be simultaneously optimized; (ii) each model parameter is optimized only once (in one hierarchical level) during the optimization process. When the processes (objective functions) are separated in hierarchical levels, parameters that describe the processes should be distributed and optimized in only one hierarchical level. The selection of the parameters each process is most sensitive to is done through a sensitivity analysis. The hierarchical optimization technique was implemented in an algorithm (OPTIS), which was applied to the LSEM IBIS – Integrated Biosphere Simulator [Foley et al., 1996].

Assuming the mono-objective calibration is the best possible calibration for each individual output, the performance of the model after the multi-objective calibration can be compared to the performance of the mono-objective calibration of each independent variable by using a relative performance index (D) [Varejão-Jr. et al., 2012]. The index D is defined as the average of the individual ratios between the objective function output of the mono-objective calibration and the objective function output of the multi-objective calibration. The multi-objective hierarchical calibration efficiency is proportional to the value of the index D.

The way the performance of the optimization algorithm varies according to the number of optimized objective functions (n) is crucial to the optimization of LSEMs using intensive sites data. If the index D drops rapidly with number of objective functions then the optimization attempt may be useless, the LSEM calibration using all data collected at intensive sites is impossible and only the model validation is possible.

The behavior of the performance index D is not well known in the literature. In the only test performed, Varejão-Jr. et al. [2012] obtained a small drop of index D associated to the increase of the number of objective functions, with the index D ranging from 0.925 (n = 2) to 0.801 (n = 9), representing a satisfactory result for the multi-objective calibration problem. However, neither slope and concavity of the curve D versus n nor the variation of these particularities according to the way the hierarchical levels are organized have been studied yet. This information is essential

for determining the feasibility of the multi-objective optimization of LSEMs, in order to aid the design of model-improvement oriented intensive experimental sites and define future steps in land-surface/ecosystem modeling.

This study has two main objectives related to a better understanding of the advantages of the hierarchical calibration. First, compare the hierarchical calibration to the simultaneous technique. Second, evaluate the basic characteristics (slope and concavity) of the relationship between the index D and the number of objective functions.

## **2. LITERATURE REVIEW**

### **2.1 Sensitivity analysis**

Sensitivity Analysis (SA) is an essential tool in modeling that helps to understand the importance of each parameter on the model output. An optimization problem can be very complex when it involves many parameters and multiple objective functions. Through SA, it is possible to understand how the variability in the model output may be attributed to the variability in the input [Rosolem et al., 2012]; and thus eliminate non-influential parameters in the optimization of some outputs, reducing the dimensions of the parameters space (screening method). Therefore, the search process becomes easier and the optimization results are improved.

A special case of screening sensitivity analysis is the Morris method (1991) for global sensitivity analysis, which is a one-step-at-a-time method (OAT), i.e., only one input factor is changed in each run and all other factors are kept fixed to their baseline value. The Morris method provides qualitative sensitivity analysis results, which, in addition to identifying non-influential parameters, enables the ranking of



the parameters in order of importance. However, such method does not quantify exactly the relative importance of each one.

Morris (1991) proposed the use of two different measures: one ( $\mu$ ) to estimate the mean effect of the model output parameters and another ( $\sigma$ ) to estimate second and higher order effects (nonlinear effects and interactions). In order to avoid erroneous interpretations in non-monotonic models, Saltelli et al. (2005) proposed the use of the absolute value of the first measure proposed by Morris ( $|\mu| = \mu^*$ ).

The sensitivity results are never unique, because the SA algorithm uses a random value to begin the test. Furthermore, in nonlinear models, OAT could fail to account for the interactions among different input factors [Czitrom, 1999]. Thus multiple analyses are necessary to confirm the results, as described in Section 3.5.

## **2.2 Model evaluation**

The model evaluation makes use of an objective function to yield numerical values which quantify how the predicted values fit the observed data. The choice of the adjustment metrics between prediction and observation for model calibration is even more important than the selection of the optimization method itself [Trudinger et al., 2007]. Several studies discuss the choice of adjustment metrics to evaluate accuracy and effectiveness of models [Fox, 1980; Willmott, 1981, 1982; Willmott et al., 1985; Legates and McCabe Jr., 1999; Willmott and Matsuura, 2005].

Mean error measures such as the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean bias error (MBE), provide appropriate and comprehensive information about how the model fits the observed data. Among these measures, RMSE and MAE are notably interesting because they exhibit the model error in the same unit of the simulated variable. Willmott [1982] recommended that the evaluation model procedure uses at least one of these two metrics, but later Willmott and Matsuura [2005] recommended using MAE rather than RMSE, because the former provides an easier interpretation, as it only represents the mean error, without the ambiguity involved in the calculation of the RMSE.

### **2.3 Multi-objective optimization**

The multi-objective optimization deals with solving problems that have many objective functions. This technique finds a set of parameters which minimize or maximize simultaneously all the objective functions, generally satisfying some constraints. In multi-objective optimization problems, conflicting objectives require a set of optimal solutions (largely known as the Pareto set), instead of a single optimal solution, which is typically the case in mono-objective optimization problems [Veldhuizen and Lamont, 2000; Deb et al., 2002; Vrugt et al., 2003; Fu et al., 2005]. When an optimal solution cannot be replaced by one which improves an objective without worsening another, it is called a nondominated solution or Pareto-optimal solution. The plot of the objective functions whose nondominated solutions are in the Pareto set is called the Pareto frontier [Coello, 2006]. One of these Pareto-optimal solutions cannot be considered better than other without additional information, demanding the definition of as many Pareto-optimal solutions as possible [Deb et al., 2002].

### **2.4 Genetic algorithms**

Genetic algorithms (GAs) belong to the class of evolutionary algorithm (EAs), which generate solutions to optimization problems using techniques that mimic natural evolution (according to Darwin's theory of evolution) and simulate the natural selection process, by selecting the set of parameters that produces the fittest output to the data. Traditional GAs are customized to accommodate multi-objective problems by using specialized fitness functions and introducing methods to promote solution diversity [Konak et al., 2006]. EAs are suitable to solve multi-objective optimization problems because these algorithms deal simultaneously with a set of possible solutions (called population), which enables the computation of multiple Pareto-optimal solutions in a single algorithm run. Some multi-objective optimization algorithms are susceptible to the shape of the Pareto frontier and may not work when it is concave or discontinuous, but EAs have the advantage of being less susceptible to such characteristics [Coello, 2006].

## **2.5 Hierarchical optimization**

The hierarchical optimization is used for solving decentralized problems with multiple influencing factors, such as the optimization of ecosystem models [Wu et al., 2002; McMahon et al., 2007]. These are complex models that represent nonlinear processes and self-organized elements/components such as soil, plants and atmosphere. In order to be solved hierarchically, optimization problems should be separated into groups according to common characteristics and these groups should be hierarchically organized. For example, ecosystem functioning can be separated by the different temporal scales of the processes within and between components. The optimization decisions are sequential and independent at each level. Despite this and the optimization done for each problem own benefit, all the other model results are affected by the previous optimizations [Lai, 1996].

### **3. METHODOLOGY**

#### **3.1 IBIS model**

The IBIS (Integrated Biosphere Simulator) was designed to connect land and hydrologic processes, land biogeochemical cycles and vegetation dynamics into a single modeling framework [Foley et al., 1996; Kucharik et al., 2000]. The model represents a wide range of processes, including land surface physics (solar and infrared radiative transfer through the canopy, turbulent processes, water interception and heat and mass transfer by canopy), canopy physiology, plant phenology, vegetation dynamics and competition, and carbon and nutrient cycling. All processes are hierarchically organized and operate on different time scales, from minutes (radiative fluxes) to years (carbon allocation). The model is forced by hourly inputs of air temperature and specific humidity, precipitation, wind speed, short- and long-wave incoming radiative fluxes.

Among the different processes and variables simulated by IBIS, seven output variables were chosen to be optimized, according to the availability of field data: outgoing photosynthetically active radiation (PARo), fraction of absorbed

photosynthetically active radiation (fAPAR), net radiation (Rn), friction velocity ( $u^*$ ), sensible heat flux (H), latent heat flux (LE) and net ecosystem exchange (NEE). IBIS has 43 potential parameters for calibration that are described in Table 1.

### **3.2 Site description and observed data**

This study used data from an Amazon micrometeorological site: Tapajós National Forest (K67), an intensive field site located near km 67 of the Santarém-Cuiabá highway ( $02^{\circ} 51' 25''\text{S}$ ,  $54^{\circ} 58' 15''\text{W}$ ), south of Santarém, Pará, Brazil. The typical land cover is evergreen broadleaf forest, formed mainly by evergreen and some semideciduous species. The forest spans 5 km to the east, 8 km to the south and 40 km to the north, before bordering pasture [Saleska et al., 2003]. Canopy height is 40 m on average, with emerging trees reaching up to 55 m [Costa et al., 2010].

This is an intensive data collection site, including measurements of meteorological variables (air temperature and specific humidity, precipitation, wind speed, short- and long-wave incoming radiative fluxes) and fluxes of mass, energy and momentum (PARo, fAPAR, Rn,  $u^*$ , NEE, H, LE) [Saleska et al., 2003; Costa et al., 2010]. Other ecological measurements like leaf area index, net primary production and biomass were also taken at this site, but were not included in this study.

### **3.3 Data gap-filling and data filtering**

Observed data used as model input usually present gaps. These data gaps were filled using interpolation, according to Senna et al. [2009]. This method uses three different equations: the first one is used when the gap duration is less than or equal to 3 hours, the second equation is used when the gap duration is greater than 3 hours or less than 24 hours and the third is used when the gap duration is greater than or equal to 24 hours.

Three different types of data filtering were performed before the model's evaluation [Senna et al., 2009]. The first was the input data filtering, which eliminated periods with missing data in order to avoid inconsistencies produced by data gap-filling. The

Table 1: Potential parameters for the calibration of IBIS

Parameter name	Computational notation	Description
$\rho_{VIS}$	rhoveg_vis	Leaf reflectance on upper canopy (visible) (dimensionless)
$\rho_{NIR}$	rhoveg_NIR	Leaf reflectance on upper canopy (NIR) (dimensionless)
$\tau_{VIS}$	tauveg_vis	Leaf transmittance on upper canopy (visible) (dimensionless)
$\tau_{NIR}$	tauveg_NIR	Leaf transmittance on upper canopy (NIR) (dimensionless)
$\chi_u$	chifuz	Leaf orientation factor on upper canopy (-1 vertical, 0 random, 1 horizontal)
$V_{max}$	vmax_pft	Maximum capacity of enzyme Rubisco at 15 °C on upper canopy (mol[CO <sub>2</sub> ]m <sup>-2</sup> s <sup>-1</sup> )
$m$	coefmub	Coefficient for stomatal conductance (dimensionless)
$C_S$	chs	Heat capacity of stems on upper canopy (J kg <sup>-1</sup> m <sup>-2</sup> )
$C_U$	chu	Heat capacity of upper leaves (J kg <sup>-1</sup> m <sup>-2</sup> )
$C_L$	chl	Heat capacity of lower canopy leaves and stems (J kg <sup>-1</sup> m <sup>-2</sup> )
$\beta_2$	beta2	Parameter of fine roots distribution (Jackson rooting profile) (dimensionless)
$f_a$	funca_coef	Temperature function for aboveground biomass (stems) (dimensionless)
$f_b$	funcb_coef	Temperature function for belowground biomass (roots) (dimensionless)
$R_{mr}$	rroot_coef	Maintenance respiration coefficient for root (s <sup>-1</sup> )
$R_{mw}$	rwood_coef	Maintenance respiration coefficient for wood (s <sup>-1</sup> )
$R_g$	rgrowth_coef	Growth respiration coefficient (fraction) (dimensionless)
$tv$	tempvm_coef	Vmax thermal stress (dimensionless)
$ms$	stressf_coef	Factor of the soil moisture stress (dimensionless)
$c_{ll}^{lit}$	clitll_coef	Initial value of carbon in leaf litter pool (lignin) (kg C m <sup>-2</sup> )
$c_{lm}^{lit}$	clitlm_coef	Initial value of carbon in leaf litter pool (metabolic) (kg C m <sup>-2</sup> )
$c_{ls}^{lit}$	clitls_coef	Initial value of carbon in leaf litter pool (structural) (kg C m <sup>-2</sup> )
$c_{rl}^{lit}$	clitrl_coef	Initial value of carbon in fine root litter pool (lignin) (kg C m <sup>-2</sup> )
$c_{rm}^{lit}$	clitrm_coef	Initial value of carbon in fine root litter pool (metabolic) (kg C m <sup>-2</sup> )
$c_{rs}^{lit}$	clitrs_coef	Initial value of carbon in fine root litter pool (structural) (kg C m <sup>-2</sup> )
$c_{wl}^{lit}$	clitwl_coef	Initial value of carbon in woody litter pool (lignin) (kg C m <sup>-2</sup> )
$c_{wm}^{lit}$	clitwm_coef	Initial value of carbon in woody litter pool (metabolic) (kg C m <sup>-2</sup> )
$c_{ws}^{lit}$	clitws_coef	Initial value of carbon in woody litter pool (structural) (kg C m <sup>-2</sup> )
$c_{pas}^{soi}$	csoipas_coef	Initial value of carbon in soil (passive humus) (kg C m <sup>-2</sup> )
$c_{slon}^{soi}$	csoislon_coef	Initial value of carbon in soil (slow unprotected humus) (kg C m <sup>-2</sup> )
$c_{slop}^{soi}$	csoislop_coef	Initial value of carbon in soil (slow protected humus) (kg C m <sup>-2</sup> )
$wsoi$	wsoi_coef	Initial soil moisture (m <sup>3</sup> m <sup>-3</sup> )
$k$	kfactor	Multiplication factor of decay constant for carbon pools (dimensionless)
$\tau_L$	tauleaf	Foliage biomass turnover time constant (years)
$\tau_R$	tauroot	Fine root biomass turnover time constant (years)
$\tau_W$	tauwood0	Wood biomass turnover time constant (years)
$SLA$	specla	Specific leaf area (m <sup>2</sup> kg <sup>-1</sup> )
$a_L$	aleaf	Carbon allocation fraction to leaves (dimensionless)
$a_R$	aroot	Carbon allocation fraction to fine roots (dimensionless)
$a_W$	awood	Carbon allocation fraction to wood (dimensionless)
$d$	dispu_coef	Zero-plane displacement height for upper canopy (m)
$\log(Zol)$	alogl_coef	Natural logarithm of roughness length of lower canopy (dimensionless)
$\log(Zou)$	alogu_coef	Natural logarithm of roughness length of upper canopy (dimensionless)
	avmuir_coef	Average diffuse optical depth (m <sup>3</sup> m <sup>-3</sup> )

second was the NEE filtering by a friction velocity threshold. Data was also eliminated from the analysis if the condition  $0.6 < + < 1.4$  was not satisfied for daily averages.

### **3.4 OPTIS**

OPTIS, developed by Varejão-Jr. et al. (2012), is a hierarchical optimization algorithm based on the multi-objective genetic algorithm NSGA-II (Nondominated Sorted Genetic Algorithm II) [Deb et al., 2002]. This algorithm has limitations when dealing with a large number of objective functions at once, which accounts for loss of optimization efficiency [Luan et al., 1996]. Therefore, the maximum number of simultaneous objective functions considered in this study was three. In the multi-objective optimization, we can define the best solution for the optimization problem using different metrics [Fonseca et al., 1996; Deb et al., 2002]. This study considered as the best solution, the point with the smallest Euclidean distance from the origin in the Pareto frontier [Varejão-Jr. et al., 2012].

OPTIS was originally implemented to provide an automatic and multi-objective hierarchical calibration of the LSEM IBIS. The interaction between IBIS and OPTIS is minimal (Figure 1). The optimization algorithm changes the parameters of the model and reads its output data. In this way, IBIS is executed externally to the optimization algorithm and OPTIS is practically independent from the IBIS version or even from the model to be calibrated itself [Varejão-Jr. et al., 2012].

### **3.5 IBIS' sensitivity analysis**

The SA is able to detect which model parameters mostly influence the model results [Hamby, 1994] and how the variation in the output of a model can be apportioned in qualitative or quantitative terms [Saltelli et al., 2000; 2008]. In addition to the identification of each influential factor, the SA also tries to qualify their relative importance.

To overcome the limitation of the Morris method (due to the random initiation) and to eliminate outliers, the Morris algorithm was applied ten times and, to choose and

quantify the most sensitive parameters, the average of the sum of the measures  $\mu^*$  [Saltelli et al., 2005] and  $\sigma$  [Morris, 1991] were used.

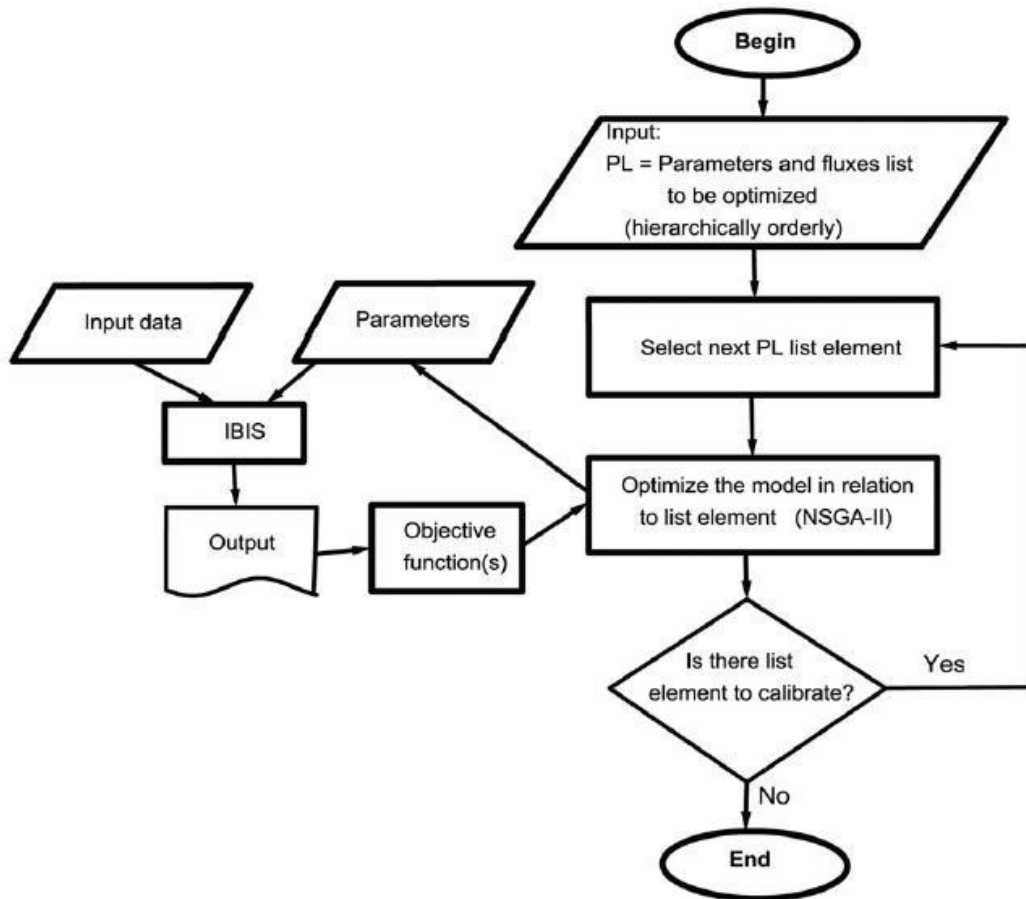


Figure 1: OPTIS structure [Varejão et al., 2012].

### 3.6 Mono-objective calibration

Two types of mono-objective calibrations were performed. First, seven overfit calibrations were run [Hawkins, 2004], optimizing all 43 available parameters for each of the seven output variables. A second set of seven mono-objective calibrations was also run. In this set, the only parameters used were those to which the model is most sensitive, selected by the SA.



### 3.7 Multi-objective hierarchical calibration

In a multi-objective hierarchical calibration each parameter can be associated to a single hierarchical level. Thus different results are expected for different arrangements of the same variables, since when the model is optimized at a certain hierarchical level, the parameters of the previous level have already been determined and remain fixed.

The best result (smallest MAE) between the two mono-objective runs (overfit and using SA) is the reference to analyze and understand the other calibrations, according to Equation 1. The comparison between mono-objective and hierarchical calibration provides a relative performance index (D) of the calibration method. The multi-objective calibration efficiency is proportional to the index D value.

The relative performance index D is defined as the average of the individual ratios between the objective function of the best mono-objective overfit calibration output ( $f^{mono}$ ) and the objective function of the multi-objective calibration output ( $f^{multi}$ ). Mathematically,

$$D = \frac{1}{n} \sum_{i=1}^n \frac{f_i^{mono}}{f_i^{multi}} \quad (1)$$

where n is the number of objective functions and the D value should be between 0 and 1.

In order to organize the variables in hierarchical levels, the output variables were separated into two groups: the first related to radiative fluxes (PAR<sub>o</sub>, fAPAR, R<sub>n</sub>) and the second related to turbulent fluxes (u\*, H, LE, NEE). Each group may be arranged in one or more hierarchical levels. Even though the order of the output variables within a group may be changed, the order in which the objective functions are optimized must be respected by the groups, because a necessary condition of the hierarchical approach is that the optimization moves from the fastest to the slowest process.

Two types of analyses were performed. First, the simultaneous (traditional) multi-objective calibration was compared to the hierarchical multi-objective calibration,

when such comparison was possible ( $n \leq 3$ ). Then the hierarchical calibration optimizing all model outputs was performed as described below.

To facilitate the interpretation of the results, the multi-objective hierarchical calibration was divided in three parts. Initially, only output variables related to radiative fluxes were optimized, as described in section 3.7.1. Two or three variables were hierarchically optimized, considering one or two hierarchical levels (Table 2).

Afterwards, only output variables related to turbulent fluxes were optimized, as shown in section 3.7.1. A number from two to four variables were hierarchically optimized, considering one or two hierarchical levels as well (Table 3).

Finally, the output variables of any of the previous groups were optimized (section 3.7.2). A number from two to seven variables were hierarchically optimized, with the number of hierarchical levels varying from two to four.

### **3.7.1 Comparison between simultaneous and hierarchical calibration**

Two methods of calibration were compared for the group of variables related to radiative fluxes (PARo, fAPAR and Rn) as well as for the group of variables related to turbulent fluxes ( $u^*$ , H, LE, NEE). Hereinafter, curly brackets  $\{\}$  were used to identify output variables that were simultaneously optimized. Hence the simultaneous calibration is denoted by all variables in the same bracket (e.g., {PARo-fAPAR-Rn}), whereas the hierarchical calibration is denoted by a sequence of brackets, e.g. {PARo-fAPAR}{Rn}, in which the variables were optimized following the brackets sequence.

There are ten different simulation combinations that can be separated in four simulations using one hierarchical level (simultaneous calibration) and six simulations using two hierarchical levels (Table 2).

It should be noted that the order of variables in the same bracket ({PARo-Rn} or {Rn-PARo}) is irrelevant, since the optimization is simultaneous. On the other hand, the order of hierarchical levels ({PARo}{Rn} or {Rn}{PARo}) may affect the results, thus all combinations were tested.

It is also important to highlight in Table 2 that PARo and fAPAR are present either individually or in the same hierarchical level, but not in two different hierarchical levels. This is explained by the fact the parameter  $\chi_u$  is as important to PARo as to fAPAR, so they must either be optimized simultaneously or individually, but not sequentially.

To optimize the turbulent fluxes variables, there are twenty-four different simulation combinations that can be separated in ten simulations using one hierarchical level (simultaneous calibration) and fourteen simulations using two hierarchical levels (Table 3). As explained for radiative flux variables, because the order of hierarchical levels may affect the results, all combinations were tested.

Table 2: List of optimization experiments considering the group of variables related to radiative fluxes.

Simultaneous	Hierarchical
{PARo-fAPAR}	
{PARo-Rn}	{PARo} {Rn}
	{Rn} {PARo}
{fAPAR-Rn}	{fAPAR} {Rn}
	{Rn} {fAPAR}
{PARo-fAPAR-Rn}	{PARo-fAPAR} {Rn}
	{Rn} {PARo-fAPAR}

The variables H, LE and NEE cannot be separated in different hierarchical levels because the most sensitive parameters for them are the same ( $\beta_2, V_{\max}, R_g$  and others), thus they must be optimized simultaneously in the same hierarchical level.

### 3.7.2 Hierarchical calibration

When all variables (related to radiative and turbulent fluxes) are taken into account, there are 397 simulation combinations that can be separated in 31 simulations using two output variables, 85 using three output variables, 122 using four output variables,

113 using five output variables, 45 using six output variables and 6 simulations using all seven output variables.

The combination of variables was carried out as described in section 3.7.1, emphasizing that variables from different groups (radiative and turbulent fluxes) cannot be calibrated simultaneously in the same hierarchical level because the main assumption of the hierarchical optimization is that it is performed from the fastest to the slowest processes.

Table 3: List of optimization experiments considering the group of variables related to turbulent fluxes.

Simultaneous	Hierarchical
{H-LE}	
{H-NEE}	
{LE-NEE}	
{u*-H}	{u*} {H}
	{H} {u*}
{u*-LE}	{u*} {LE}
	{LE} {u*}
{u*-NEE}	{u*} {NEE}
	{NEE} {u*}
{H-LE-NEE}	
{u*-H-LE}	{u*} {H-LE}
	{H-LE} {u*}
{u*-H-NEE}	{u*} {H-NEE}
	{H-NEE} {u*}
{u*-LE-NEE}	{u*} {LE-NEE}
	{LE-NEE} {u*}
	{u*} {H-LE-NEE}
	{H-LE-NEE} {u*}

## 4. RESULTS AND DISCUSSION

### 4.1 IBIS' sensitivity analysis

A rank determining the sensitivity of each variable to each parameter was built using the average of ten sensitivity measures ( $\mu^* + \sigma$ ) (Table 4). The greater the sum, the more sensitive is the model to that parameter. As shown in Table 4, PARo, fAPAR, Rn, H and LE are sensitive to 30 parameters,  $u^*$  to 25 parameters and NEE to 37 parameters.

When the variables are separated into different hierarchical levels, it is possible to choose which parameters will be in each hierarchical level, considering that one parameter can be calibrated only once. Such decisions were made according to the degree of sensitivity shown in Table 4.

Some examples about how the division of the parameters was carried out are shown in Table 5. Despite the fact that the variables are the same, simulations with different order of the hierarchical levels can result in a different parameter configuration, e.g. {fAPAR}{Rn} and {Rn}{fAPAR} (Table 5). In the former, fAPAR is calibrated using only 5 parameters because a calibration using only these parameters is enough

Table 4: Degree of sensitivity ( $\mu^* + \sigma$ ) of all parameters for each output variables according to the Morris method. Unit of the degree of sensitivity is the same of the output variable.

Parameter	Output variables						
	PARo	fAPAR	Rn	$u^*$	H	LE	NEE
$\rho_{VIS}$	2.7741	0.0238	1.0061	0.0014	1.2781	0.2770	0.0866
$\rho_{NIR}$	0.3338	0.0165	2.9942	0.0034	4.3265	0.6197	0.0830
$\tau_{VIS}$	0.4307	0.0088	0.1844	0.0003	0.2798	0.0986	0.0304
$\tau_{NIR}$	0.1302	0.0058	0.7311	0.0007	1.0033	0.1357	0.0368
$\chi_u$	2.5591	0.0388	2.6206	0.0066	3.2947	1.3362	0.2181
$V_{max}$	1.8075	0.0697	8.5504	0.0186	8.2274	2.9153	0.3867
$m$	0.1847	0.0146	0.5654	0.0013	1.2764	1.2890	0.2162
$C_S$	0.0074	0.0013	0.0847	0.0015	0.1541	0.1920	0.0126
$C_U$	0.0049	0.0003	0.0300	0	0.0510	0.0470	0.0056
$C_L$	0.0040	0.0007	0.0469	0	0.1041	0.2111	0.0044
$\beta_2$	0.4456	0.0228	1.1070	0.0031	1.0951	0.9740	0.1314
$fa$	0.1990	0.0165	0.2727	0.0012	0.3332	0.1459	0.0407
$fb$	0.2031	0.0176	0.8200	0.0025	0.7640	0.2567	0.0659
$R_{mr}$	0.2204	0.0250	1.2808	0.0035	1.0604	0.4059	0.0964
$R_{mw}$	0.2028	0.0107	0.4380	0.0017	0.5719	0.3015	0.0400
$R_g$	0.8858	0.0656	4.6257	0.0146	5.2825	1.1712	0.5643
$tv$	1.0953	0.0366	2.7070	0.0065	2.9048	1.4507	0.3595
$ms$	0.0203	0.0025	0.1052	0	0.1594	0.0905	0.0210
$c_{ll}^{lit}$	0	0	0	0	0	0	0
$c_{lm}^{lit}$	0	0	0	0	0	0	0
$c_{ls}^{lit}$	0	0	0	0	0	0	0.0022
$c_{rl}^{lit}$	0	0	0	0	0	0	0
$c_{rm}^{lit}$	0	0	0	0	0	0	0
$c_{rs}^{lit}$	0	0	0	0	0	0	0.0035
$c_{wl}^{lit}$	0	0	0	0	0	0	0
$c_{wm}^{lit}$	0	0	0	0	0	0	0
$c_{ws}^{lit}$	0	0	0	0	0	0	0.0023
$c_{pas}^{soi}$	0	0	0	0	0	0	0.0006
$c_{slon}^{soi}$	0	0	0	0	0	0	0.0014
$c_{slop}^{soi}$	0	0	0	0	0	0	0.0527
$wsoi$	0.2179	0.0068	0.5588	0.0013	0.4516	0.5251	0.0833
$k$	0	0	0	0	0	0	0.0184
$\tau_L$	0.7217	0.0377	1.8662	0.0054	2.5818	1.0698	0.2470
$\tau_R$	0.1342	0.0136	0.7323	0.0028	0.9929	0.3675	0.0811
$\tau_W$	0.0942	0.0061	0.1839	0.0043	0.3480	0.1141	0.0438
SLA	0.1681	0.0181	1.3231	0.0020	1.8983	0.3192	0.0845
$a_L$	0.1334	0.0270	1.3143	0.0028	1.8930	0.4402	0.0814
$a_R$	0.1823	0.0093	1.0189	0.0022	1.0221	0.5253	0.0743
$a_W$	0.0150	0.0004	0.0443	0.0002	0.0764	0.0169	0.0026
$d$	0.0166	0.0014	0.1338	0.0090	0.4498	0.1562	0.0225
$\log(Zol)$	0.0062	0.0011	0.0279	0	0.0502	0.0192	0.0026
$\log(Zou)$	0.0076	0.0005	0.0367	0.0009	0.0724	0.0313	0.0053
$\mu$	0.0068	0.0008	0.3909	0	0.1989	0.0874	0.0043

Table 5: Parameters optimized in the calibration experiments related to radiative flux variables. Each pair of bracket represents one hierarchical level. Some parameters are not shown in the table because they are not used in any calibration below.

Parameter	Output Variables														
	One hierarchical level				Two hierarchical levels										
	{PARo-fAPAR}	{PARo-Rn}	{fAPAR-Rn}	{PARo-fAPAR-Rn}	{PARo}{Rn}	{Rn}{PARo}	{fAPAR}{Rn}	{Rn}{fAPAR}	{PARo-fAPAR}{Rn}	{Rn}{PARo-fAPAR}					
$\rho_{VIS}$	X	X	X	X	X		X	X		X	X	X	X		X
$\rho_{NIR}$	X	X	X	X		X	X		X	X			X	X	
$\tau_{VIS}$	X	X	X	X	X			X	X		X	X			X
$\tau_{NIR}$	X	X	X	X		X	X		X	X			X	X	
$\chi_u$	X	X	X	X	X		X	X			X	X			X
$V_{max}$	X	X	X	X		X	X		X	X			X	X	
$m$	X	X	X	X		X	X		X	X			X	X	
$C_S$	X	X	X			X	X		X	X			X	X	
$C_U$	X	X	X			X	X		X	X			X	X	
$C_L$	X	X				X	X		X				X	X	
$\beta_2$	X	X	X	X		X	X		X	X			X	X	
$fa$	X	X	X	X		X	X		X	X			X	X	
$fb$	X	X	X	X		X	X		X	X		X			X
$R_{mr}$	X	X	X	X		X	X		X		X	X			X
$R_{mw}$	X	X	X	X		X	X		X		X	X			X
$R_g$	X	X	X	X		X	X		X		X		X	X	
$tv$	X	X	X	X		X	X		X		X		X	X	
$ms$	X	X	X	X		X	X		X		X		X	X	
$wsoi$	X	X	X	X		X	X		X		X		X	X	
$\tau_L$	X	X	X	X		X	X		X		X		X	X	
$\tau_R$	X	X	X	X		X	X		X		X		X	X	
$\tau_W$	X	X	X	X		X	X		X		X		X	X	
$SLA$	X	X	X	X		X	X		X		X		X	X	
$a_L$	X	X	X	X		X	X		X		X		X	X	
$a_R$	X	X	X	X		X	X		X		X		X	X	
$a_W$	X	X	X	X		X	X		X		X		X	X	
$d$	X	X	X	X		X	X		X		X		X	X	
$\log(Zol)$	X	X		X					X				X	X	
$\log(Zou)$	X	X		X		X	X		X		X		X	X	
$\mu$	X	X	X	X		X	X		X	X			X	X	

to get a satisfactory result of fAPAR. Furthermore, Rn does not present a good fit if simulated with many parameters calibrated to fAPAR. The same happens when Rn is in the first hierarchical level and is calibrated using only three parameters.

## 4.2 Mono-objective calibration

Mono-objective calibrations were performed in two different ways: calibrations using all 43 available parameters (overfit) and calibrations using only the sensitive parameters (Table 4). The MAE results for both mono-objective calibrations are shown in Table 6. For the seven optimized variables, the optimizations using only the SA parameters produced the smallest MAE in all cases.

Table 6: MAE results for the mono-objective calibrations.  $f^{\text{mono,overfit}}$  is the simulation using all 43 available parameters and  $f^{\text{mono,SA}}$  is the simulation using only the parameters selected according to the sensitivity analysis validation

Output	$f^{\text{mono,overfit}}$	$f^{\text{mono,SA}}$	$f^{\text{mono,best}}$
PARo ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ )	1.37125	1.36715	1.36715
fAPAR	0.00592	0.00583	0.00583
Rn ( $\text{W m}^{-2}$ )	16.53813	16.52972	16.52972
$u^*$ ( $\text{m s}^{-1}$ )	0.10018	0.10008	0.10008
H ( $\text{W m}^{-2}$ )	16.90152	16.76446	16.76446
LE ( $\text{W m}^{-2}$ )	36.36789	35.97890	35.97890
NEE ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{s}^{-1}$ )	3.81659	3.80435	3.80435

When a simulation uses more optimized parameters than what is necessary (overfitting), it can result in worse predictions for two reasons: first, the use of a non-influential parameter adds random variation to the other parameters; second, fewer parameters generate a smaller sample space, so the search for the Pareto frontier becomes more efficient, yielding better results.



In order to understand the MAE results, dispersion and typical day plots were produced for each mono-objective calibration. According to Figures 2-8, most outputs (except fAPAR and Rn) fit better to the data when only the sensitive parameters were used.

Figure 2 shows the nearly perfect fit of the hourly data of PARo. The model slightly overestimated PARo between 9 and 12 hours and underestimated PARo in the first hours of the night (Figure 2-b and d). The regression coefficients were  $a = 1.01$  and  $b = -0.19$  for the overfit calibration (Figure 2-a) and  $a = 1.00$  and  $b = -0.14$  for the SA calibration (Figure 2-c).

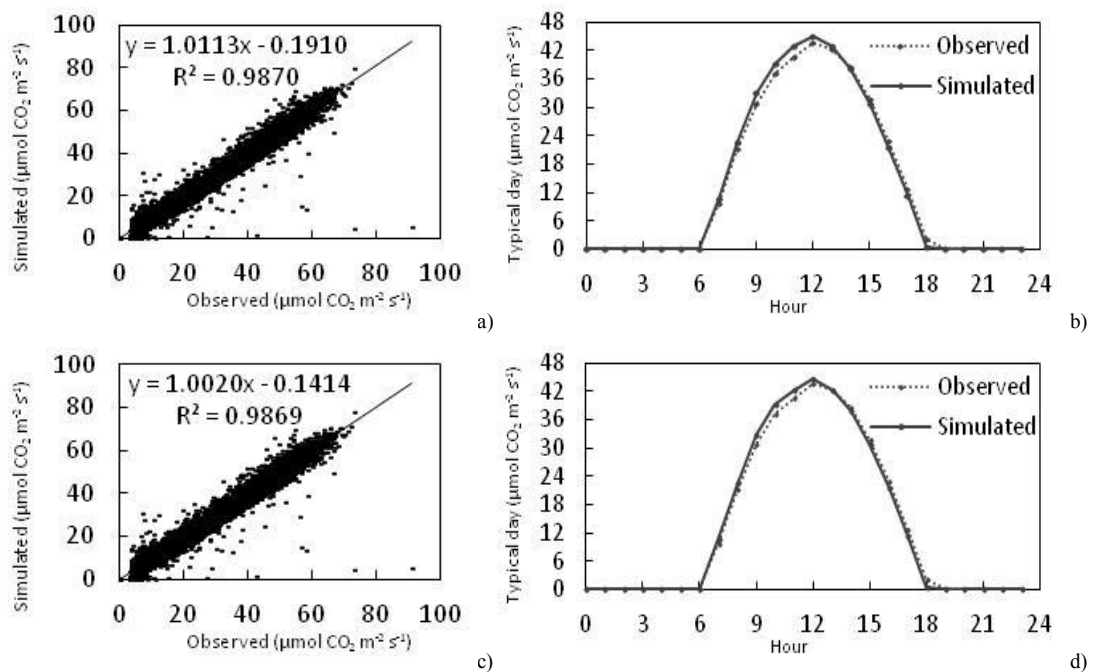


Figure 2: PARo results after the mono-objective calibration: a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 3 shows the mono-objective calibration of fAPAR. The bias between observed and simulated data was small, increasing in February (mid of summer) and August (mid of spring) (Figure 3-b and d). The regression coefficients were  $a = 0.84$

and  $b = 0.15$  for the overfit calibration (Figure 3-a) and  $a = 0.88$  and  $b = 0.11$  for the SA calibration (Figure 3-c).

Figure 4 shows the nearly perfect fit of the hourly data of  $R_n$ . The difference between both mono-objective calibrations of  $R_n$  was small. The regression coefficients were  $a = 0.99$  and  $b = 7.39$  for overfit calibration (Figure 4-a) and  $a = 0.99$  and  $b = 7.72$  for the SA calibration (Figure 4-c).

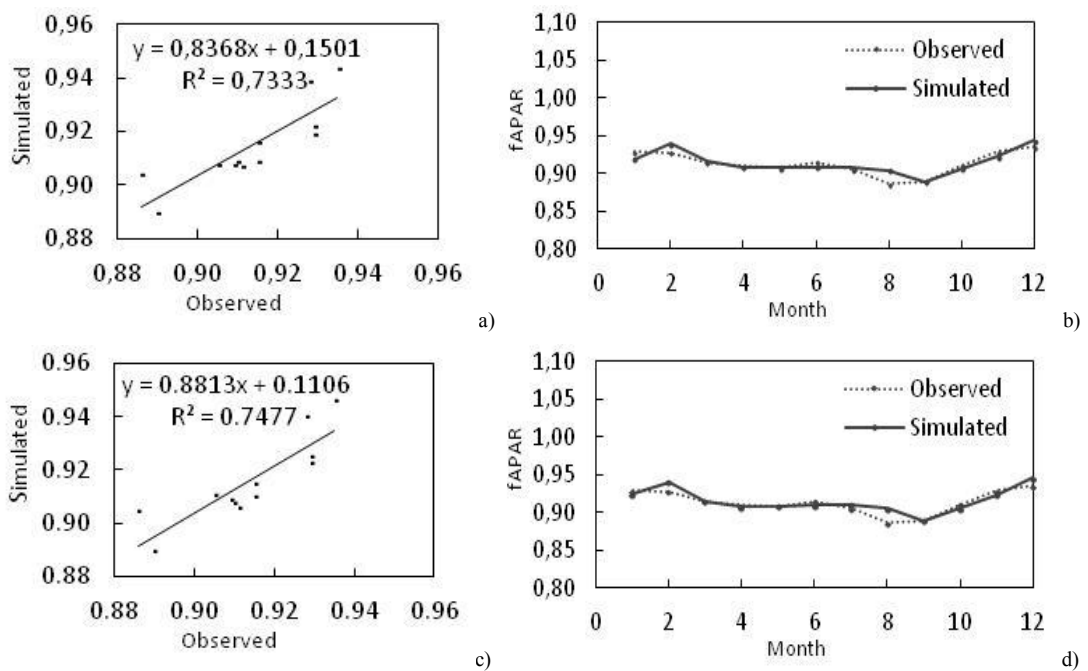


Figure 3: fAPAR results after the mono-objective calibration a) Dispersion plot and b) seasonal variation using all 43 parameters and c) Dispersion plot and d) seasonal variation using only sensitive parameters.

Figure 5 shows the mono-objective calibration of  $u^*$ . The model overestimated  $u^*$  during the day (unstable condition) and underestimated  $u^*$  during the night (stable condition), with a delay of one hour between the simulated and the observed maximum values (Figure 5-b and d). This is an indication that the model formulation for turbulence is incomplete, and mere adjustments in the parameters are insufficient to match the results. The regression coefficients were  $a = 0.52$  and  $b = 0.14$  for the

overfit calibration (Figure 5-a) and  $a = 0.53$  and  $b = 0.14$  for the SA calibration (Figure 5-c).

Figure 6 shows the mono-objective calibration of H. The bias between observed and simulated data was small and the overestimation of H was noticeable during the night (Figure 6-b and d). On the other hand, H simulations during daytime were satisfactory. The regression coefficients were  $a = 0.73$  and  $b = 13.13$  for the overfit calibration (Figure 6-a) and  $a = 0.76$  and  $b = 13.06$  for the SA calibration (Figure 6-c). This was a fairly surprising result, considering that the model does not generate enough turbulence (Figure 5) and means that OPTIS has found a way to improve the IBIS simulation for H, by excessively heating up the canopy during daytime, through changes in the net radiation and leaves heat capacity (results not shown). This is a common behavior in mono-objective optimizations, indicating the need for multiple output optimizations.

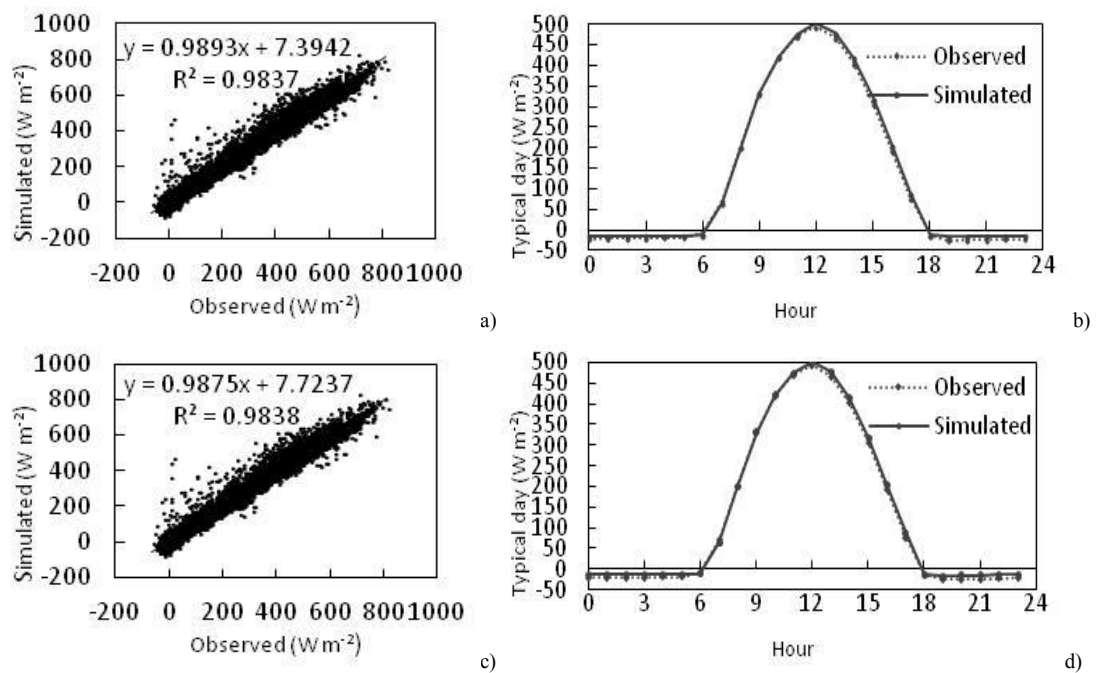


Figure 4: Rn results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

Figure 7 shows the mono-objective calibration of LE. The model was not able to estimate the high values during the day, causing a bias of about  $80 \text{ W m}^{-2}$  in the daily peak (Figure 7-b and d). This is a direct consequence of the low turbulence generated by the model, although, contrary to H, in this case the optimizer apparently could not find alternative ways to increase LE values. The regression angular coefficients were  $a = 0.59$  for the overfit calibration (Figure 7-a) and  $a = 0.60$  for the SA calibration (Figure 7-c), just slightly higher than the angular coefficient for the  $u^*$  optimization ( $a = 0.52$ ).

Figure 8 shows the mono-objective calibration of NEE. The model overestimated NEE during the day and there was an advance of one hour in the minimum value (Figure 8-b and d). The regression coefficients were  $a = 0.76$  and  $b = -1.02$  for the overfit calibration (Figure 8-a) and  $a = 0.76$  and  $b = -0.99$  for the SA calibration (Figure 8-c).

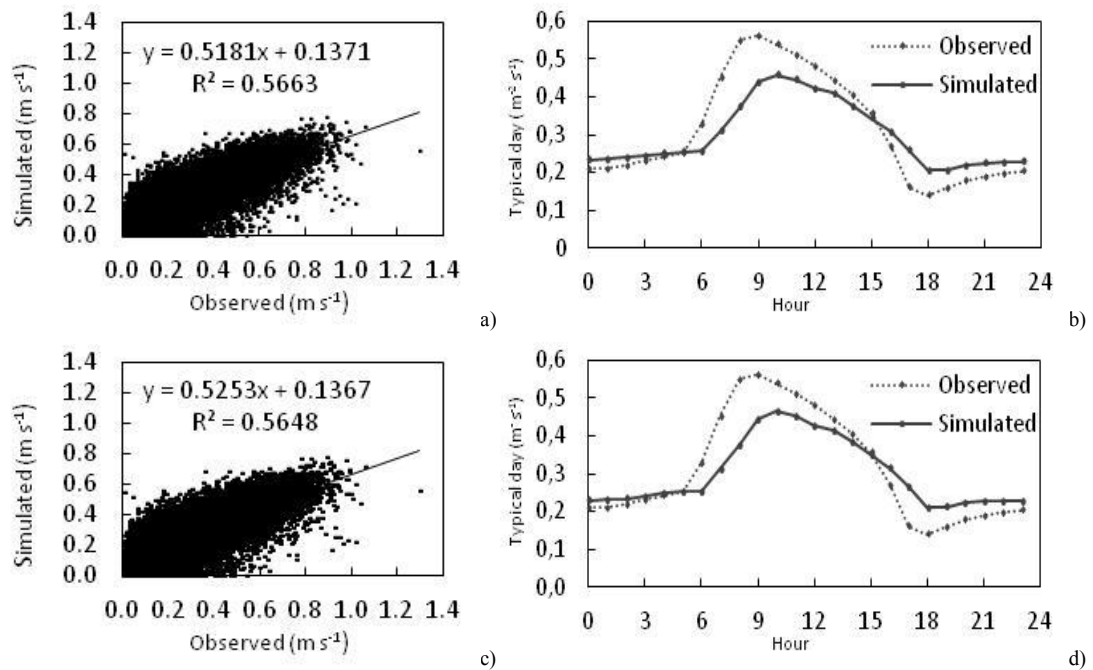


Figure 5:  $u^*$  results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

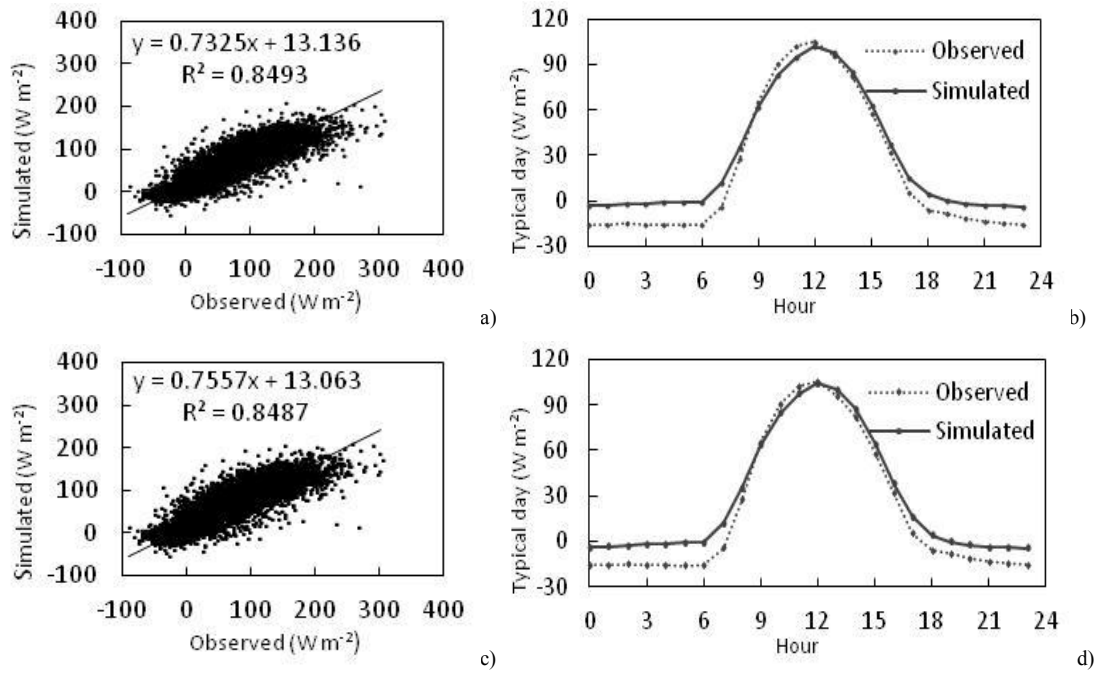


Figure 6: H results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

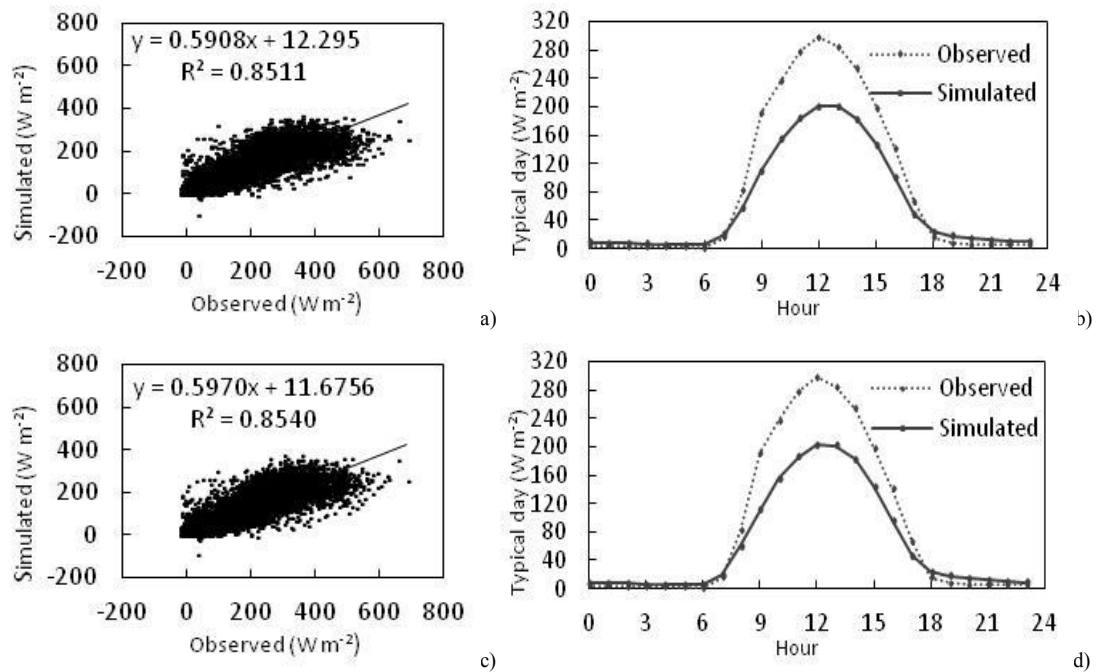


Figure 7: LE results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

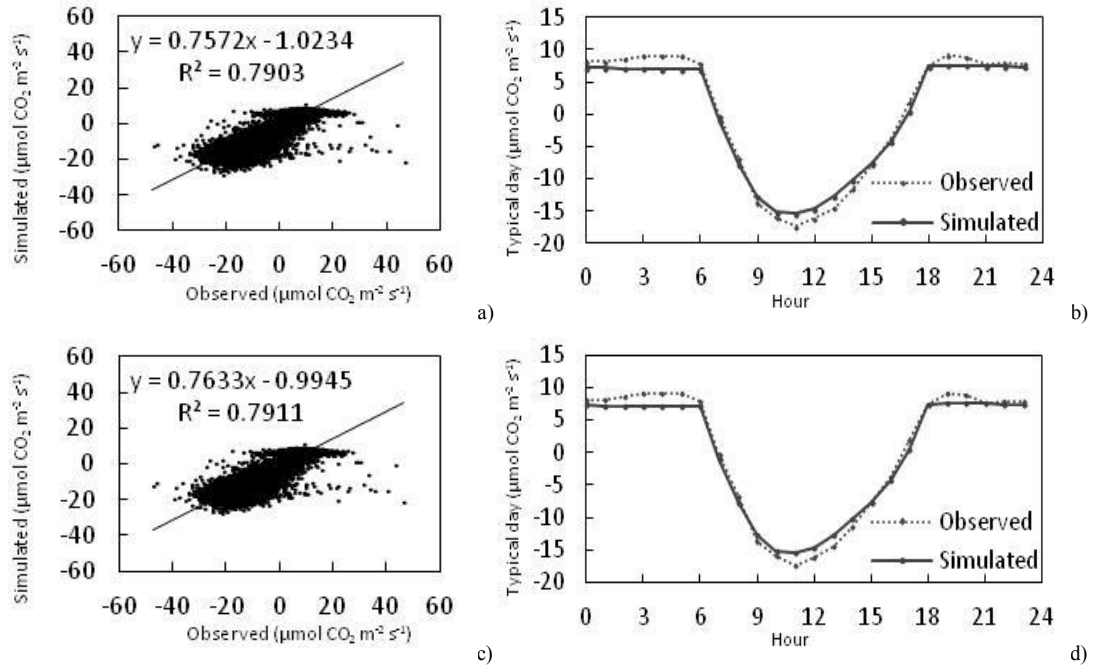


Figure 8: NEE results after the mono-objective calibration a) Dispersion plot and b) typical day using all 43 parameters and c) Dispersion plot and d) typical day using only sensitive parameters.

### 4.3 Multi-objective calibration

#### 4.3.1 Comparison between simultaneous and hierarchical calibration

All combinations of radiative flux variables were tested and the results were separated by number of hierarchical levels (Table 7 and Figure 9). One hierarchical level is equivalent to the simultaneous (traditional) optimization. For all situations (using one or two hierarchical levels), the higher the number of hierarchical levels, the better the results: when optimizing PAR<sub>o</sub> and R<sub>n</sub>, the values obtained for the index D were 0.9787 for the configuration {R<sub>n</sub>} {PAR<sub>o</sub>}, 0.9680 for the configuration {PAR<sub>o</sub>} {R<sub>n</sub>}, and only 0.9129 for the simultaneous configuration {PAR<sub>o</sub>-R<sub>n</sub>}. A similar behavior was observed when optimizing fAPAR and R<sub>n</sub>, with D values of 0.8770 for the configuration {R<sub>n</sub>} {fAPAR}, 0.8669 for the configuration {fAPAR} {R<sub>n</sub>}, and the smallest D value of 0.7600 for the simultaneous configuration {fAPAR-R<sub>n</sub>}. When optimizing all the three radiative flux variables (PAR<sub>o</sub>, fAPAR and R<sub>n</sub>), the index D values obtained were 0.8948 for the configuration {R<sub>n</sub>} {PAR<sub>o</sub>-fAPAR}, 0.8443 for the configuration {PAR<sub>o</sub>-

{fAPAR} {Rn}, and only 0.8222 for the simultaneous configuration {PARo-fAPAR-Rn}.

Table 7: Calibration results using only radiative flux output variables. The bracket is used to identify the hierarchical level, h is the number of hierarchical levels and n is the output quantity.

h	n	Optimization configuration	$\frac{f^{mono,best}}{f^{multi}}$			Index D
			PARo	fAPAR	Rn	
1	2	{PARo-fAPAR}	0.9726	0.7447	-	0.8586
1	2	{PARo-Rn}	0.8582	-	0.9676	0.9129
2	2	{PARo} {Rn}	0.9699	-	0.9661	0.9680
2	2	{Rn} {PARo}	0.9707	-	0.9867	0.9787
1	2	{fAPAR-Rn}	-	0.6796	0.8405	0.7600
2	2	{fAPAR} {Rn}	-	0.8235	0.9103	0.8669
2	2	{Rn} {fAPAR}	-	1.0294	0.7246	0.8770
1	3	{PARo-fAPAR-Rn}	0.8888	0.7527	0.8250	0.8222
2	3	{PARo-fAPAR} {Rn}	0.9334	0.6250	0.9746	0.8443
2	3	{Rn} {PARo-fAPAR}	0.9587	0.7955	0.9304	0.8948

The same procedure was carried out for turbulent flux variables (Table 8 and Figure 10). In general, the model fitted better when two hierarchical levels were used, instead of one. When optimizing  $u^*$  and H (or LE), the highest D value was obtained for the configuration {H} { $u^*$ } (or {LE} { $u^*$ }), the second highest D value was presented by the simultaneous configuration { $u^*$ -H} (or { $u^*$ -LE}) and the smallest D value was found for the configuration { $u^*$ } {H} (or { $u^*$ } {LE}). It is important to highlight the use of the variables  $u^*$  and NEE together: they fit better in the same hierarchical level (simultaneous optimization), but it is not possible to calibrate NEE in a different hierarchical level of H and LE, and the algorithm cannot support more than three variables in the same hierarchical level. For three variables, the simultaneous configuration was responsible for the smallest D value in two cases ({ $u^*$ -H-NEE} and { $u^*$ -LE-NEE}) and the highest D value was obtained for the

configuration using  $u^*$  in the second hierarchical level ( $\{H-LE\}\{u^*\}$ ,  $\{H-LE\}\{u^*\}$ ,  $\{LE-NEE\}\{u^*\}$ ).

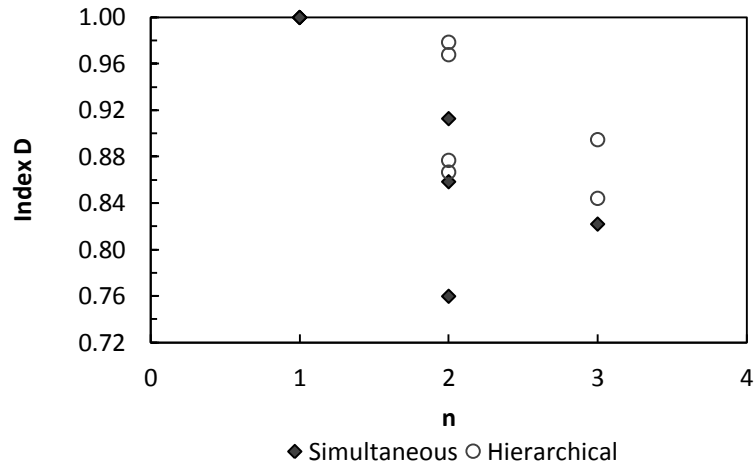


Figure 9: Index D for radiative flux output variables as a function of the number of optimized outputs ( $n$ ), separated by the type of optimization (simultaneous or hierarchical).

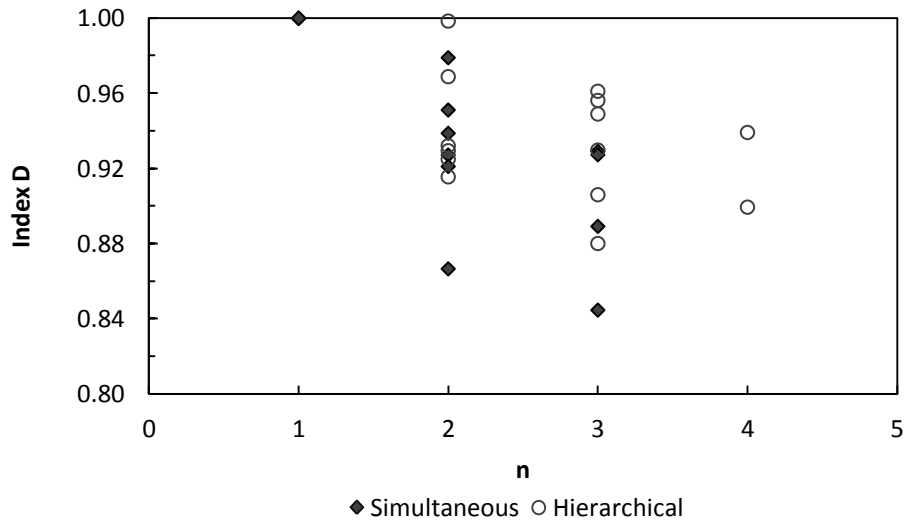


Figure 10: Index D for turbulent flux output variables as a function of the number of optimized outputs ( $n$ ), separated by the type of optimization (simultaneous or hierarchical).



Table 8: Calibration results using only turbulent flux output variables. The bracket is used to identify the hierarchical level, h is the number of hierarchical levels and n is the output quantity.

h	n	Optimization configuration	$\frac{f^{mono,best}}{f^{multi}}$				Index D
			u*	H	LE	NEE	
1	2	{H-LE}	-	0.9451	0.9093	-	0.9272
1	2	{H-NEE}	-	0.7243	-	1.0090	0.8666
1	2	{LE-NEE}	-	-	0.8757	0.9666	0.9211
1	2	{u*-H}	0.9806	0.9774	-	-	0.9790
2	2	{u*}{H}	0.8510	0.9802	-	-	0.9156
2	2	{H}{u*}	1.0000	0.9970	-	-	0.9985
1	2	{u*-LE}	0.9517	-	0.9507	-	0.9512
2	2	{u*}{LE}	0.8623	-	0.9967	-	0.9295
2	2	{LE}{u*}	0.9459	-	0.9919	-	0.9689
1	2	{u*-NEE}	0.9011	-	-	0.9765	0.9388
2	2	{u*}{NEE}	0.8510	-	-	0.9991	0.9250
2	2	{NEE}{u*}	0.8178	-	-	1.0465	0.9322
1	3	{H-LE-NEE}	-	0.9075	0.8990	0.9757	0.9274
1	3	{u*-H-LE}	0.9726	0.8713	0.9436	-	0.9292
2	3	{u*}{H-LE}	0.8510	0.9464	0.9209	-	0.9061
2	3	{H-LE}{u*}	0.9808	0.9563	0.9101	-	0.9491
1	3	{u*-H-NEE}	0.8764	0.7180	-	0.9398	0.8447
2	3	{u*}{H-NEE}	0.8623	0.8707	-	0.9074	0.8801
2	3	{H-NEE}{u*}	0.9715	0.9158	-	0.9962	0.9612
1	3	{u*-LE-NEE}	0.8711	-	0.8170	0.9797	0.8892
2	3	{u*}{LE-NEE}	0.8510	-	0.9831	0.9555	0.9299
2	3	{LE-NEE}{u*}	0.9484	-	0.9467	0.9736	0.9563
2	4	{u*}{H-LE-NEE}	0.8510	0.8720	0.9407	0.9343	0.8995
2	4	{H-LE-NEE}{u*}	0.9759	0.8827	0.9153	0.9830	0.9392

### 4.3.2 Results for the hierarchical calibration

The analysis of the hierarchical calibration results indicates that it is possible to underline some combinations of variables that generate the best results and also to identify the position of each variable and the distribution of parameters among the hierarchical levels to provide for the best fit of observed data.

For radiative flux variables, the best result for D was always obtained when Rn was in the first hierarchical level and PARo or/and fAPAR were in the second hierarchical level (Table 7). These results may be dependent on the choice of the baseline parameters and thus may not be generic. When optimizing fAPAR and Rn

in the configuration {Rn} {fAPAR}, the ratio  $f^{\text{mono}}/f^{\text{multi}}$  for fAPAR was greater than one, i.e., the mono-objective calibration was not the best possible calibration. There are two possible explanations: either the number of parameters used was more than necessary or the values used for the baseline parameters were not the most appropriate.

For turbulent flux variables, the best result was always obtained when  $u^*$  was in the second hierarchical level and H or/and LE or/and NEE were in the first hierarchical level (Table 8). The same discussion presented for fAPAR can be applied to NEE in two cases: configurations {H-NEE} and {NEE} { $u^*$ }.

For radiative and turbulent fluxes, confirming the trend of the previous results, the best result considering the index D (Table 9) occurred for the order below:

1° hierarchical level: Rn

2° hierarchical level: PARo and fAPAR

3° hierarchical level: H, LE and NEE

4° hierarchical level:  $u^*$

Hierarchical calibration performs a global calibration of the model and optimizes, as well as possible, all variables simultaneously. As expected, the individual results for hierarchical calibration are worse than the results for the mono-objective calibration.

In order to understand the hierarchical calibration results, the dispersion and typical day plots were produced for each individual variable of the best seven-objective calibration (Figures 11-17). These results can be compared to those obtained for the mono-objective calibration, presented in Figures 2 to 8.

According to Figures 11 and 13, PARo and Rn still had a nearly perfect fit and Rn underestimated the higher values during the day (Figure 13-b). The regression coefficients were  $a = 0.99$  and  $b = 0.09$  for PARo (Figure 11-a) and  $a = 0.91$  and  $b = 22.63$  for Rn (Figure 13-a).

Table 9: Calibration results using radiative and turbulent flux output variables. The bracket is used to identify the hierarchical level, h is the number of hierarchical levels and n is the output quantity.

h	n	Output	$\frac{f^{mono,best}}{f^{multi}}$							Index D
			PARo	fAPAR	Rn	u*	H	LE	NEE	
3	7	{PARo-Rn-fAPAR} {ustar} {H-LE-NEE}	0.6706	0.5556	0.7468	0.8856	0.2626	0.7031	0.7298	0.6506
4	7	{PARo-fAPAR} {Rn} {ustar} {H-LE-NEE}	0.9476	0.6140	0.7192	0.8882	0.2720	0.7524	0.6555	0.6927
4	7	{Rn} {PARo-fAPAR} {ustar} {H-LE-NEE}	0.7944	0.6481	0.7246	0.8863	0.2733	0.7242	0.8479	0.6998
3	7	{PARo-Rn-fAPAR} {H-LE-NEE} {ustar}	0.6706	0.5556	0.7468	0.9136	0.2723	0.7228	0.7901	0.6674
4	7	{PARo-fAPAR} {Rn} {H-LE-NEE} {ustar}	0.9476	0.6140	0.7192	0.8920	0.2761	0.7174	0.8037	0.7100
4	7	{Rn} {PARo-fAPAR} {H-LE-NEE} {ustar}	0.7944	0.6481	0.7246	0.9025	0.2813	0.7254	0.8248	0.7002

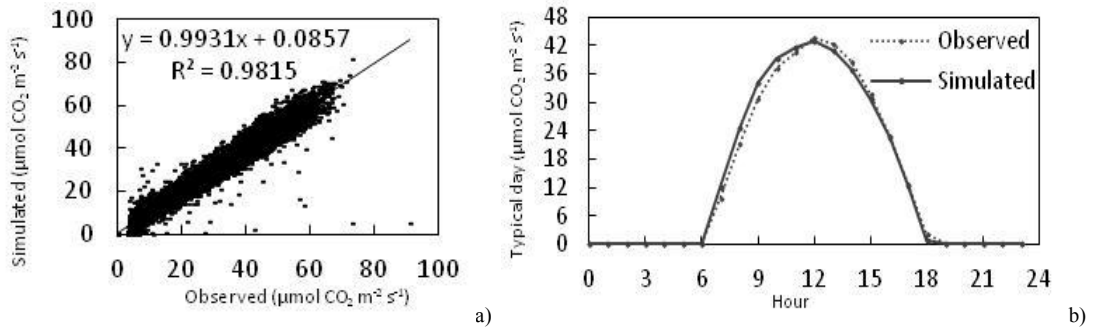


Figure 11: PARo results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figure 12 shows fAPAR after the hierarchical calibration. Monthly data presented good fit, with overestimated results during the winter (Figure 12-b). The regression coefficients were  $a = 0.76$  and  $b = 0.23$  (Figure 12-a).

Figure 14 shows the multi-objective calibration of  $u^*$ . As observed for the mono-objective calibration, the model overestimated  $u^*$  during the day and underestimated  $u^*$  during the night, with a delay of one hour in the maximum value (Figure 14-b). The regression coefficients were  $a = 0.44$  and  $b = 0.16$  (Figure 14-a).

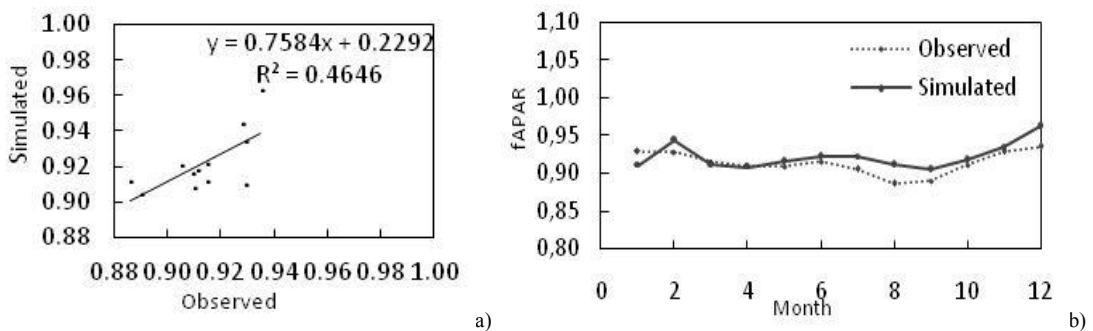


Figure 12: fAPAR results after the multi-objective calibration a) Dispersion plot and b) seasonal variation.

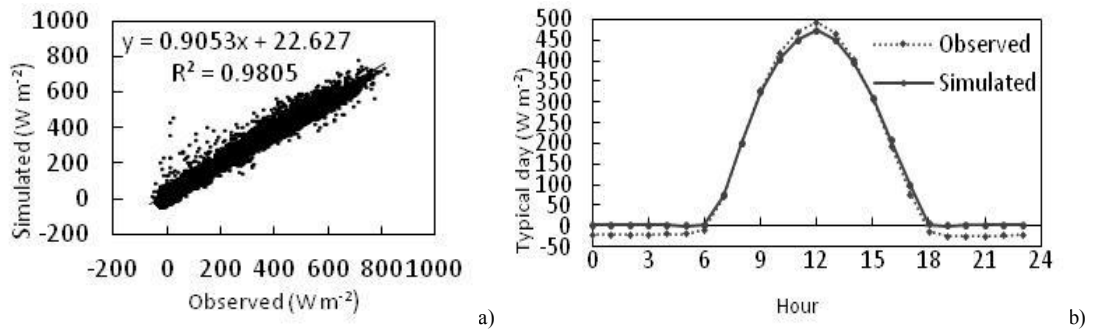


Figure 13: Rn results after the multi-objective calibration a) Dispersion plot and b) typical day.

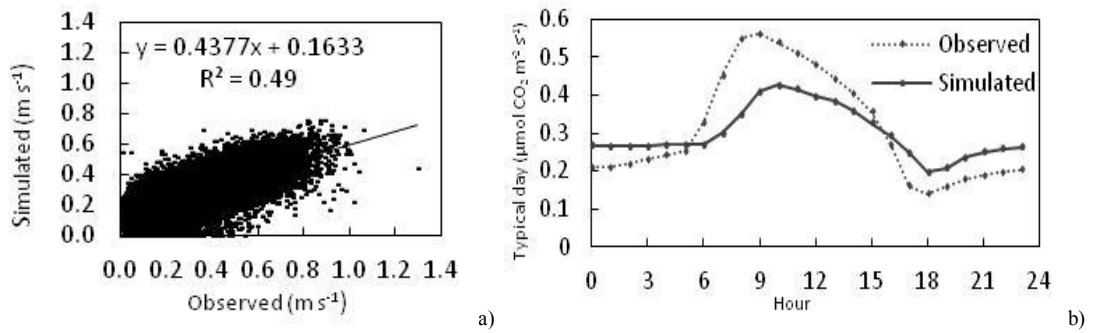


Figure 14:  $u^*$  results after the multi-objective calibration a) Dispersion plot and b) typical day.

Figures 15 and 16 show the multi-objective calibration of H and LE, respectively. The bias between observed and simulated data was noticeable and the overestimation of H and LE during the day was approximately  $140 \text{ W.m}^{-2}$  (Figures 15-b and 16-b). The regression coefficients were  $a = 1.68$  and  $b = 47.60$  for H (Figure 15-a) and  $a = 0.43$  and  $b = 24.02$  for LE (Figure 16-a).

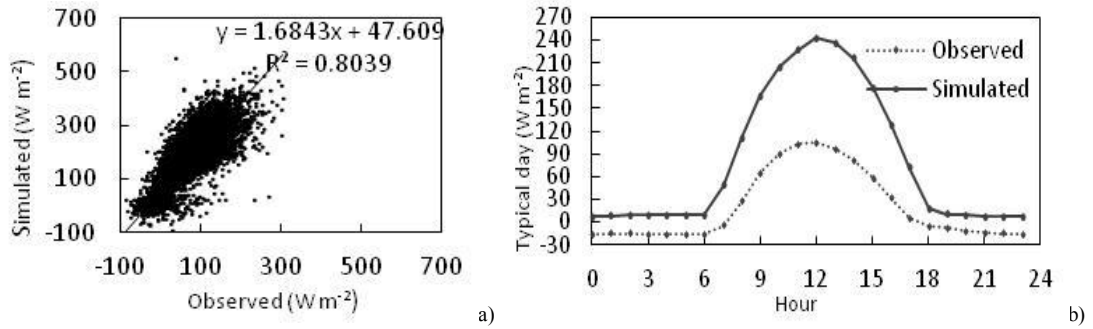


Figure 15: H results after the multi-objective calibration a) Dispersion plot and b) typical day.

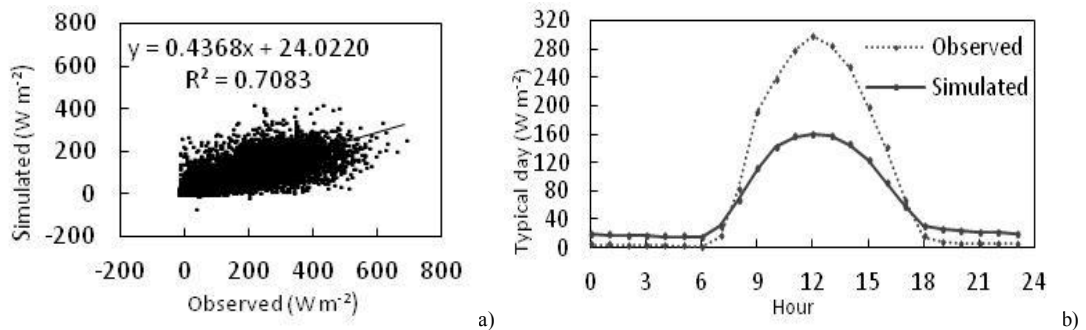


Figure 16: LE results after the multi-objective calibration a) Dispersion plot and b) typical day.

The results for H, LE and NEE were remarkably different from those obtained for the mono-objective optimization. All three variables must be optimized simultaneously, in the third hierarchical level, because the three output variables are sensitive to the same parameters. It was observed earlier that the simultaneous optimization usually decreases the algorithm's performance, and this became very clear when comparing Figures 15-17 to Figures 6-8. Because of several constraints in this optimization (including H, LE and NEE in the same hierarchical level and energy conservation by the model), the optimizer chose parameters that reduced stomatal conductance and increased canopy temperature (results not shown).

A curious result became apparent when radiative fluxes were optimized together with turbulent fluxes: the performance of the model to simulate H dropped considerably ( $f_H^{\text{mono}}/f_H^{\text{multi}}$  was in the range of 0.27 in Table 9, whereas it ranged from 0.71 to 0.99

in Table 8). Although the drop in the ratio  $f^{\text{mono}}/f^{\text{multi}}$  is an expected characteristic of the hierarchical multi-objective optimization as the number of objectives increase, in no other optimized variable the performance presented such a significant drop. In addition to the constraints already discussed, an important characteristic of the observed and the simulated data was co-responsible for this result: whereas the model must conserve energy ( $R_n - G = LE + H$ ), the observed energy flux data obtained from net radiometers and eddy covariance systems is rarely conservative, and  $R_n - G$  is usually greater than  $LE + H$  (See Twine et al. [2000] for a general discussion, or Costa et al. [2010] for a discussion specific for this site).

Given these errors in the data, the optimizer tends to make different choices in the mono or multi-objective optimizations. In the H mono-objective optimization, there is no constraint on  $R_n$  simulation, so the optimizer can freely modify it to provide the adequate net radiation for the correct simulation of H. In the multi-objective optimization, simulated  $R_n$  matches the observations, which eliminates the possibility of any energy adjustments, and more radiative energy must be portioned between the two turbulent fluxes ( $G$  is about 1% of  $R_n$ , so it was not considered in this discussion).

An attempt to compensate this problem in energy balance closure was made by using only  $R_n$ ,  $LE$  and  $H$  values in days when  $LE + H$  was within 40% of  $R_n$  (see discussion in Costa et al. [2010]). However even after filtering, in most days,  $R_n$  was significantly different from  $LE + H$ . A better approach to correct this problem may be to correct  $LE$  and  $H$  according to the observed Bowen ratio, so that long-term  $LE + H$  matches long-term  $R_n$ , as proposed by Twine et al. [2000]. This would ensure energy conservation in the long-term observed data.

Figure 17 shows the multi-objective calibration of NEE. The model bias at midday was about  $7 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  and there was an advance of one hour in the minimum value (Figure 17-b). The regression coefficients were  $a = 0.59$  and  $b = 0.12$  (Figure 17-a). Being influenced by the stomatal conductance in the same way as  $LE$ , the factors that affect the correct partition of energy between  $LE$  and  $H$  also affect NEE.

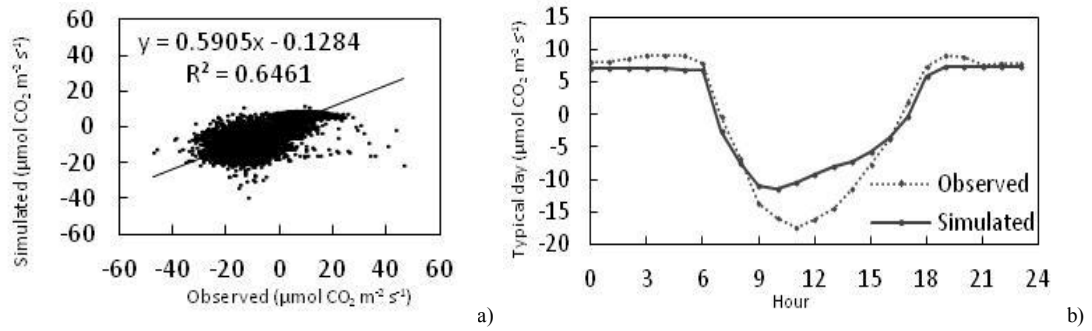


Figure 17: NEE results after the multi-objective calibration a) Dispersion plot and b) typical day.

#### 4.4 Discussion on the performance index

Figure 18 shows all results of the performance index  $D$  versus the number of optimized outputs, separated by the number of hierarchical levels, from 1 to 4. Four conclusions can be drawn from this figure.

First, it is generally true that, for the same number of optimized outputs, the higher the number of hierarchical levels used, the better the performance of the optimized model (higher  $D$ ). This is in agreement with the detailed analysis provided in Section 4.3.1.

Second, the best possible model's performance is very high ( $D > 0.98$ ) for selected configurations for  $n \leq 3$ , and when  $n = h$ . For  $n > 3$ , it was not possible to avoid constraints in the experiment design, and these constraints reduced model's performance. Maximum model's performance (maximum  $D$ ) dropped about 1% per additional objective function added (from 1.00 to 0.98) in the range from  $n = 1$  to  $n = 3$ , about 4.5% per additional objective function added (from 0.98 to 0.89) in the range from  $n = 3$  to  $n = 5$ , and about 9% per additional objective function added (from 0.89 to 0.71) in the range from  $n = 5$  to  $n = 7$  (Figure 18). Overall, when compared to mono-objective calibrations, the index  $D$  of multi-objective calibrations dropped about 30% for seven variables. This performance is satisfactory, considering that the use of many variables provides a more reliable model, because all the variables are optimized to correctly simulate all the implemented processes.



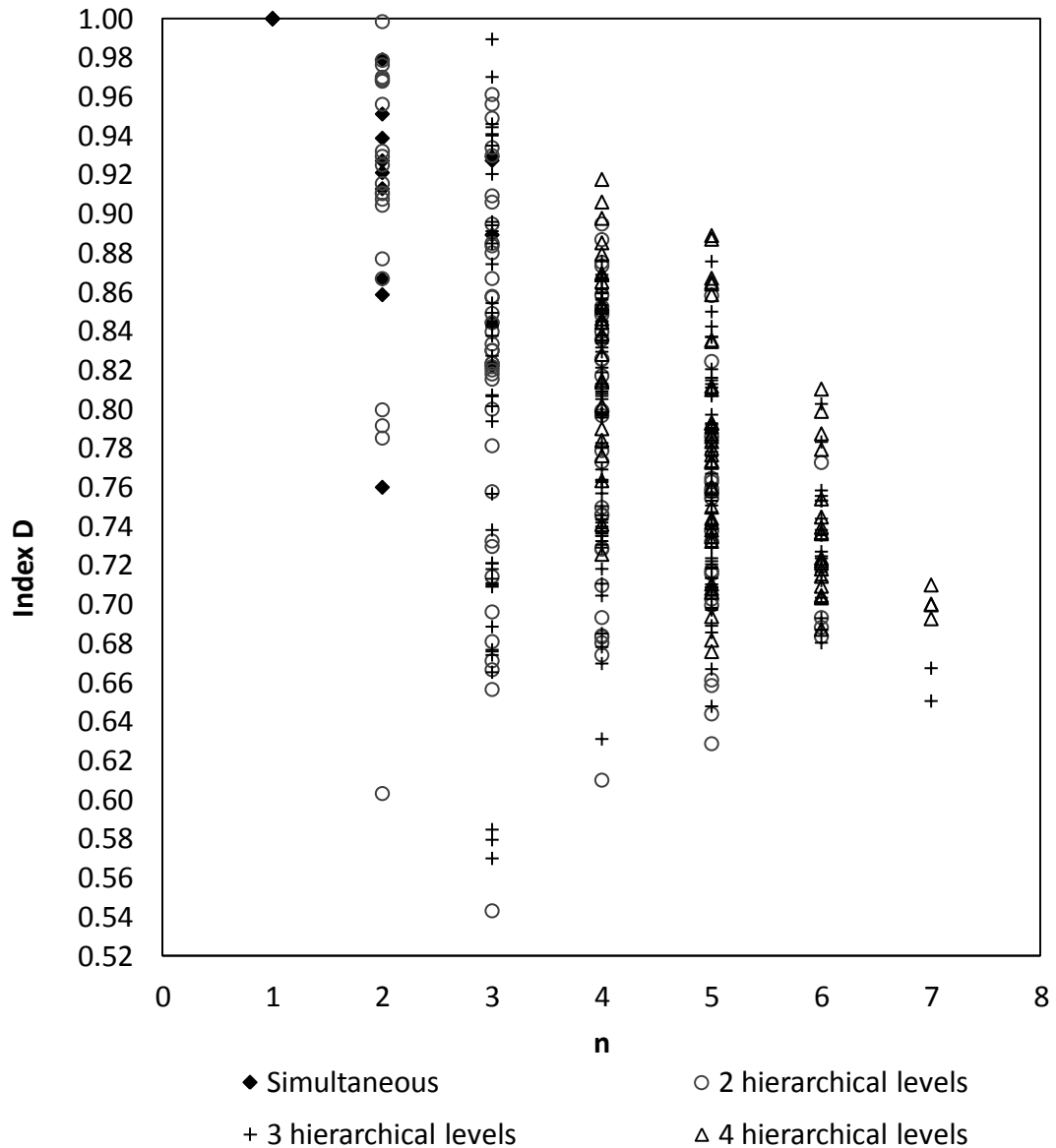


Figure 18: Index D for radiative and turbulent flux output variables versus the number of optimized outputs (n), separated by the type of optimization (simultaneous or hierarchical).

Third, results indicate that D drops at faster rates if n increases. This seems to be a consequence of the constraints introduced by the need to simultaneously optimize specific variables, notably fAPAR and PAR<sub>o</sub>, and NEE, H and LE, which are sensitive to the same parameters. Further research on the criteria to select the parameters to be optimized at each hierarchical level may remove some of these constraints, and may help improve the model results for a large value of n.

Fourth, as previously discussed, constrains and a poor optimization configuration can result in poor model's performance (as low as 50% of the mono-objective one).

## 5. CONCLUSIONS

An evaluation of the efficiency of an algorithm of hierarchical calibration, which is based in the temporal organization of natural ecosystems and of the IBIS land surface/ecosystem model was carried out. The automatic system of calibration (OPTIS) is based in the multi-objective genetic algorithm NSGA-II and enables the calibration of many (more than 3) output variables. The application of OPTIS allows for the general calibration of the model, adjusting the variables simulated by IBIS in a more reliable way. This study used the sensitivity analysis developed by Morris to select the most important parameters for calibration. Sensitivity analysis is the most important step for a well fit model, because the results are highly dependent on the distribution of the parameters in the hierarchical levels. Furthermore, in a mono-objective calibration, the variable fits better to the data if it is calibrated using only the sensitive parameters and not all the available ones.

The index D was used to evaluate the performance of the multi-objective calibrations independently of the errors embedded to the results due to model limitations. First, the hierarchical calibration was compared to the simultaneous (traditional) technique and the results showed that hierarchical calibration outperformed the simultaneous calibration in most of the situations tested. In general, the higher the number of hierarchical levels, the better the model's performance, quantified by the index D. As a second step, the basic characteristics (slope and concavity) of the curve  $D \times n$  were determined. Some constraints introduced by the need of simultaneous optimization

(radiative flux variables PAR<sub>0</sub> and fAPAR, and turbulent flux variables H, LE and NEE) and the difficulty in selecting parameters to be optimized in each hierarchical level, resulted in a faster drop of D as n increased. The decrease in the slope of the D x n curve is desirable, since the use of more measurements at different time scales is the only way to comprehensively improve models.

Model's performance depends on the optimization configuration. As a first rule, this study identified that the configuration responsible for the best model fit occurs if R<sub>n</sub> is calibrated before PAR<sub>0</sub> and fAPAR, and u\* is calibrated after H, LE and NEE. However, the model's performance after optimization is extremely sensitive to variations in the allocation of parameters to each hierarchical level. The choice of these parameters is in principle governed by the results of the sensitivity analysis. Nevertheless, the qualitative nature of the Morris method, its inability to evaluate the interactions among parameters, as well as its random nature, limited the role of the SA in the parameter decision process, and considerable modeler intuition was used to allocate the parameters to each hierarchical level. Although it is likely that the index D for model's performance will always decrease with the increase of n, it may be possible to keep the decrease rate in the range of 1% of D per objective added to the problem (per unitary increase in n). This was the rate obtained in this study before constraints started to decrease the optimizer's performance, and it may be extended to higher values of n if the following limitations are overcome:

First, given the importance of choosing the right parameters at each hierarchical level, further research is necessary about the criteria of parameter selection at each hierarchical level, including other sensitivity analyses techniques that account for the interaction among parameters;

Second, a sensitivity analysis that provides quantitative, comparable measures across parameters could also be the solution for the constraints that arise when two or more variables are sensitive to the same parameters;

Third, the performance of the objective functions calibrated at the initial hierarchical levels depends on the model baseline parameters. This dependence could be minimized by using an iterative hierarchical calibration. Multiple interactions may be necessary to evaluate solution convergence. A first iteration would choose well-

adjusted parameters to be the baseline for the second iteration; the process could be iterated until increases in D are considered negligible;

Fourth, other sources of error in LSEMs must be identified, quantified, and corrected. In particular, model limitations such as the lack of representation of important processes may weaken the entire calibration process. A poor model performance during the mono-objective calibration could be indicative of poor model formulation. After improvements in the code are made, new parameters can be included in the list of calibrated parameters;

Fifth, some sites may present a single measurement of several important variables, usually slowly-varying variables that are important for slow processes (NPP, biomass, etc.). Although the model's long-term performance may improve if these variables are included as objective functions, such procedure considerably decreases the index D (the mono-objective optimization has no problem in exactly matching a single observation, and hence the ratio  $f^{\text{mono}}/f^{\text{multi}}$  becomes zero). A possible solution could be to include these measurements collected at several sites.

Finally, reference data must meet strict quality criteria to avoid contamination of model parameters. In particular, flux data should be corrected to ensure energy conservation.

Based on this study, it is possible to identify several subjects for further research. First, at site-level scales, all the considerations made in this study regarding the SA must be thoroughly studied to reduce the constraints that limit optimization performance. Second, the multi-site optimization enables the use of variables that have a single measurement at each site such as NPP or biomass. Third, the study of the effect of vegetation composition considering the comparison between single species sites (pastures) versus multiple species sites (forests) [Fischer et al., in press] is recommended. One additional topic for future research using the hierarchical optimizer is the integration of site-level regional optimization using regional estimates from remote sensing and field data networks, such as the RAINFOR network - an international effort to monitor structure, composition and dynamics of the Amazonian forest in order to better understand their relationship to soil and climate [Malhi et al., 2002]. Data from hydrologic networks could also be used, if the LSEM is coupled to a river routing model.

The power of the hierarchical calibration to use a high number of objective functions is highly dependent on the removal of constraints for model's performance. After these constraints are removed, the model may be able to achieve a high performance for large values of  $n$ .

## REFERENCES

- Coello, C.A.C., 2006. Evolutionary multi-objective optimization: A Historical view of the field. *IEEE Comput. Intell. Mag.* 1(1), 28-36.
- Costa, M.H., Biajoli, M.C., Sanches, L., Hutyra, L.R., da Rocha, H.R., Aguiar, R.G. de Araújo, A.C., 2010. Atmospheric versus vegetation controls of amazonian tropical rainforest evapotranspiration: Are the equatorial and tropical rainforests any different?. *J. Geophys. Res.* 115, G04021.
- Czitrom, V., 1999. One-Factor-at-a-Time versus designed experiments. *The American Statistician* 53 (2), 126-131.
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multi-objective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation* 6 (2).
- Fischer, G.R., Costa, M.H., Murta, F.Z., Malhado, A.C.M., Aguiar, L.J.G., Ladle, R.J., 2013. Multi-site land surface model optimization: An exploration of objective functions. *Agric. Forest Meteorol.*, [dx.doi.org/10.1016/j.agrformet.2012.11.021](https://doi.org/10.1016/j.agrformet.2012.11.021).
- Foley, J.A., 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.

Fonseca, C.M., Fleming, P.J., 1996. On the performance assessment and comparison of stochastic multiobjective optimizers. *Parallel Problem Solving from Nature IV*, H.-M. Voigt, W.Ebeling, I. Rechenberg, and H.-P. Schwefel, Eds. Berlin, Germany: Springer-Verlag, 584-593.

Fox, D.G., 1980. Judging air quality model performance, *Bulletin of American Meteorological Society* 62, 599–609.

Fu, M.C., Glover, F.W., April, J., 2005. Simulation optimization: A review, new developments, and applications. In: *Proceedings Of The 2005 Winter Simulation Conference*, edited by Kuhl M, Steiger N, Armstrong F, Joines J. Institute of Electrical and Electronics Engineers, 83–95.

Goldberg, D.E., 1989. *Genetic Algorithms in Search Optimization and Machine Learning*. Addison Wesley, 41. ISBN 0-201-15767-5.

Groenendijk, M., Dolman, A.J., Van Der Molen, M.K., Leuning, R., Arneth, A., Delpierre, N. et al., 2011. Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. *Agricultural and Forest Meteorology* 151, 22–38.

Gupta, H.V., Sorooshian, S., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable metrics of information. *Water Resour. Res.* 34(4), 751–763.

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment* 32, 135-154.

Hawkins, D.M., 2004. The problem of overfitting. *Journal of Chemical Information and Computer Sciences* 44(1), 1-12.

Konak, A., Coit, D.W., Smith, A.E., 2006. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety* 91, 992–1007.

Kucharik, C.J., Foley, J.A., Delire, C., Fisher, V.A., Coe, M.T., Lenters, J.D., Young-Molling, C., Ramankutty, 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.

Lai, Y., 1996. Hierarchical optimization: A satisfactory solution. *Fuzzy Sets and Systems* 77, 321-335.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35(1), 233–241.

Luan, J., Muetzelfeldt, R.I., Grace, J., 1996. Hierarchical approach to forest ecosystem simulation. *Ecol. Model.* 86, 37–50.

Mackay, D.S., Samanta, S., Ahl, D.E., Ewers, B.E., Gower, S.T., Burrows, S.N., 2003. Automatic parameterization of land surface process models using Fuzzy logic. *Transactions in GIS* 7(1), 139–153.

Malhi, Y., Phillips, O.L., Lloyd, J., Baker, T., Wright, J., Almeida, S., Arroyo, L., Frederiksen, T., Grace, J., Higuchi, N., Killeen, T., Laurance, W.F., Leño, C., Lewis, S., Meir, P., Monteagudo, A., Neill, D., Núñez Vargas, P., Panfil, S.N., Patiño, S., Pitman, N., Quesada, C.A., Rudas-LI, A., Salomão, R., Saleska, S., Silva, N., Silveira, M., Sombroek, W.G., Valencia, R., Vásquez Martínez, R., Vieira, I.C.G., Vinceti, B., 2002. An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR). *Journal of Vegetation Science* 13(3), 439-450.

McMahon, S.M., Diez, J.M., 2007. Scales of association: hierarchical linear models and the measurement of ecological systems. *Ecology Letters* 10, 1-16.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33(2), 161–174.

Pitman, A.J., 2003. The evolution of, and revolution in, land surface schemes designed for climate models. *Int. J. Climatol.* 23(5), 479–510.

Praditwong, K., Yao, X., 2007. How well do multi-objective evolutionary algorithms scale to large problems. *Evolutionary Computation. CEC 2007. IEEE Congress on*, 3959–3966.

- Rosolem, R., Gupta, H.V., Shuttleworth, W.J., Zeng, X., de Gonçalves, L.G.G., 2012. A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis. *J. Geophys. Res.* 117, D07103. doi:10.1029/2011JD016355.
- Saltelli, A., Chan, K., Scott, M., 2000. *Sensitivity Analysis*. John Wiley and Sons Ltd.
- Saltelli, A., 2002. Sensitivity analysis for importance assessment. *Risk Anal.* 22(3), 579–590. doi:10.1111/0272-4332.00040.
- Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models. *Chem. Rev.* 105(7), 2811–2827.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global sensitivity analysis: the primer*. Wiley-Interscience.
- Sampaio, G., Nobre, C., Costa, M.H., Satyamurty, P., SoaresFilho, B.S., Cardoso, M., 2007. Regional climate change over eastern Amazonia caused by pasture and soybean cropland expansion. *Geophys. Res. Lett.* 34, L030612.
- Senna, M.C.A., Costa, M.H., Pires, G.F., 2009. Vegetation-atmosphere-soil nutrient feedbacks in the Amazon for different deforestation. *J. Geophys. Res.* 114 (D4), D04104. doi:10.1029/2008JD010401.
- Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F.G., Berry, J.A., Collatz, G.J, Denning, A.S., Mooney, H.A., Nobre, C.A., Sato, N., Field, C.B., Henderson-Sellers, A., 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science* 275(5299), 502–509.
- Schutze, O., Lara, A., Coello, C.A.C., 2011. On the influence of the number of objectives on the hardness of a multi-objective optimization problem. *Evolutionary Computation. IEEE Transactions on* 15(4), 444-455.
- Srinivas, N., Deb, K., 1995. Multi-objective function optimization using nondominated sorting genetic algorithms. *Evol. Comput.* 2(3), 221–248.
- Trudinger, C.M., Raupach, M.R., Rayner, P.J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A.D., Roxburgh, S.H., Styles, J., Wang, Y.P., Briggs, P., Barrett, D., Nikolova, S., 2007. OptIC project: An intercomparison

of optimization techniques for parameter estimation in terrestrial biogeochemical models. *J. Geophys. Res.* 112(G02027), 17. doi:10.1029/2006JG000367.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. *Agr. and Forest Met.* 103, 279-300.

Varejão-Jr., C.G., Costa, M.H., Camargos, C.C.S., 2012. A multi-objective hierarchical calibration procedure for land surface/ecosystem models. *Inverse Problems in Science and Engineering*, ahead-of-print, 1-30.

Veldhuizen, D.A.V., Lamont, E.G.B., 2000. Multi-objective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary Computation* 8 (2), 125–147.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multi-objective optimization of hydrologic models. *Water Resources Research* 39 (8), 1214. doi:10.1029/2002WR001746.

Willmott, C.J., 1981. On the validation of models. *Physical Geography* 2 (2), 184–194.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin American Meteorological Society* 63 (11), 1309–1313.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research* 30, 79–82.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90 (C5), 8995–9005.

Wu, J., David, J.L., 2002. A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling* 153, 7-26.