Modeling sugar cane yield with a process-based model from site to continental scale: uncertainties arising from model structure and parameter values.

Valade A.[a], Ciais P. [a], Vuichard N. [a], Viovy N. [a], Huth N. [b] , Marin F. [c], Martiné J.-F. [d]

[a]LSCE, CEA-CNRS, Gif-sur-Yvette, 91191 France

[b]CSIRO. Ecosystem Sciences, PO Box 102, Toowoomba, Qld, 4350, Australia

[c] EMBRAPA Informatica Agropecuria, Barão Geraldo, 13083-886 Campinas SP, Brazil

[d] CIRAD, UR SCA, Saint-Denis, La Réunion, F-97408 France

Abstract

Agro-Land Surface Models (agro-LSM) have been developed from the integration of specific crop processes into large-scale generic land surface models that allow calculating the spatial distribution and variability of energy, water and carbon fluxes within the soil-vegetation-atmosphere continuum. When developing agro-LSM models, a particular attention must be given to the effects of crop phenology and management on the turbulent fluxes exchanged with the atmosphere, and the underlying water and carbon pools. A part of the uncertainty of Agro-LSM models is related to their usually large number of parameters. In this study, we quantify the parameter-values uncertainty in the simulation of sugar cane biomass production with the agro-LSM ORCHIDEE-STICS, using a multi-regional approach with data from sites in Australia, La Réunion and Brazil. In ORCHIDEE-STICS, two models are chained: STICS, an agronomy model that calculates phenology and management, and ORCHIDEE, a land surface model that calculates biomass and other ecosystem variables forced by STICS phenology. First, the parameters that dominate the uncertainty of simulated biomass at harvest date are determined through a screening of 67 different parameters of both STICS and ORCHIDEE on a multi-site basis. Secondly, the uncertainty of harvested biomass attributable to those most sensitive parameters is quantified and specifically attributed to either STICS (phenology, management) or to ORCHIDEE (other ecosystem variables including biomass) through distinct Monte-Carlo runs. The uncertainty on parameter values is constrained using observations by calibrating the model independently at seven sites. In a third step, a sensitivity analysis is carried out by varying the most sensitive parameters to investigate their effects at continental scale. A Monte-Carlo sampling method associated with the calculation of Partial Ranked Correlation Coefficients is used to quantify the sensitivity of harvested biomass to input parameters on a continental scale across the large regions of intensive sugar cane cultivation in Australia and Brazil. Ten parameters driving most of the uncertainty in the ORCHIDEE-STICS modeled biomass at the 7 sites are identified by the screening procedure. We found that the 10 most sensitive parameters control phenology (maximum rate of increase of LAI) and root uptake of water and nitrogen (root profile and root growth rate, nitrogen stress threshold) in STICS, and photosynthesis (optimal temperature of photosynthesis, optimal carboxylation rate), radiation interception (extinction coefficient), and transpiration and respiration (stomatal conductance, growth and maintenance respiration coefficients) in ORCHIDEE. We find that the optimal carboxylation rate and photosynthesis temperature parameters contribute most to the uncertainty in harvested biomass simulations at site scale. The spatial variation of the ranked correlation between input parameters and modeled biomass at harvest is well explained by rain and temperature drivers, suggesting climate-mediated different sensitivities of modeled sugar cane yield to the model parameters, for Australia and Brazil. This study reveals the spatial and temporal patterns of uncertainty variability for a highly parameterized agro-LSM and calls for more systematic uncertainty analyses of such models.

1 Introduction

In the recent years, many governments have set targets in terms of biofuels consumption for transportation fuel (Sorda et al., 2010), resulting in a large increase in bioenergy cropping area around the world. Concerns about energy shortage, policy to reduce CO_2 emissions, and the search for new income for farmers can explain why energy policies have considered biofuels as a serious alternative to fossil fuel in many countries (Demirbas, 2008). Yet, the claimed benefits of biofuels for fossil fuel substitution have been questioned in terms of their net effect on atmospheric CO_2 and climate, and even of their economic return (Doornbosch and Steenblik; Naylor et al., 2007). In particular, the conditions of biofuel cultivation, such as the type of crop, practice, previous land use, and local climate, have emerged as key factors that determine the effectiveness of their carbon emissions reduction (Fargione et al., 2008; Hill et al., 2006; Searchinger et al., 2008). At the heart of biofuel cultivation is ethanol that represents today 74% of the energy content of the world production of liquid biofuels (Howarth et al., 2008) and whose production is expected to double between 2011 and 2021 (OECD, 2012), hence the urgency to better quantify and understand regional potentials of bioethanol crops. Based on recent life cycle analysis studies (de Vries et al., 2010; Schubert, 2006; von Blottnitz and Curran, 2007), ethanol from sugar cane is the most competitive in terms of energy use and net carbon balance and the energy use projections from the International Energy Agency foresee that by 2050, sugar cane is the only 1st generation biofuel that that will keep expanding (IEA, 2011).

The impact of sugar cane expansion on climate and carbon balance is under scrutiny with different approaches. Satellite observation data have been used to study biophysical effects of sugar cane expansion on local temperature in the Brazilian Cerrado (Loarie et al., 2011) Survey for agricultural and industrial performances from sugar cane mills have allowed Macedo et al. (2008) to establish the carbon balance of sugar cane ethanol production in the Center-South of Brazil. Georgescu et al. (2013) simulate the hydroclimatic impacts of sugar cane expansion by forcing sugar cane land cover characteristics into a regional climate model. All approaches provide useful information on impacts and potentials but are impractical to apply outside of the regions and conditions (climate, management) where they have been conducted.

In parallel with empirical approaches, significant progress has been made towards mechanistic modeling of sugar cane yields using models. Crop models are generally used to simulate sugar cane production at site scale, with specific parameters (Cheeroo-Nayamuth et al., 2000). Land surface models (LSM) are rather used to estimate the spatial distribution of crop productivity under different soil and climatic conditions, over a region or even over the globe but with a simpler and generic description of sugar cane plants (Black et al., 2012; Cuadra et al., 2012; Lapola et al., 2009). Agro-LSM models stand at the interface between plot-scale crop models and global LSMs. Yet, as highlighted by Surendran Nair et al. (2012) if the development of agro-LSM models for biofuels has been the subject of much interest recently, detailed parameterization, validation and uncertainty quantification is still very limited in regional and global applications, and efforts must be made in that direction. The importance of evaluating and communicating about global models uncertainty was as well emphasized within the framework of the model inter-comparison project

AgMIP - providing insights for IPCC AR5 report – in which crop models uncertainty is identified as a key theme of interest that was only little explored so far (Rosenzweig et al., 2013). ORCHIDEE-STICS (Gervois et al., 2004) is an agro-LSM model that has been developed from the coupling of the agronomical model STICS (Brisson et al., 1998) and the Land Surface Model ORCHIDEE (Krinner et al., 2005) and that has been applied for studies from site to continent mainly for temperate crops in Europe (Gervois et al., 2008) and has been recently adapted to sugar cane simulation (Valade et al., 2013).

Four uncertainty sources affect the simulation of sugar cane biomass with ORCHIDEE-STICS: 1) input uncertainty on boundary conditions used for climate drivers and soil properties, 2) structure uncertainty related to model equations and parameterizations, 3) parameters value uncertainty, and 4) uncertainty associated with the measurements used for model evaluation or calibration. Here we focus on structure and parameters uncertainty and try to estimate how these two sources of uncertainties affect the simulations of sugar cane harvest biomass. We want to determine which parameters are responsible for most of the uncertainty in harvest biomass (screening analysis) and to what extent this is related to the model's structure (uncertainty analysis). In addition, we want to quantify this uncertainty and examine its temporal and spatial variability (sensitivity analysis).

In the following, we first present the sites and regions considered in this study (section 2.1) and the main features of the ORCHIDEE-STICS model (section 2.2). We then describe the screening algorithm used to sort the most important parameters (section 2.3), and the uncertainty and the sensitivity analyses (sections 2.4 and 2.5). Then we discuss the results of the screening analysis, in terms of the parameters identified by the screening as the most important for controlling harvested sugar cane biomass (section 3.1). We describe the results for the measure of the uncertainty calculated for 7 sites in section 3.2 to 3.4 and present maps of the sensitivity of the model to its main parameters in section 3.5.

2 Materials and methods

In this study, we aim to quantify the uncertainty related to the parameter values of a chain of two process-based models (STICS-ORCHIDEE) to simulate sugar cane yield (biomass at harvest date). This is a difficult task because this model is a detailed and complex model that contains over 100 plant specific parameters within the primitive equations of phenology, energy and water balance, photosynthesis and allocation. We perform the uncertainty analysis in three steps, illustrated in Figure 1 and consisting of screening, uncertainty and sensitivity analyses, all described in more details in section 2. These three steps are sequential and complementary. The first step is a screening to sort the most important parameters controlling yield, and to reduce the dimension of the parameter space from a large number of parameters to few key parameters, allowing a moderate number of sensitivity simulations. The screening allows the restriction of the two further steps to a smaller parameter subset. The second step is an uncertainty analysis that considers all retained parameters together with their probability distributions and determines the probability distribution for the output variable (biomass). The third step is a sensitivity analysis of the modeled spatial distribution of sugar cane yield to the model parameters for two large regions, in Brazil and Australia, at a spatial resolution of 0.7°. The sensitivity is established from the spatial distribution of ranked correlations between each parameter and yield in each grid point. Along the study steps, we address several problems inherent to uncertainty and sensitivity evaluation such as the determination of the uncertainty on the input parameters and the spatial (regional) differences of the sensitivity of the model to its key parameters.

2.1 Sites and study areas

This study is based on sugar cane field trials in three regions (figure 2) where sugar cane is of economical importance, Brazil (1 site), Australia (4 sites), and La Reunion Island (2 sites). These sites, already used by Valade et al. (2013) span different climatic conditions and agricultural practices, as shown in Table 1, which makes them useful for our purpose to provide continental-scale sugar cane yield uncertainty estimates. More details about the four sites from Australia and La Réunion can be found respectively in Keating et al. (1999); Muchow et al. (1994); Robertson et al. (1996) and in Martiné (unpublished). The site from Brazil is described in(Marin et al., 2011). The sensitivity analysis of the yield spatial distribution to the model parameters is carried out for two continental-scale areas where sugar cane is cultivated at large scale. In Brazil, we consider the region encompassing partly the Sao Paulo and Mato Grosso states,and in Australia the sugar cane cultivation belt of the northeastern coast (Figure 2).

2.2 Model & parameters considered

We use the agro-Land Surface Model ORCHIDEE-STICS (Gervois et al., 2004) in a version that was already calibrated for sugar cane for Leaf Area Index at the same sites than used here (Valade et al., 2013). This model chains the crop model STICS with sugar cane specific phenology and management with the generic process-based land surface model ORCHIDEE that can be applied either at a site, or on a grid for regional runs.

STICS (Brisson et al., 1998) is an agronomical model designed for site-scale operational applications, which describes in details the soil and crop processes associated with specific crop varieties and with management practices, such as aboveground biomass, and biomass nitrogen content, water and nitrogen content in the soil, yield, root density. Yet, STICS is a generic crop model, because from a set of common equations it can describe a large number of crop species through specific parameterizations. Similarly, specific vectors of parameters define crop cultivars. STICS has been validated for a variety of cropping situations (Brisson et al., 2003)

ORCHIDEE (Krinner et al., 2005) is a land surface model developed for global applications, standing now as the land surface model of the IPSL Earth System Model. It has been developed from the association of a surface energy and water balance scheme (SECHIBA) with a biogeochemistry module (STOMATE) and as such simulates the short time scale exchanges of water and energy between the land surface and the atmosphere, as well as the processes of the carbon cycle including photosynthesis, respiration, carbon allocation, soil decomposition. The vegetation is represented in ORCHIDEE with the Plant Functional Type (PFT) concept, by grouping species into a few categories based on the similarities of their traits and resulting in an average plant. For example, sugar cane would fall in the generic 'C4 crop' PFT in the standard version of ORCHIDEE, and this un-calibrated version of model fails to reproduce site-level phenology, as shown by Valade et al. (2013)

The chaining of STICS with ORCHIDEE was performed to improve the ability of ORCHIDEE to simulate specific crops, for which the PFT concept was not appropriate, as it lacks representation of crop phenology and crop management practices (Gervois et al., 2004). In the chain-like structure (Figure 3), STICS calculates phenology, water and nitrogen requirements, and passes the key variables of Leaf Area Index (LAI), root profile and nitrogen stress as well as the input data concerning irrigation requirements to ORCHIDEE that uses them to calculate carbon assimilation and allocation, water balance, and energy-related variables. The one-way coupling between the two models can generate some inconsistencies, such as the soil status that is different between ORCHIDEE and STICS. This type of inconsistencies, inherent to the structure of the model is considered as part of the structural uncertainty and is not covered in this study. However, this particular one-way structure will have a consequence in the uncertainty that we are analyzing in this study.

ORCHIDEE and STICS each have a large number of parameters involved at every step of a simulation over the course of a growing season. The values of these parameters - often empirically prescribed - are not easy to measure or are not measurable at all, calling in many cases for expert judgment to set their values, when it is impractical to find reference values. The uncertainty of these parameters is propagated onto the output variables of ORCHIDEE STICS and has impacts which strength depends on the structure of both STICS and ORCHIDEE. Because of the chain-type structure of ORCHIDEE-STICS (fig.3), the parameters from STICS that control LAI and nitrogen stress are expected to have a weaker and more indirect effect on downstream variables such as biomass compared with parameters from ORCHIDEE that directly control carbon assimilation processes and the development of biomass to produce yield at the date of harvest.

2.3 Parameter screening

In this section, we describe the screening step that allows us to select the most influential parameters upon which the model uncertainty is investigated. An initial set of 17 parameters from ORCHIDEE and 50 parameters from STICS is considered for the screening, according to their influence on the simulation of biomass production, based on expert knowledge and literature as listed in Table 2. The screening analysis procedure is the same as described in (Valade et al., 2013). It is based upon the method of Morris (Campolongo et al., 2007; Morris, 1991; Pujol, 2009) often used to explore the parameters space for complex models with a large number of parameters. Like all screening methods, the Morris method gives qualitative information on the sensitivity of the output variables to the parameters, since it only discriminates parameters based on their importance, but does not provide information on the relative difference of importance (Cariboni et al., 2007). Its aim is to reduce the dimensionality of the problem for further use of quantitative, computationally heavier methods (Saltelli et al., 2004).

The advantage of the Morris method is that it is computationally efficient and easy to implement and interpret. It is based on a one-at-a-time approach, in which only one parameter is changed between two runs, allowing for the calculation of a local partial derivative of the output variable with respect to the input parameter, called an elementary effect. The Morris method is considered to be a “global” screening method, because the algorithm is repeated several times to calculate the elementary effects of each parameter in several locations of the parameters space so that the average and standard deviation of all elementary effects associated with each parameter are representative of the behavior of this parameter in its whole range of variation. The results of the Morris screening algorithm can be represented by a 2-D plot of standard deviation versus mean value of the elementary effects on the output variable (here harvested biomass) of each parameter. A parameter with a high mean elementary effect (called μ, or μ^* for mean of absolute values) is interpreted as a parameter with high influence on the output harvested biomass variable. A parameter with a high standard deviation of its elementary effects (σ) is interpreted as inducing non-linearities in the model output, and/or as having interactions with other parameters.

Here, we apply the Morris method as implemented in the R 'sensitivity' package (Pujol et al., 2013) using site-scale simulations of ORCHIDEE STICS across the 7 field trial sites listed in Table 1. For each site, we identify the most influential parameters for the output variable harvested biomass. The parameters identified as important at least at two sites are selected for the rest of the study.

2.4 Uncertainty analysis (UA)

The goal of the UA is to quantify the overall uncertainty in the harvested biomass output variable that results from uncertain input parameter values. Firstly, based on the a priori probability of each parameter's value, a Probability Density Function is assigned to each parameter in order to generate sample parameter sets according to the Latin Hypercube Sampling (LHS) method. Secondly, an ensemble of model runs is performed using those samples. Thirdly, the uncertainty on the output variables is obtained from the statistical properties of the distribution of simulated harvested biomass from the ensemble runs by defining the uncertainty as one standard deviation of the distribution.

The first step is thus to generate parameters samples constrained with prior parameters ranges and statistical distributions that are then used as inputs for ensemble simulations.

The parameters considered for the uncertainty (UA) for both STICS and ORCHIDEE are those selected by the screening analysis, allowing a reduction in the parameters space hypercube dimensionality and therefore in the required computing resources. Starting from the initial set of 17 and 50 parameters respectively for the screening of ORCHIDEE and STICS parameters, the Morris algorithm result (see Section 3.1) allows us to reduce the parameter numbers to 8 and 3 parameters for ORCHIDEE and STICS, respectively.

For the UA, we use Monte-Carlo methods, which are less computationally expensive than variance-based approaches (Marino et al., 2008), making them a frequent choice in environmental sciences (Poulter et al., 2010; Verbeeck et al., 2006; Zaehle et al., 2005). The Monte-Carlo sampling scheme used here is the stratified LHS, which is an efficient scheme for generation of multivariate samples of statistical distributions (McKay et al., 1979) In LHS, the range of each of the k parameters X_1,X_2, …X_k included in the study is divided into N intervals of equal probability. One value is randomly selected from each interval. The N values obtained for the X_1 parameter are then paired at random, without replacement, with the N values obtained for the X_2 parameter, then to the N values obtained for the X_3 parameter and so on until the k^{th} parameter. The procedure results in N sets of k parameters, or samples, that can be used for input to the model. In this study, from the 11 parameters identified by the screening, the N value is set to 250 resulting in 250 simulations for exploring the uncertainty around modeled biomass for each site.

In order to get insights on the part of the uncertainty attributable to each of the two models chained together, STICS and ORCHIDEE (fig.1), first, only the uncertainty coming from ORCHIDEE parameters is evaluated (fig.1), secondly, only the uncertainty propagated from STICS parameters (fig.1), and last, uncertainties propagated from both ORCHIDEE and STICS parameters are considered together through the chained model ORCHIDEE-STICS.

An important difficulty in the utilization of sampling-based UA methods is the lack of literature about a priori probability distribution of most parameters, given the dependency of output upon a priori assigned values (Marino et al., 2008) If most studies rely on a thorough literature search and expert judgment (Medlyn et al., 2005; Verbeeck et al., 2006; Wang et al., 2005), this approach might result in an overestimation of the model output uncertainty due to combinations of extreme parameters values that are not realistic and therefore excessively decrease the estimated reliability of the models. Some studies have addressed this issue by trying to rationalize the parameters ranges through benchmarking outputs (removing parameter sets resulting in values for output variables outside of a given benchmark range) or by prescribing hypothesized correlations between parameters (Poulter et al., 2010; Zaehle et al., 2005). Here, after a first estimation of uncertainty based on expert opinion for the a priori parameters range (overestimation of uncertainty), we propose a second approach to overcome the scarcity of information about parameters reference distributions by reducing the parameters a priori range based on site-optimized values, thus providing narrower and more realistic a priori ranges that are constrained by observations (likely underestimation of uncertainty).

For the first a priori estimation of parameters range, ranges and distributions are assigned to parameters based on expert knowledge and previous parameterization studies (Kuppel et al., 2012) and centered on their a priori values. The *a priori* ranges prescribed using this approach are considered as overestimations of the likely ranges for parameters' values for sugar cane because they are adapted from studies in which parameters' ranges were assigned for plant functional types instead of a single crop as is the case here and sometimes used for optimization studies therefore requiring wide enough ranges within the model's domain of applicability (Groenendijk et al., 2011; Kuppel et al., 2012). By using overestimated ranges for input parameters, we estimate an upper bound for the value of the uncertainty on output variables.

The second (site-constrained) a priori estimation is a refinement of the uncertainty estimation based on the idea that the 'real' probability distribution of the parameters can be approached by the distribution of optimal parameters over all the possible case studies (sites, weather, management). It is of course not possible to determine the model's optimal parameters for an infinite number of eco-climatic and land-management conditions, but a sample of representative case studies can provide a rough estimate of the parameters plausible range. Building on this hypothesis, the model is calibrated independently at 7 sites using an iterative method, seeking to constrain the uncertainty analysis with observation-based parameters ranges. For this, we performed a Bayesian calibration of the model parameters, using a standard variational method based on the iterative minimization of a cost function that measures both the model data misfit as well as the parameters' deviations from a prior knowledge. The iterative scheme is described in (Tarantola, 1987) with the hypothesis of Gaussian error on the observations and the parameters. At each site, parameter values are varied iteratively until the best match between simulation and observation is found. More details on the calibration results can be found in the Supporting Information. We are aware that the optimization of the parameters at 7 sites only to obtain a representative a priori range of the parameters distributions likely results into an optimistic estimate of this range even though the sites chosen cover different climatic, edaphic and management conditions making them well suited for applying our method. This observations-constrained range is highly dependent on growing conditions. When the model is applied to the context of climate change, these ranges may then be out of their domain of significance and the first wider estimate of prior parameters distribution, based on literature, must be preferred.

For both a priori parameters range estimations (expert judgment vs. site constrained), when no parameter value appears to be more likely than another, a uniform *a priori* uncertainty distribution is prescribed. When there is some level of confidence that the a priori value is more likely, we use a beta distribution. This type of distribution is often used for uncertainty analyses, because of its adjustable shape (parameterized equation) yet having the advantage of bounded tails (Monod et al., 2006; Wyss and Jorgensen, 1998). The successive analysis of both techniques provides an improvement in the estimation of the uncertainty from the first (expert-judgment based, likely too pessimistic) to the second (observation-based, perhaps too optimistic) approach.

2.5 Spatial sensitivity analysis (SA)

The first step in the sensitivity analysis also consists in generating parameters samples. The same parameters are considered for the SA as for the UA (section 2.4), i.e. the 11 parameters (8 parameters from ORCHIDEE and 3 parameters from STICS) selected by the screening analysis.

As opposed to the UA where all parameters are considered together for their effect on the distribution of the harvested biomass output variable, the goal of the sensitivity analysis is to rank the influence of parameters based on their impact on the biomass and its spatial distribution obtained in the continental-scale 0.7° runs. The partial correlation coefficient (PCC) measures the correlation between an output variable and a parameter after the correlation with other parameters has been eliminated (Marino et al., 2008). However, for monotonic but non-linear relationships, these measures perform poorly and a rank transformation needs to be applied to the data first to linearize the relationship. The correlation calculated between the rank-transformed data is then called partial rank correlation coefficients (PRCC). PRCC has been found to be an efficient indicator for the influence of parameters, because it is a measure of the sensitivity of the output to parameters (Saltelli and Marivoet, 1990). The larger the PRCC, the more important the parameter is with respect to the output variable. Here, the relationship between modeled biomass on a grid, and parameters is diagnosed through the calculation of the Partial Ranked Correlation Coefficients (PRCC) on each grid point between the output and parameter assuming a monotonic behavior of the model.

The SA is implemented from the results of the 0.7° simulations over Brazil and Australia (see fig.1 and section 3.5). In this regional sensitivity analysis, ORCHIDEE-

STICS is run for each region on a grid of 20 by 15 grid points and 13 by 20 grid points respectively, driven by gridded climate forcing fields from the reanalysis products ERA-Interim (Dee et al., 2011), with varying parameter values from a sampling where only bounds and no distributions were assigned to the parameters. The management information (date of planting, date of harvest, fertilization, irrigation) and the soil properties (as described in Valade et al. (2013)) are assumed to be uniform across each region and were defined as typical of each area. The a priori bounds used for the parameters in the SA correspond to the first version of the parameters ranges considered in the uncertainty analysis (i.e. derived from expert knowledge). As cited by Wang et al. (2005), for sensitivity analyses, Bouman (1994) advises to use parameters ranges as broad as possible within the limits of the model validity domain. Once the parameters' a priori bounds have been set, ensemble runs are performed with all the parameter sets. From the distributions of input parameters and output variables obtained at each pixel, a spatial distribution of PRCC is obtained, which is interpreted in section 3.5 in terms of regional differences of each parameter on modeled sugar cane yield.

The interest of carrying out such a regional sensitivity analysis is that it provides maps of the geographic patterns of the importance of each parameter, leading to a better comprehension of the mechanisms behind the parameter-related model sensitivity. These results can be very useful for planning purposes, for instance to quantify what are the different factors that control sugar cane yield and ethanol production over a large region under future climatic conditions as compared to present-day conditions.

3 Results and discussion

3.1 Screening

From the Morris screening method, we obtain for each parameter two indices μ^* and σ, that measure the influence of each parameter and its degree of involvement in non-linearities and interactions with other parameters, respectively. We first made sure that no parameter with a significant value for μ^* was above the line $\sigma=2\mu^*$ which would imply that non-linearities and/or interactions would be so strong that the uncertainty propagation from the parameter to the model output could not be clearly established. None of our parameters selected for their significant values of μ* was above this line (Supporting information figure 2). From μ* and σ values, we establish a ranking of the parameters by only considering parameters involved in limited interactions and/or non-linearities (σ<2μ*) and then we rank the remaining parameters based on their μ* index, a larger μ* being interpreted as a more influential parameter. The Morris parameters ranks for ORCHIDEE and STICS are respectively shown in Figure 5a and 5b where each radar plot corresponds to one model. The axes refer to the parameters and the line colors to the sites. For STICS, for the sake of readability, not all of the initially selected 50 parameters are represented on the radar plot but only those parameters that pertain to the 10 top-ranked parameters at least at one site. The maximum number of 10 parameters was fixed based on examination of Morris indices μ* and σ at individual sites that only revealed 3 to 5 sensitive parameters each time. The positions and roles in the model of the parameters identified as most important are shown in Figure 3. Figure 4 gives more details, with the main equations through which these parameters affect the output variables of STICS and of ORCHIDEE.

The 3 most influential parameters of STICS (fig.3a) reflect the way STICS and ORCHIDEE are chained (fig.3). Indeed, from the chained model structure, the indirect impact of STICS parameters on harvested biomass occurs through their effect on processes related to LAI, root growth and nitrogen stress, the only STICS variables passed to ORCHIDEE for calculating biomass. This chaining of the models through three variables is reflected in the identification of the 3 most important STICS parameters, which control the daily maximum rate of foliage production δ_{LAI}^{max}, the growth rate of the root front,κ_{root} and the threshold of nitrogen nutrition index INN_{min}. δ_{LAI}^{max} and INN_{min}parameters are both involved in LAI calculation. Indeed, the LAI equation has four members describing four processes of the sugar cane foliage development. First, the LAI-development (Δ_{LAI}^{dev} in fig.4) describes the potential LAI increase through the scaling of the daily maximum rate of foliage production by a function of the development stage (k_{LAI}), and is logically directly controlled by the value of parameter δ_{LAI}^{max}. The second member in equation (*) represents the temperature effect on LAI growth through the accumulation of degrees above a temperature threshold (T_{min} in fig.3). The last two members of the equation represent processes that can limit LAI development, competition for light between plants due to planting density (Δ_{LAI}^{dens}in fig.4) and a limitation from trophic stress emerging from competition between plant components for nitrogen based in calculation of a nitrogen nutrition index limited by parameter INN_{min}. The root growth rate κ_{root}has a less direct impact on LAI since it intervenes in the calculation of the root front depth, which then impacts the availability of nitrogen and water and therefore the stress status of the crop (impact on C_N^{plant} and $W_s$in fig.4).

The 8 most influential parameters that control harvested biomass in ORCHIDEE, are identical for all sites except at the Colimaçons site (where only 7 parameters are identified as influential by the Morris method). The Morris top ranked parameters of ORCHIDEE control photosynthesis and water budget equations as well as respiration processes (fig.4). Three of those (the minimum and optimal temperatures for photosynthesis, T_{min}, T_{opt} , the maximum rate of carboxylation V_{Cmax}^{opt}) affect directly the rate of carboxylation V_c that is calculated from the maximum rate of carboxylation weighted by a mean leaf efficiency and scaled by a limiting factor depending on the optimum and minimum temperatures for photosynthesis. The stomatal conductance g_s that links assimilation and transpiration is defined by the Ball-Berry equation (Ball et al., 1987) as a function of assimilation and depends on the air relative humidity and CO_2 concentration, scaled by a slope factor, called the Ball-Berry slope (β). The root profile constant (κ_{hum}) describes the exponential distribution of root density in the soil and is involved in the definition of available water and root temperature. Finally, the extinction coefficient (k_{ext}) intervenes in an equation derived by Monsi and Saeki (1953), similar to Beer's law, which describes the attenuation of light with depth in the canopy.

Two ORCHIDEE parameters controlling autotrophic respiration also stand out, with the maintenance respiration coefficient (α_{Mresp}) and the fraction of biomass allocated to growth respiration (f_{Gresp}). The $leaf_{age}^{crit}$ parameter that is involved in the biomass allocation also ranked high (5th most important) but only for one site and is therefore not retained for the rest of the study.

For the chained model STICS-ORCHIDEE, the 11 most influential parameters show a good agreement between sites for the most important parameters as seen on fig.5 where ranking lines overlap for most of the parameters. Building on the results of the Morris screening analysis, we select the 8 top ranked parameters for ORCHIDEE and 3 for STICS that were revealed as influential for biomass for further uncertainty and sensitivity analysis.

3.2 Uncertainty analysis: Parameters controlling biomass uncertainty at a typical site

In this section, we attribute the harvested biomass uncertainty to the uncertainty of the ORCHIDEE vs. STICS parameters. The simulated biomass uncertainty is a function of time during the growing season, and it differs between sites. In Figure 6, we show the contributions of ORCHIDEE and STICS parameters respectively to the total uncertainty for one typical site, Grafton, Australia, during the 1994-95 growing season, which has climate conditions within the range of other sites. Fig.6 a-c displays the normalized frequency distributions of simulated biomass obtained from ensemble runs for three times in the growing season: 1) very early in the cycle in fig.6a, at 100 days after planting (DAP), 2) during the peak growing season in fig.6b, at 200 DAP and 3) short before harvest in fig.6c, at 350 DAP. We distinguish between the normalized frequency distributions of simulated biomass when considering the uncertainty propagated from STICS parameters alone (green), ORCHIDEE parameters alone (yellow), and from ORCHIDEE and STICS parameters together (brown), along with their best-fit normal distributions overlaid. These distributions were obtained by Monte Carlo LHS ensemble runs (section 2.4) with a sampling of parameters of STICS alone, ORCHIDEE alone and of both models together. We consider uncertainties starting from the time when biomass reaches 50 $gC.m^{-2}$ in order to discard the emergence phase during which biomass is very low and uncertainties are therefore not significant.

At 100 DAP (Fig 6a), the uncertainty distribution of biomass related to ORCHIDEE parameters U(O) spans a slightly larger range than the distribution related to STICS, U(S), and it has more extreme values. The U(O) distribution is symmetrical around the mean value, with a standard deviation of 86.9 $gC.m^{-2}$. The U(S) distribution is non-symmetric, skewed towards larger values of biomass, and it has a slightly smaller standard deviation (76.5 $gC.m^{-2}$) than that of U(O). Combining U(O) and U(S) in Monte Carlo runs by varying the parameters of both models at the same time gives the total uncertainty distribution, U(O+S), shown in brown in fig.6. This distribution has more extreme values and a higher standard deviation (112.0 $gC.m^{-2}$), i.e. U(O+S) > U(O) + U(S).

At 200 DAP (Fig 6b), and later at 350 DAP (Fig 6c), the picture has changed. First, all uncertainties distributions are wider than at 100 DAP. Secondly, the means of U(O) and U(S) are no longer in agreement, with the asymmetric U(S) distribution being even more shifted towards high values of the harvested biomass. The reason for this shift is that among the variables transmitted from STICS to ORCHIDEE in the chain of models, the only one that can act to increase the biomass calculated by ORCHIDEE in the later phase of the growing season, near 350 DAP, is LAI. This is because a higher LAI will result into increased photosynthesis and therefore biomass in ORCHIDEE. However, passed a certain threshold, the LAI impact saturates when the foliage is sufficient for all incoming light to be captured, and therefore, uncertainty on the STICS parameters that impact LAI will not increase the uncertainty of biomass any longer. Unlike LAI, the nitrogen stress and root profile variables controlled by the parameters of STICS continue to act as limiting factors on biomass throughout the peak and late growing season. The saturation of the biomass uncertainty associated with STICS parameters is stronger at 200 DAP than at 300 DAP, when biomass increase has slowed down and the role of LAI for driving biomass is less important.

On fig.6d, the total uncertainty U(O+S) is given for the reference simulation (with parameters at their maximum likelihood values, red line) and the uncertainty on harvested biomass can be defined as a percentage of the harvested biomass in the reference simulation. For the Grafton site, at harvest, the overall uncertainty is 26.%. The relative contributions of ORCHIDEE and STICS to the total uncertainty, α_O and α_S respectively, are defined by $\alpha_O = \frac{U(O)}{U(O+S)}$,, $\alpha_S = \frac{U(S)}{U(O+S)}$. The evolution of these contributions to the total uncertainty is shown in fig.6e. We can see in this example that $U(O) > U(S)$ during the entire growing season, but with a decrease of $U(S)$, and an increase of $U(O)$ such that the increase in biomass uncertainty seen on fig.6d becomes increasingly dominated by uncertain ORCHIDEE parameters. The progressive increase in the weight of ORCHIDEE parameters uncertainties is due to the reduction in the role played by LAI for biomass increase along the growing season. Indeed, if early in the season the foliage is crucial to allow photosynthesis and carbon allocation, later in the cycle, other processes become important as well and passed a certain LAI for which all incoming light is captured, it might not even play a role anymore and then the STICS parameters only impact biomass accumulation through nitrogen stress index and root depth.

3.3 Uncertainty analysis: role of ORCHIDEE vs. STICS parameters in controlling biomass uncertainty at 7 sites

Table 3 summarizes the results of the overall parametric uncertainty analysis at the 7 sites, including Grafton. The total uncertainty U(O+S) ranges between 25.5% of biomass at Piracicaba, Brazil during 2004-05 and 44.26% of harvested biomass at Tirano, La Réunion in 1998-99 yielding an average uncertainty on biomass at harvest due to uncertain parameter values of the chained model ORCHIDEE-STICS of 34.0% of harvested biomass across the 7 sites, in the order of previous results on different variables in similar studies using process-based models such as (Dufrêne et al., 2005) who found an uncertainty of 30% on modeled NEE for a forest sites in France with the CASTANEA model.

As for the ORCHIDEE vs. STICS relative contributions to the uncertainty of simulated biomass at all sites, the results at each site are not identical but display a similar general pattern shown by figure 7. For all sites, the ORCHIDEE parameters contribution to total uncertainty increases during the cycle, or remains approximately constant for Ingham in 1992-93, and increases during the growing cycle to dominate entirely the total uncertainty at the end of the cycle compared to STICS parameters. The STICS contribution to overall uncertainty decreases during the growing season to reach a minimum by the end of the growing season. For sites Piracicaba during 2004-05, Tirano in 1998-99 and Colimaçons during 1994-95, during the beginning of the cycle the U(S) is even larger than U(O). The results for Ayr in 1991-92 display a less clear pattern. Indeed, at the end of the cycle, the contributions of ORCHIDEE and STICS to the total uncertainty are almost equal,due to an increase in STICS contribution during the second half of the cycle. This result confirms a hypothesis made in Valade et al. (2013) where the difficult calibration of LAI at this site was attributed to the simulation by STICS of an important stress. Indeed if a large stress is simulated by the phenological module, this can impede ORCHIDEE processes of biomass growth and therefore increases the weight of STICS parameters with respect to ORCHIDEE ones.

3.4 Uncertainty analysis: constraining uncertainty from sites optimization

Optimizing the 11 ORCHIDEE-STICS parameters selected from the screening analysis at 7 sites leads to a reduction of the width of the a priori uncertainty distribution of the parameters (Table 2). Carrying out the same uncertainty analysis with a narrower uncertainty range of parameters (thanks to their site calibration) leads to an important reduction of uncertainties of biomass both for the STICS and ORCHIDEE components of uncertainty. This can be seen by comparing Figure 6 (initial range of parameters) with figure 8 (narrower range after parameters calibration at the sites). For site Grafton during 1994-95 for example, U(O+S) gets reduced from 26% to 17% of the reference harvested biomass, U(O) from 24% to 15% and U(S) from 14% to 10%. Figure 9 and Table 3 (bottom section) show the uncertainty contributions and overall uncertainty estimates for the 7 sites after observation-based reduction of the a priori uncertainty on parameters. The overall parametric uncertainty of biomass defined as the 1-sigma standard deviation of the (O+S) distribution has thus been reduced to 21% in average, to 11.48% when attributed to STICS alone, and to 17.15% when attributed to ORCHIDEE alone, (Table 3).

The ORCHIDEE and STICS contributions to the total uncertainty keep the same general pattern as with the initial parameters uncertainty distribution, with a domination of ORCHIDEE parameters in the uncertainty towards the end of the growing season (fig.9). Compared with the first uncertainty budget with expert-based parameters uncertainties (fig.8), there is generally a slight decrease in the STICS contribution at the end of the season.

We have thus established full uncertainty budgets for the two components of the ORCHIDEE-STICS chain of models, which has revealed variations in the uncertainty in the biomass simulation from site to site. The next step is to discriminate between the different parameters the ones that contribute most to the overall uncertainty through a sensitivity analysis at regional scale.

3.5 Spatial sensitivity analysis: sensitivity of sugar cane yields to the model parameters for Brazil and Australia

The overall parametric uncertainties have been quantified at 7 sites and attributed to either STICS or ORCHIDEE. The sensitivity analysis (SA) in this section will go a step further and leads to discriminate the different parameters that contribute to the spatial distribution of uncertainty over the two regions considered. This sensitivity analysis is performed at regional scale because from the previous section, we have seen that the uncertainty in the biomass simulation varies from site to site.

Ensemble runs at regional scale were realized over Brazil and Australia each with different value combinations for the 11 parameters previously selected through the Morris screening analysis (Table 1). The Partial Rank Correlation Coefficients (PRCC) were then calculated for each pixel in each of the two regions (see section 2.5), and the SA results are discussed for two dates during the growing season, 200 and 350 days after planting (DAP). The SA results express the strength of the relationship between an uncertain parameter and the simulated biomass at harvest at each pixel. The statistical significance of the PRCC calculated for each grid cell is tested with the associated p-values, and non-significant PRCC are removed (p-value<0.05). The first date 100 DAP examined for site scale UA studies (section 2.3) is not shown here, because no statistical significance was found in the correlations between the parameters and the harvested biomass at 100 DAP. Then, the pixels statistically significant PRCC calculated for each parameter can be analyzed both in a geographical projection (latitude, longitude) (fig. 11 & 12, columns 1-2 and 4-5) and in a (Temperature, Precipitation) climatic space projection (fig 11 & 12, columns 3 and 6). The regional sensitivity analysis thus carried out for sugar cane growing areas in Brazil and Australia shows the magnitude, spatial distribution and climatic dependency of the sensitivity of harvested biomass to the 11 parameters previously selected through the Morris screening analysis (Table 2).

Across both regions in Brazil and Australia, we find that the sensitivity of biomass to the model parameters is not uniformly distributed. This means that the simulated yield depends on different parameters within different parts of the same region. This result shows that applying a model at one site to determine the most important parameters, and generalizing its conclusion across a region generates biased conclusions. Considering only the first most important parameter in each pixel (fig. 10), we can see that early in the cycle (200 DAP, Figure 10a) four parameters dominate the spatial distribution of the U(O+S) uncertainty of biomass at 200 DAP, both over Brazil and Australia. These parameters are three ORCHIDEE parameters involved in the photosynthesis process, the minimum and optimum temperature for photosynthesis T_{min}, T_{opt} , and the maximum rate of carboxylation V_{Cmax}^{opt} , and one parameter from STICS δ_{LAI}^{max}, defining the maximum rate of increase of LAI and only appearing in the Australian region. In Brazil, the parameter V_{Cmax}^{opt} is the first most important parameter for 93% of the area, whereas the optimum and minimum photosynthesis temperatures parameters only dominate in respectively 3 and 4% of the area. In Australia, the parameters' domination is more balanced with 37.5% for each of V_{Cmax}^{opt} and δ_{LAI}^{max} and 25% for T_{min}.

Later in the growing season (350DAP, fig.10b), consistently with the results of the site-scale uncertainty analysis, the influence of the STICS parameters decreases until no STICS parameters appear any longer as a dominant parameter in any of the regions. At this later stage in the season, two parameters stand out as explaining most of the uncertainty in most pixels of both regions, V_{Cmax}^{opt} and T_{min}. In Brazil, V_{Cmax}^{opt} is still the most sensitive parameter for most of the region, but T_{opt} disappeared and the area dominated by T_{min} expanded and now covers the cooler area of the southeast coastal zone, which is likely to result from the growing calendar of sugarcane in Brazil since the later part of the growing season takes place during winter in this region. In Australia, the area dominated by V_{Cmax}^{opt} expanded into most of the region and now covers 83% of the area. In the coolest pixels, the soil-related parameters appear with the two root profile parameters from STICS and from ORCHIDEE, κ_{root} and κ_{hum}.

Figures 11 and 12 focus on the values of the PRCC for each parameter as well as their spatial distribution. Their projection in a Temperature-Precipitation space for a given time (fig.11 for 200 DAP, fig.12 for 350 DAP) give more insights on the dependency of the sensitivity to the climatic conditions along the growing cycle. As an example, the sensitivity of the simulated biomass to T_{min} is highly sensitive to the average temperature of the location. At low-temperature sites, where temperature is a limiting factor for crop growth (below 17˚C), the PRCC is higher than 0.8, whereas at high-temperature sites (above 22˚C) the PRCC is below 0.3. Sites with temperatures above 25˚C do not even show significant correlations (grey symbols on the scatter plot).

For the parameter κ_{hum}, which describes the root profile of the cane (inverse of root depth), the dependency is most obvious on precipitation amount. For annual precipitations above 2500mm, no significant correlation is found.

Comparing the regional sensitivities at two times in the growing season shows again the decrease in the importance of STICS parameters whereas all of most important ORCHIDEE parameters have larger RPCC than earlier in the season.

4 Concluding remarks

In the perspective of applying spatially explicit mechanistic vegetation models such as ORCHIDEE-STICS to biofuel yield simulations we have sought the quantification and understanding of parametric uncertainty propagation in the model, both at site level and at sub-continental scale over two large regions, Australia and Brazil. For this, a rigorous analysis of the uncertainty budget of simulated sugar cane biomass has been established, using a step by step tracking of uncertainty in the model.

The main parameters from the two chain components of the model responsible for most of the uncertainty propagation have been identified through a Morris screening analysis. For the ORCHIDEE carbon, water and energy model, the most influential parameters are those involved in photosynthesis equations, T_{min}, T_{opt}, V_{Cmax}^{opt}, the radiation interception parameter k_{ext}, the root profile constant κ_{hum}, the parameters for respiration, slope of the Ball-Berry relation β, maintenance and growth respiration parameters f_{Gresp} and α_{Mresp}. For the STICS model, the most influential parameters are those responsible for simulation of phenology, nitrogen and water stress. The parameters describing the maximum rate of carboxylation, the maximum growth rate of the root front and the threshold for nitrogen stress have been found to have the greatest role. The parameters identified are closely related to the structure of the coupling since the key variables transmitted from STICS to ORCHIDEE each convey one key parameter.

We used two approaches for estimating the total uncertainty propagated from the parameters into the model by assigning uncertainties on parameters with two methods, one ‘pessimistic’, in which a-priori parameter uncertainty bounds are set based on expert judgment, and one optimistic where smaller uncertainty is derived by an optimization of the model parameters at several sites thus providing a smaller, arguably more realistic, a-priori uncertainty range.

We found that all these parameters together contribute to an overall uncertainty of 21% on sugar cane biomass simulations with an agro-LSM model and that this amount is variable among sites with different climatic, edaphic and management situations. We also analyzed this uncertainty separately for each component of the model and found that whatever estimate chosen for the parameters uncertainty, by the end of the growing season, the uncertainty propagated from the phenology module

STICS decreases and the overall uncertainty is almost totally explained by the ORCHIDEE uncertainty. The lower uncertainty from STICS parameters compared to ORCHIDEE ones is likely related with the lower number of processes solved by STICS in its configuration with ORCHIDEE, and to some extent to the lower number of parameters propagating their uncertainties. The decrease in the weight of the STICS' parameters to the overall uncertainty is linked to the canopy closure (LAI sufficient to capture all incoming light) and would therefore probably happen at a different timing in the growing season for different crops. For example, soybean experiences a later canopy closure and would probably show a later diminution of the STICS contribution to overall uncertainty, therefore remaining relatively high by the end of the cycle.

The overall origin of uncertainty has then been diagnosed in even more detail through a regional sensitivity analysis allowing the identification of the parameter for which harvested biomass is most sensitive for each pixel within regions of Australia and Brazil. We revealed a strong heterogeneity of the results based on climatic conditions and also variability in time that confirms the results of the uncertainty analysis, by showing a decrease in the importance of the STICS parameters along the growing season.

We believe that our results for the sugar cane crop simulated with the model ORCHIDEE-STICS are relevant to other agro-LSM with different crops. All these results prove the importance of establishing clear uncertainty budgets for highly parameterized models such as agro-LSM, especially when applying these models to answer questions related to political decisions such as biofuels burning topics.

As an example, combining our optimistic uncertainty estimation with the estimations from (Lapola et al., 2009) for irrigated sugar cane (obtained with the model LPJml, very similar to ORCHIDEE-STICS), we can evaluate the range assorted with their estimation of land requirements to fulfill the demand in ethanol in Brazil. Similarly to our study they use a multi-continental approach, focusing on Brazil and India. They simulate with a single parameterization the sugarcane productivity over both considered countries, spanning a wide range of climatic conditions. They found a mean yield of 68.8 t/ha over Brazil and 73.3 t/ha over India, and conclude that to fulfill government targets, the sugar cane areas would need to expand by 2.8 million hectares in Brazil and 1 million hectare in India. Because the yield estimates derived in (Lapola et al., 2009) are retrieved with an global agro-LSM parameterized for global applications and used in a range of climatic conditions (whole Brazil and India), we make the hypothesis that our uncertainty calculation is applicable to the LPJml results. We can then take into account the parametric uncertainty of the model and translate the potential mean production into a range of [54-83t/ha] for Brazil and [58-89t/ha] for India. The land requirements when including parameters uncertainty would then becomes [2.6– 3.9 million hectares],for Brazil and [0.9 – 1.4 million hectares] for India. To go further in the application of this result, and assuming that sugar cane expansion results in deforestation through direct or indirect land use change, we can translate the land expansion of sugar cane for biofuels into carbon emissions from deforestation. Several estimates of carbon emissions associated with conversion of tropical forest to croplands have been published and their results span a large range revealing the large uncertainties in this area (BSI, 2008; Cederberg et al., 2011; Searchinger et al., 2008) Discussing the uncertainty on this estimate is beyond the scope of this paper so we will only consider the value from (Searchinger et al., 2008), of 604tCO2eq/ha. Using this conversion factor, the expansion of sugar cane calculated by (Lapola et al., 2009) would result in CO2eq emissions of 1,68GtCO2eq whereas including the parametric uncertainty of the model we obtain a range of 1,6 to 2,4 GtCO2eq provoked by Brazilian government's ethanol targets with our calculation of uncertainty.

With the choice of the study from Lapola et al. (2009) to apply our uncertainty estimates on, we favored the closeness of the models over the full consistency of the methodologies. If the primary goal had been to calculate estimates of uncertainty of land requirements in the specific region of Brazil, we would have constrained our parameters ranges for conditions of this region, which would have resulted in lower uncertainty ranges for area requirements. However, we want to stress that agro-LSMs like ORCHIDEE-STICS or LPJml are designed for global studies and their parameters are therefore supposed to cover the full range of climatic conditions even when they are used for regional applications. This quick application of our uncertainty calculation proves how important it is to consider the uncertainty when addressing issues aimed at decision-makers.

5 Acknowledgements

This study was performed using HPC resources from GENCI- CCRT (Grant 2012-016328). We acknowledge Arnaud Caubel for his help with the experimental setup and the development of scripts for multiple runs and Patrick Brockmann for his help with data visualization.

6 References

Ball, J.T., Woodrow, I.E., Berry, J.A., (1987) A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, Progress in photosynthesis research. Springer, pp. 221-224.
Black, E., Vidale, P.L., Verhoef, A., Cuadra, S.V., Osborne, T., Van den Hoof, C. (2012) Cultivating C4 crops in a changing climate: sugarcane in Ghana. Environmental Research Letters 7, 044027.
Bouman, B. (1994) A framework to deal with uncertainty in soil and management parameters in crop yield simulation: a case study for rice. Agricultural Systems 46, 1-17.
Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P. (2003) An overview of the crop model STICS. European Journal of Agronomy 18, 309-332.
Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.-H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.-M., Meynard, J.-M., Delécolle, R. (1998) STICS: a generic model for the simulation of crops and their water and nitrogen

balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18, 311-346.
BSI (2008) PAS 2050: 2008 Specification for the assessment of the life cycle greenhouse gas emissions of goods and services. British Standards Institution.
Campolongo, F., Cariboni, J., Saltelli, A. (2007) An effective screening design for sensitivity analysis of large models. Environmental Modelling & Software 22, 1509-1518.
Cariboni, J., Gatelli, D., Liska, R., Saltelli, A. (2007) The role of sensitivity analysis in ecological modelling. Ecological Modelling 203, 167-182.
Cederberg, C., Persson, U.M., Neovius, K., Molander, S., Clift, R. (2011) Including carbon emissions from deforestation in the carbon footprint of Brazilian beef. Environmental Science & Technology 45, 1773-1779.
Cheeroo-Nayamuth, F.C., Robertson, M.J., Wegener, M.K., Nayamuth, A.R.H. (2000) Using a simulation model to assess potential and attainable sugar cane yield in Mauritius. Field Crops Research 66, 225-243.
Cuadra, S.V., Costa, M.H., Kucharik, C.J., Da Rocha, H.R., Tatsch, J.D., Inman-Bamber, G., Da Rocha, R.P., Leite, C.C., Cabral, O.M.R. (2012) A biophysical model of Sugarcane growth. GCB Bioenergy 4, 36-48.
de Vries, S.C., van de Ven, G.W.J., van Ittersum, M.K., Giller, K.E. (2010) Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. Biomass and Bioenergy 34, 588-601.
Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137, 553-597.
Demirbas, A. (2008) Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. Energy Conversion and Management 49, 2106-2116.
Doornbosch, R., Steenblik, R. Biofuels: Is the cure worse than the disease? Revista Virtual REDESMA 2, 63-100.
Dufrêne, E., Davi, H., François, C., Maire, G.l., Dantec, V.L., Granier, A. (2005) Modelling carbon and water cycles in a beech forest: Part I: Model description and uncertainty analysis on modelled NEE. Ecological Modelling 185, 407-436.
Fargione, J., Hill, J., Tilman, D., Polasky, S., Hawthorne, P. (2008) Land Clearing and the Biofuel Carbon Debt. Science 319, 1235-1238.
Georgescu, M., Lobell, D.B., Field, C.B., Mahalov, A. (2013) Simulated hydroclimatic impacts of projected Brazilian sugarcane expansion. Geophysical Research Letters 40, 972-977.
Gervois, S., Ciais, P., de Noblet-DucoudrÈ, N., Brisson, N., Vuichard, N., Viovy, N. (2008) Carbon and water balance of European croplands throughout the 20th century. Global Biogeochem. Cycles 22, GB2022.
Gervois, S., de Noblet-Ducoudre, N., Viovy, N., Ciais, P., Brisson, N., Seguin, B., Perrier, A. (2004) Including Croplands in a Global Biosphere Model: Methodology and Evaluation at Specific Sites. Earth Interactions 8, 1-25.

Groenendijk, M., Dolman, A.J., van der Molen, M.K., Leuning, R., Arneth, A., Delpierre, N., Gash, J.H.C., Lindroth, A., Richardson, A.D., Verbeeck, H., Wohlfahrt, G. (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.
Hill, J., Nelson, E., Tilman, D., Polasky, S., Tiffany, D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy of Sciences 103, 11206-11210.
Howarth, R.W., Bringezu, S., Martinelli, L.A., Santoro, R., Messem, D., Sala, O.E. (2008) Introduction: Biofuels and the Environment in the 21st Century. Biofuels: environmental consequences and interactions with changing land use. Proceedings of the Scientific Committee on Problems of the Environment (SCOPE) International Biofuels Project Rapid Assessment, 22-25.
IEA, (2004) Biofuels for transport An international perspective. International Energy Agency, p. 210.
IEA, (2011) Technology Roadmap, Biofuels for Transport.
Keating, B.A., Robertson, M.J., Muchow, R.C., Huth, N.I. (1999) Modelling sugarcane production systems I. Development and performance of the sugarcane module. Field Crops Research 61, 253-271.
Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C. (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Global Biogeochem. Cycles 19, GB1015.
Kuppel, S., Chevallier, F., Peylin, P. (2012) Quantifying the model structural error in carbon cycle data assimilation systems. Geoscientific Model Development Discussions 5, 2259-2288.
Lapola, D.M., Priess, J.A., Bondeau, A. (2009) Modeling the land requirements and potential productivity of sugarcane and jatropha in Brazil and India using the LPJmL dynamic global vegetation model. Biomass and Bioenergy 33, 1087-1095.
Loarie, S.R., Lobell, D.B., Asner, G.P., Mu, Q., Field, C.B. (2011) Direct impacts on local climate of sugar-cane expansion in Brazil. Nature Clim. Change 1, 105-109.
Macedo, I.C., Seabra, J.E.A., Silva, J.E.A.R. (2008) Greenhouse gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020. Biomass and Bioenergy 32, 582-595.
Marin, F.R., Jones, J.W., Royce, F., Suguitani, C., Donzeli, J.L., Wander Filho, J.P., Nassif, D.S. (2011) Parameterization and evaluation of predictions of DSSAT/CANEGRO for Brazilian sugarcane. Agronomy journal 103, 304-315.
Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E., Marino, S., Hogue, I., Ray, C., Kirschner, D. (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of theoretical biology 254, 178.
McKay, M.D., Beckman, R.J., Conover, W.J. (1979) Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, 239-245.
Medlyn, B.E., Robinson, A.P., Clement, R., McMurtrie, R.E. (2005) On the validation of models of forest CO2 exchange using eddy covariance data: some perils and pitfalls. Tree Physiology 25, 839-857.
Monod, H., Naud, C., Makowski, D., (2006) Uncertainty and sensitivity analysis for crop models, in: D. Wallach, D.M., J.W. Jones (Ed.), Working with Dynamic Crop

Models. Evaluation, Analysis, Parameterization and Applications. Elsevier, Amsterdam, pp. 55-100.

Monsi, M., Saeki, T. (1953) Über den Lichtfaktor in den pflanzengesellschaften und seine bedeutung für die stoffproduktion. Japanese Journal of Botany 14, 22-52.

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

Muchow, R.C., Spillman, M.F., Wood, A.W., Thomas, M.R. (1994) Radiation interception and biomass accumulation in a sugarcane crop grown under irrigated tropical conditions. Australian Journal of Agricultural Research 45, 37-49.

Naylor, R.L., Liska, A.J., Burke, M.B., Falcon, W.P., Gaskell, J.C., Rozelle, S.D., Cassman, K.G. (2007) The Ripple Effect: Biofuels, Food Security, and the Environment. Environment: Science and Policy for Sustainable Development 49, 30-43.

OECD, (2012) OECD-FAO Agricultural Outlook 2012-2021, June 2012 ed. OECD Publishing and FAO.

Poulter, B., Hattermann, F., Hawkins, E.D., Zaehle, S., Sitch, S., Restrepo-Coupe, N., Heyder, U., Cramer, W. (2010) Robust dynamics of Amazon dieback to climate change with perturbed ecosystem model parameters. Global Change Biology 16, 2476-2495.

Pujol, G. (2009) Simplex-based screening designs for estimating metamodels. Reliability Engineering & System Safety 94, 1156-1160.

Pujol, G., Iooss, B., Janon, A., (2013) sensitivity: sensitivity analysis, 1.7 ed, p. R package.

Robertson, M.J., Wood, A.W., Muchow, R.C. (1996) Growth of sugarcane under high input conditions in tropical Australia. I. Radiation use, biomass accumulation and partitioning. Field Crops Research 48, 11-25.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M. (2013) The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. Agricultural and Forest Meteorology 170, 166-182.

Saltelli, A., Marivoet, J. (1990) Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. Reliability Engineering & System Safety 28, 229-253.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M. (2004) Sensitivity Analysis in practice. A Guide to Assessing Scientific Models. Halsted Press New York, NY, USA.

Schubert, C. (2006) Can biofuels finally take center stage? Nat Biotech 24, 777-784.

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T.-H. (2008) Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. Science 319, 1238-1240.

Sorda, G., Banse, M., Kemfert, C. (2010) An overview of biofuel policies across the world. Energy Policy 38, 6977-6988.

Surendran Nair, S., Kang, S., Zhang, X., Miguez, F.E., Izaurralde, R.C., Post, W.M., Dietze, M.C., Lynd, L.R., Wullschleger, S.D. (2012) Bioenergy crop models:

descriptions, data requirements, and future challenges. GCB Bioenergy 4, 620-633.
Tarantola, A. (1987) Inverse problem theory : methods for data fitting and model parameter estimation. Elsevier ; Distributors for the United States and Canada, Elsevier Science Pub. Co., Amsterdam; New York; New York, NY, U.S.A.
Valade, A., Vuichard, N., Ciais, P., Ruget, F., Viovy, N., Gabrielle, B., Huth, N., Martiné, J.F. (2013) ORCHIDEE-STICS, a process-based model of sugarcane biomass production: calibration of model parameters governing phenology. GCB Bioenergy.
Verbeeck, H., Samson, R., Verdonck, F., Lemeur, R. (2006) Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: a Monte Carlo analysis. Tree Physiology 26, 807-817.
von Blottnitz, H., Curran, M.A. (2007) A review of assessments conducted on bio-ethanol as a transportation fuel from a net energy, greenhouse gas, and environmental life cycle perspective. Journal of Cleaner Production 15, 607-619.
Wang, X., He, X., Williams, J., Izaurralde, R., Atwood, J. (2005) Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC. TRANSACTIONS-AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS 48, 1041.
Wyss, G.D., Jorgensen, K.H. (1998) A user's guide to LHS: Sandia's Latin hypercube sampling software. SAND98-0210, Sandia National Laboratories, Albuquerque, NM.
Zaehle, S., Sitch, S., Smith, B., Hatterman, F. (2005) Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles 19, n/a-n/a.

Figure 1: flowchart of the analysis carried out in this study. The first step is the separate screening for 7 sites of the *STICS* and *ORCHIDEE* parameters. The selection of parameters obtained from the screening are then used for two uncertainty analysis, one with the same parameters ranges of variation as for the screening, the other with parameters ranges of variation constrained by the optimization of the model at 7 sites. Each uncertainty analysis is decomposed in three parts, one including only *ORCHIDEE* parameters, one including only *STICS* parameters and one including parameters from both *ORCHIDEE* and *STICS*. Finally a sensitivity analysis is carried out for two small regions in Australia in Brazil for all parameters together.

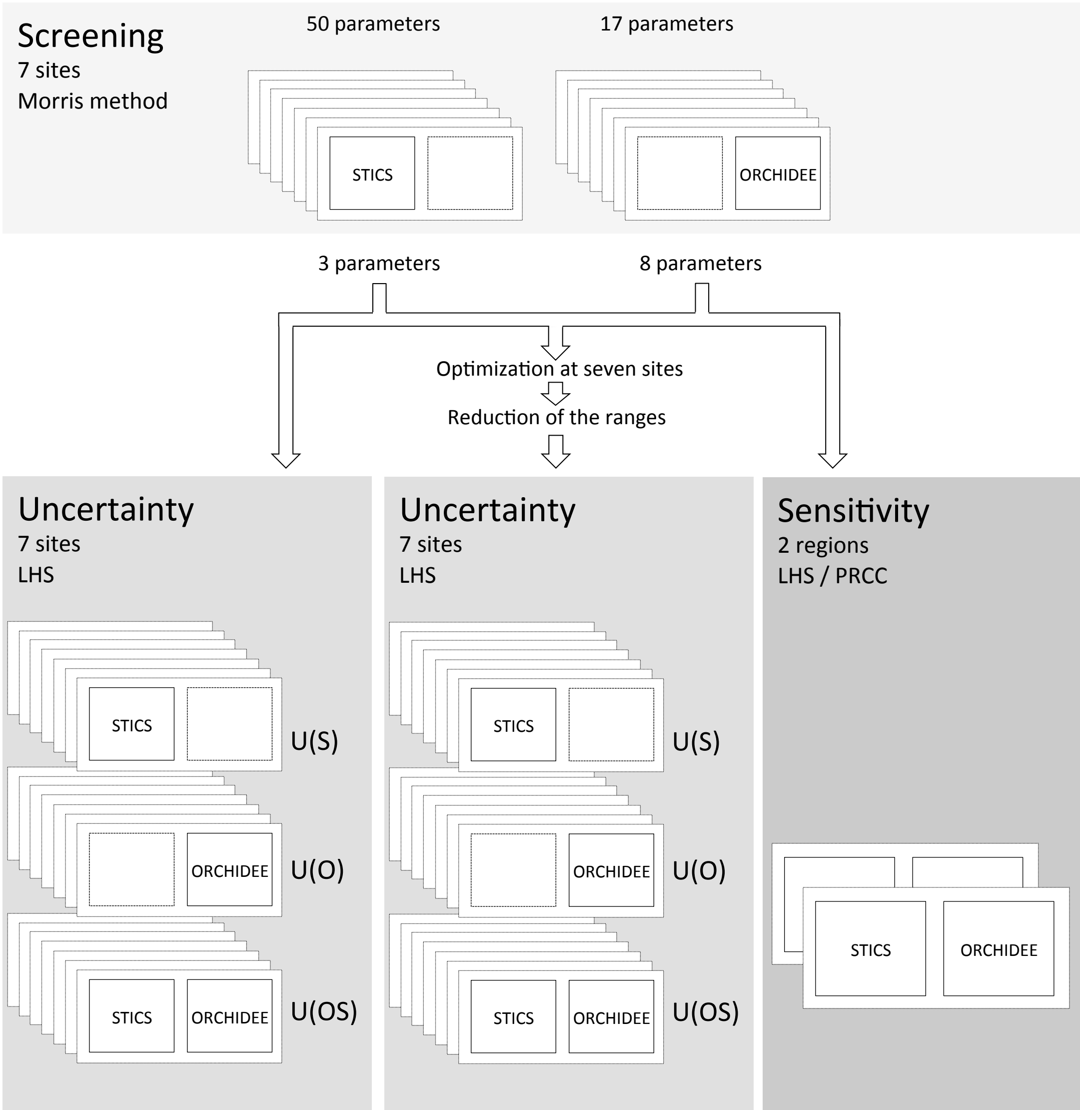

Screening
7 sites
Morris method
50 parameters
17 parameters
STICS
ORCHIDEE
3 parameters
8 parameters
Optimization at seven sites
Reduction of the ranges
Uncertainty
7 sites
LHS
STICS
U(S)
ORCHIDEE
U(O)
STICS
ORCHIDEE
U(OS)
Uncertainty
7 sites
LHS
STICS
U(S)
ORCHIDEE
U(O)
STICS
ORCHIDEE
U(OS)
Sensitivity
2 regions
LHS / PRCC
STICS
ORCHIDEE

Figure 2: Spatial distribution of the sites (dots) and regions (dashed rectangles) used in this study overlaid on a map of the distribution of sugar cane growing areas indicated in green.

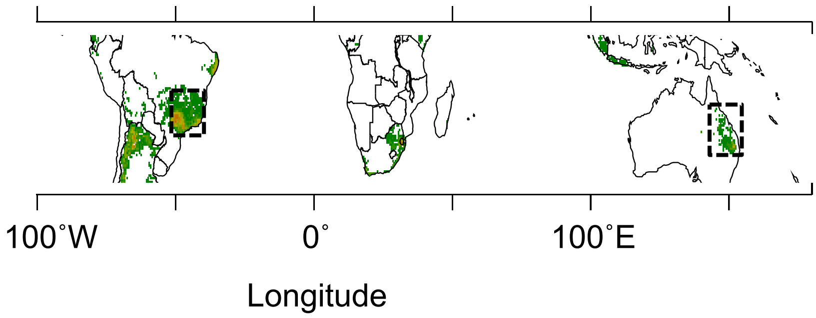


Figure 3: Structure of the *ORCHIDEE-STICS* chain model. *STICS* calculates the crop phenology, water and nitrogen requirements and passes LAI, root profile, irrigation and Nitrogen nutrition index to *ORCHIDEE*. *ORCHIDEE* consists in the coupling of two module. *SECHIBA* simulates the photosynthesis process, water and energy budgets, *STOMATE* is a carbon module and calculates carbon fluxes and to the atmosphere (respiration) and carbon accumulation in the carbon pools (biomass compartments, litter, soil).

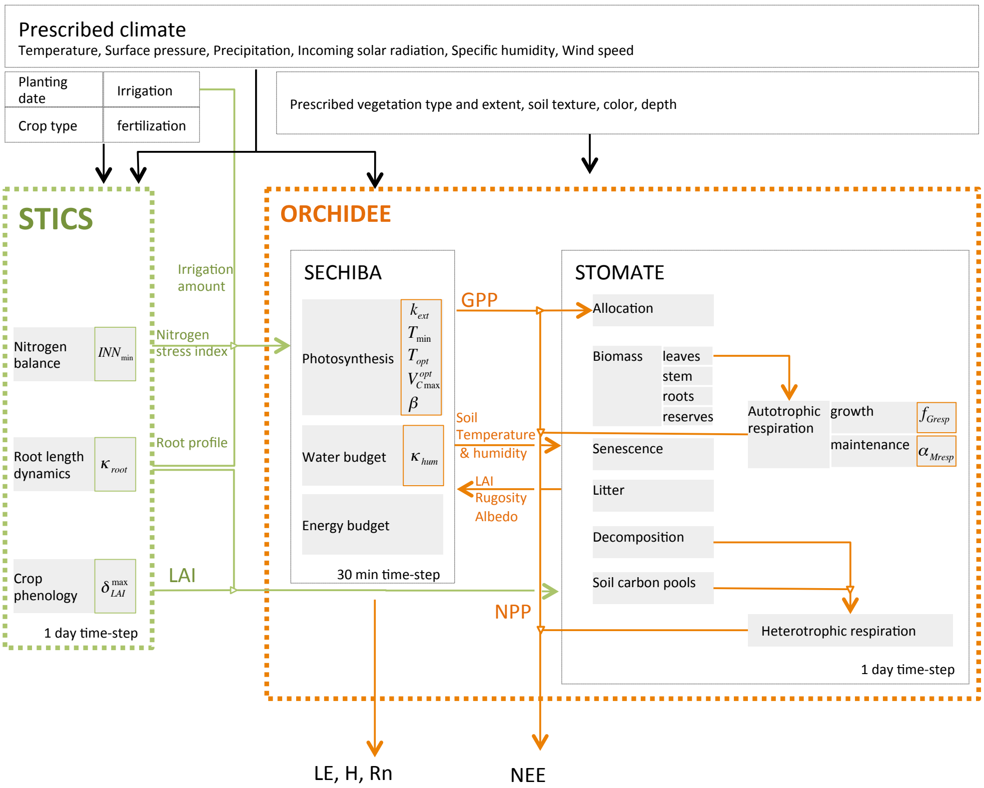


Figure 4: Main parameters for simulation of sugar cane yield with *ORCHIDEE-STICS* with the equations in which they are involved.

STICS			
δ_{LAI}^{max} : daily maximum increment of LAI ($m^2.plt^{-1}.deg\text{-}day^{-1}$) INN_{min} : threshold for nitrogen nutrition index (unitless)	$LAI = \int_0^T \Delta_{LAI}^{dev} \cdot \Delta_{LAI}^{T} \cdot \Delta_{LAI}^{dens} \cdot \Delta_{LAI}^{stress} .splai.dt$ $\Delta_{LAI}^{dens} = e^{\alpha_{dens} \log\left(\frac{d}{\beta_{dens}}\right)} \cdot d$ $\quad \Delta_{LAI}^{T} = \begin{cases} T_{crop} - T_{crop}^{min} \text{ if } T_{crop} \le T_{crop}^{max} \\ T_{crop}^{max} - T_{crop}^{min} \text{ if } T_{crop} \ge T_{crop}^{max} \end{cases}$ $\Delta_{LAI}^{dev} = \frac{\delta_{LAI}^{max}}{1+e^{f(k_{LAI})}}$ $\quad \Delta_{LAI}^{stress} = \min \begin{cases} INN = \max\left(\frac{C_N^{plant}}{C_N^{crit}}, INN_{min}\right) \\ W_s \end{cases}$	d : Sowing density ($plt.m^{-2}$) $\alpha_{dens}, \beta_{dens}$: Parameters describing competition from planting density(unitless, $plt.m^{-2}$) $T_{crop}, T_{crop}^{min}, T_{crop}^{max}$: Actual, minimum and maximum temperature for crop growth (°C) k_{LAI} : LAI development stage (unitless) LAI : Leaf area index ($m^2.m^{-2}$) INN, W_S : Nitrogen nutrition index, water stress C_N^{crit}, C_N^{plt} : Actual and critical nitrogen concentration in the plant (%) $splai$: Stress index for competition between foliage and sugar accumulation (unitless)	(Brisson et al., 2009)
κ_{root}: root growth rate	$z_{root}^{eff} = \int_0^T \kappa_{root} .f_{\Delta T} .pfz.f_{anox} .dt$	z_{root}^{eff} : Depth of the root front efficient for absorption (cm) $f_{\Delta T}$: Minimum temperature for emergence of the crop (°C) pfz : Water status of the soil layer (unitless) f_{anox} : Anoxia effect	(Brisson et al., 2009)
ORCHIDEE			
k_{ext} : extinction coefficient (unitless)	$light = e^{-k_{ext}.LAI}$	$light$: Light fraction that goes through the vegetation	(Monsi and Saeki, 1953)
T_{min} / T_{opt} : Minimum / Optimal photosynthesis temperatures (°C)	$\varepsilon_{temp} = f\left(T_{air}, T_{min}, T_{opt}\right)$	ε_{temp} : Limitation of photosynthesis capacity by temperature	(Krinner et al., 2005)
$V_{C\max}^{opt}$: Rate of carboxylation in optimal conditions ($\mu mol \cdot m^{-2} \cdot s^{-1}$)	$V_{C\max} = V_{C\max}^{opt} \cdot \varepsilon_{temp} \cdot \varepsilon_{water} \cdot \varepsilon_{leaf}$	$V_{C\max}$: Effective rate of carboxylation ($\mu mol.m^{-2}.s^{-1}$) ε_{stress} : Water and nitrogen limitation (unitless) ε_{temp} : Temperature limitation (unitless) ε_{leaf} : Limitation from leaf age (unitless)	(Ishida et al., 1999)
β : Ball-Berry slope (unitless)	$g_s = \beta \frac{h_r}{C_a} A + g_s^{offset}$	g_s : Stomatal conductance ($\mu mol.m^{-2}.s^{-1}$) A : Assimilation ($\mu mol.m^{-2}.s^{-1}$) h_r : Air relative humidity ($kg.kg^{-1}$) g_s^{offset} : Offset constant (unitless) C_a : Air CO_2 concentration (ppm)	(Ball et al., 1987)
κ_{hum}: Root profile description (m^{-1})	$T_{bg} = \frac{1}{K_{int}} + \sum_{k=1}^{n_{lay}} T_k \cdot \left(e^{-\kappa_{hum} \cdot z_{k-1}} - e^{-\kappa_{hum} \cdot z_k}\right)$ $f_w = \alpha_{sat} \cdot e^{-\kappa_{hum} \cdot h_{top}} + \left(1-\alpha_{sat}\right) \cdot e^{-\kappa_{hum} \cdot h_{bottom}}$	T_{bg} : Temperature for below-ground biomass (K) K_{int} : Integration constant (unitless) n_{lay}, z_k : Number of soil layers, depth of the k^{th} soil layer (m) T_k : Soil temperature for k^{th} layer (K) f_w : Water fraction available for the plant (unitless) α_{sat} : Normalizing coefficient related to wetness of the top soil layer (unitless) h_{top}, h_{bottom} : Depth of dry soil in the top and bottom layers (m)	(Krinner et al., 2005)
f_{Gresp} : fraction of GPP lost as growth respiration (dimensionless)	$Gresp = \frac{1}{\Delta t}\left(f_{Gresp} \cdot B_{alloc}\right)$	$Gresp$: Growth respiration ($gC \cdot m^2 \cdot dt^{-1}$) B_{alloc} : Allocatable biomass ($gC \cdot m^2$)	(Ruimy et al., 1996)
α_{Mresp} : slope of the dependance on temperature of maintenance respiration coefficient (K^{-1})	$c_i = \max \begin{cases} 0 \\ c_i^0 \left(1+\alpha_{Mresp} T_i\right) \end{cases}$ $Mresp = \frac{1}{\Delta t}\left(\sum_{i \ne leaf} c_i \cdot B_i + c_{leaf} \cdot B_{leaf} \cdot \frac{0.3 \cdot LAI + 1.4\left(1-e^{-k \cdot LAI}\right)}{LAI}\right)$	c_i : Fraction of biomass lost as maintenance respiration for part *i* (unitless) c_i^0 : Prescribed maintenance respiration coefficients at 0 Degree Celsius (unitless) T_i : Air (soil for belowground compartments) temperature (°C) M_{resp} : Maintenance respiration ($gC.m^{-2}.dt^{-1}$) B_i : Biomass in compartment i ($gC.m^{-2}$)	(Krinner et al., 2005; Ruimy et al., 1996)

Figure 5: Parameters rankings derived from the Morris screening analysis for *STICS* parameters (a) and *ORCHIDEE* parameters (b) for 7 sites (color lines). Each axis of the radar plot corresponds to the rank of a parameter, the lower the rank, the more important the parameter.

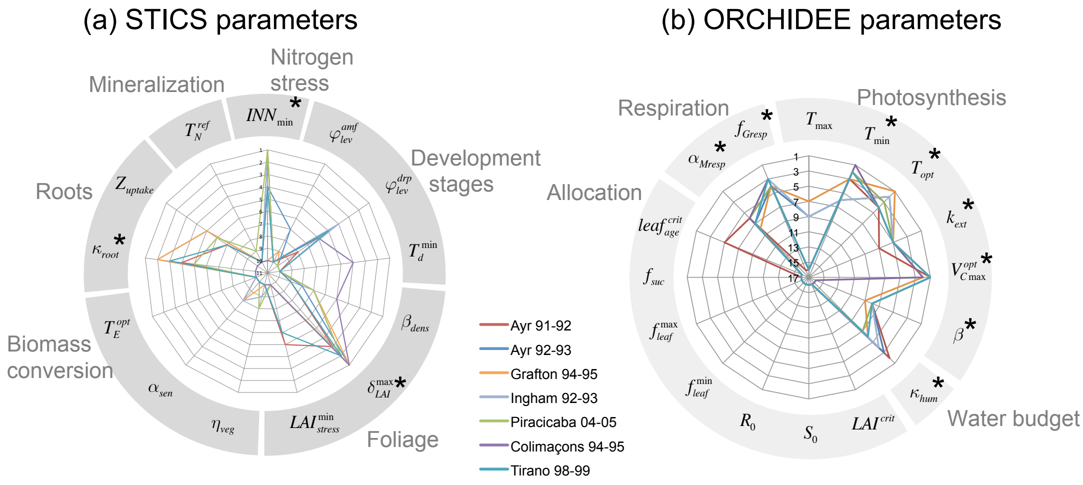


Figure 6 : Uncertainty analysis for the site Grafton 94-95. (a-c) probability distributions of harvested biomass simulated after parameters uncertainty (from *STICS*:green, from *ORCHIDEE*: yellow, from *ORCHIDEE+STICS*: brown) has been propagated into the model. (d) reference simulation of harvested biomass (red) and uncertainty from *ORCHIDEE*, *STICS*, *ORCHIDEE+STICS*. (e) Contribution (%)of *ORCHIDEE* (yellow) and *STICS* (green) to the total uncertainty (brown) over the length of the growing season.

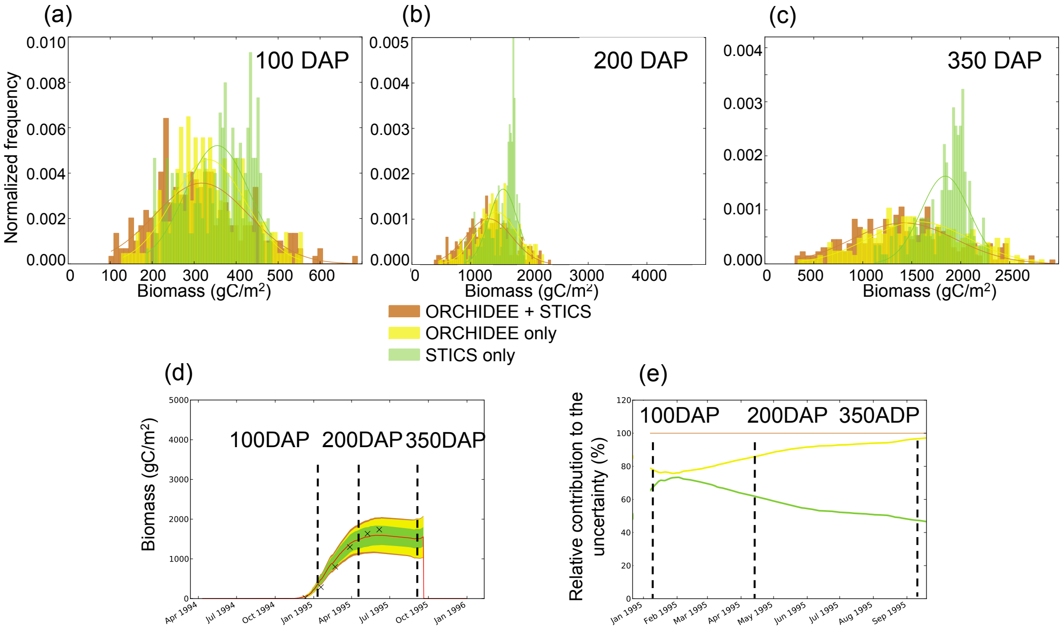


Figure 7 : Contribution (%) of *ORCHIDEE* (yellow) and *STICS* (green) to the total uncertainty (brown) over the length of the growing season for 7 sites.

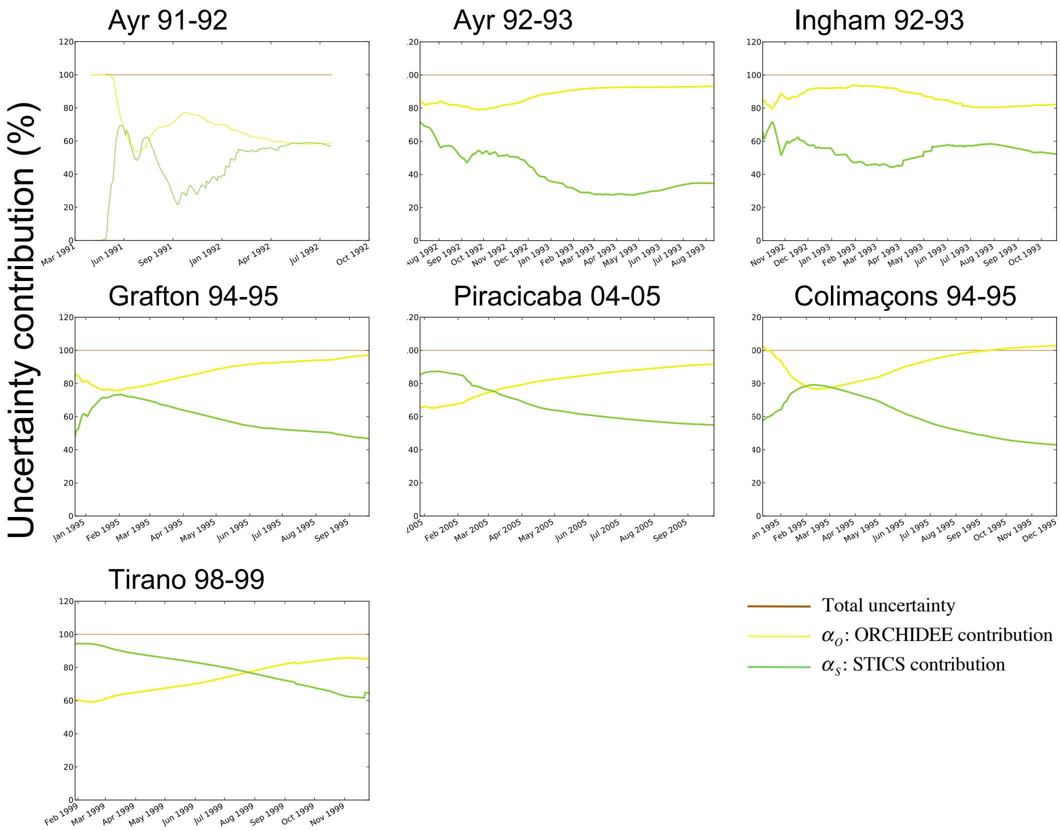


Figure 8: Uncertainty analysis for the site Grafton 94-95 after parameters uncertainty ranges have been constrained through optimization at 7 sites. (a-c) probability distributions of harvested biomass simulated after parameters un- certainty (from *STICS*: green, from *ORCHIDEE*: yellow, from *ORCHIDEE+STICS*: brown) has been propagated into the model. (d) reference simulation of harvested biomass (red) and uncertainty from *ORCHIDEE*, *STICS*, *OR- CHIDEE+STICS*. (e) Contribution (%) of *ORCHIDEE* (yellow) and *STICS* (green) to the total uncertainty (brown) over the length of the growing season.

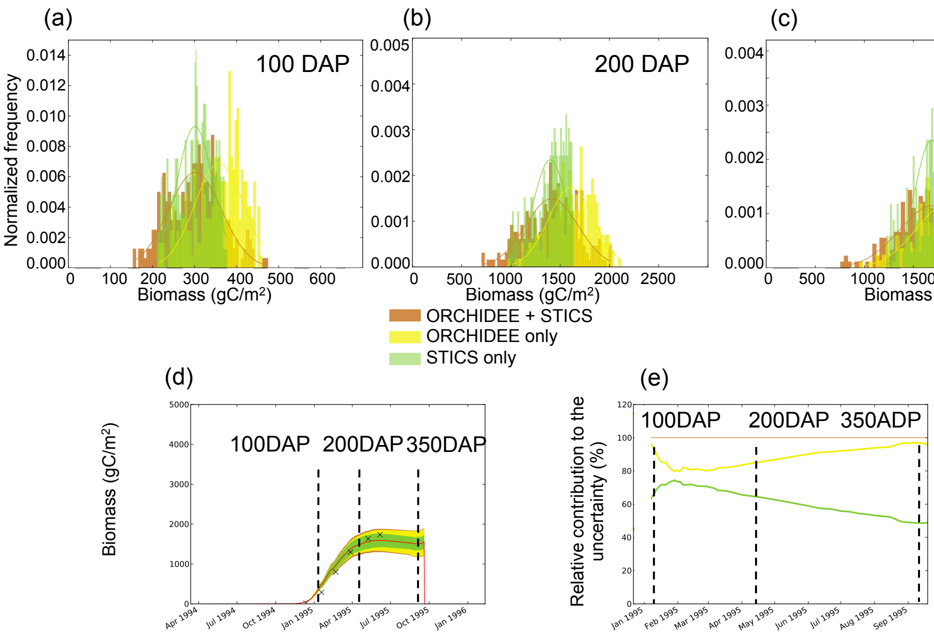


Figure 9: Contribution (%) of *ORCHIDEE* (yellow) and *STICS* (green) to the total uncertainty (brown) over the length of the growing season for 7 sites after parameters uncertainty ranges have been constrained through optimization at 7 sites.

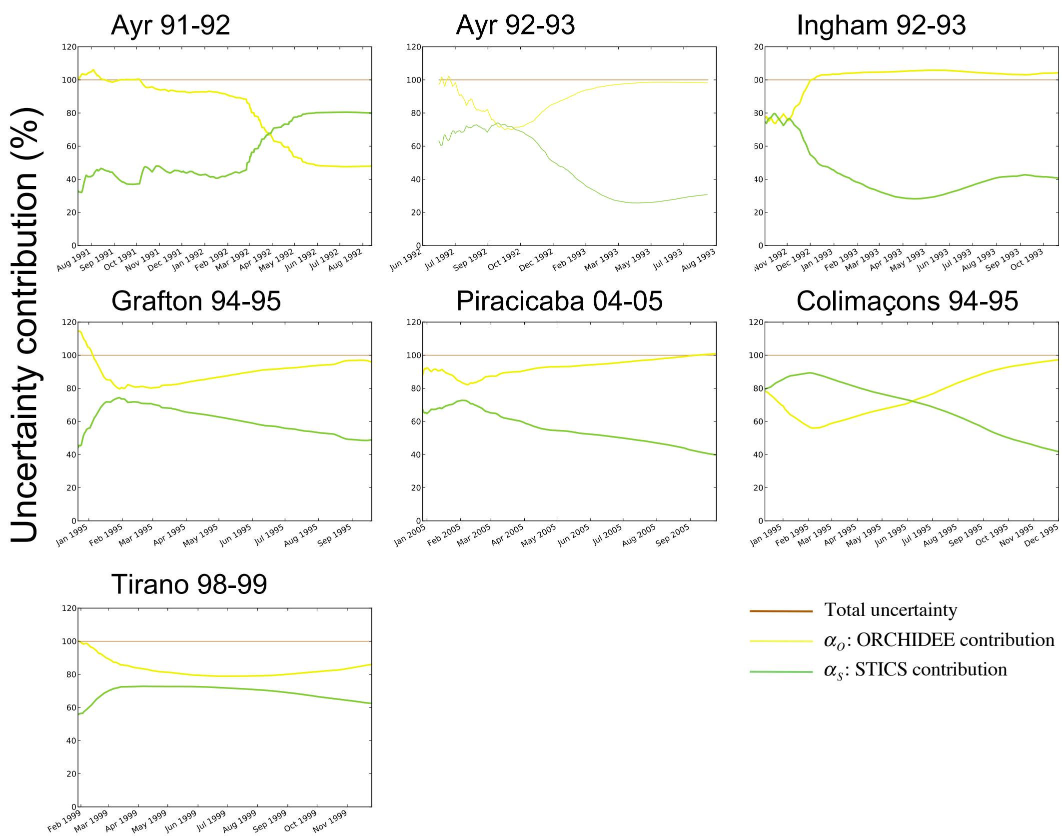


Figure 10: Spatial distribution of the most influential parameters for the simulation of harvestable biomass for two milestones during the growing season, 200 days after planting (DAP) and 350DAP

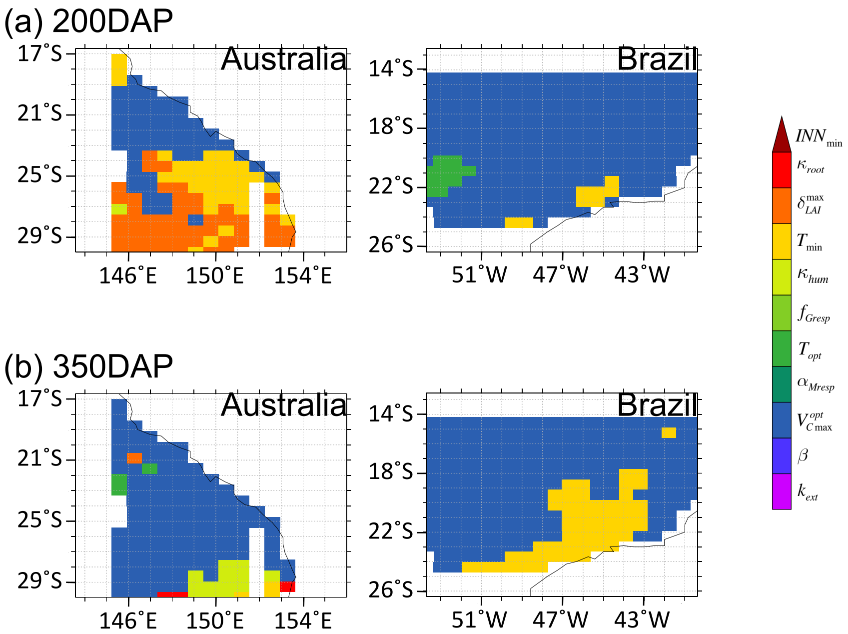
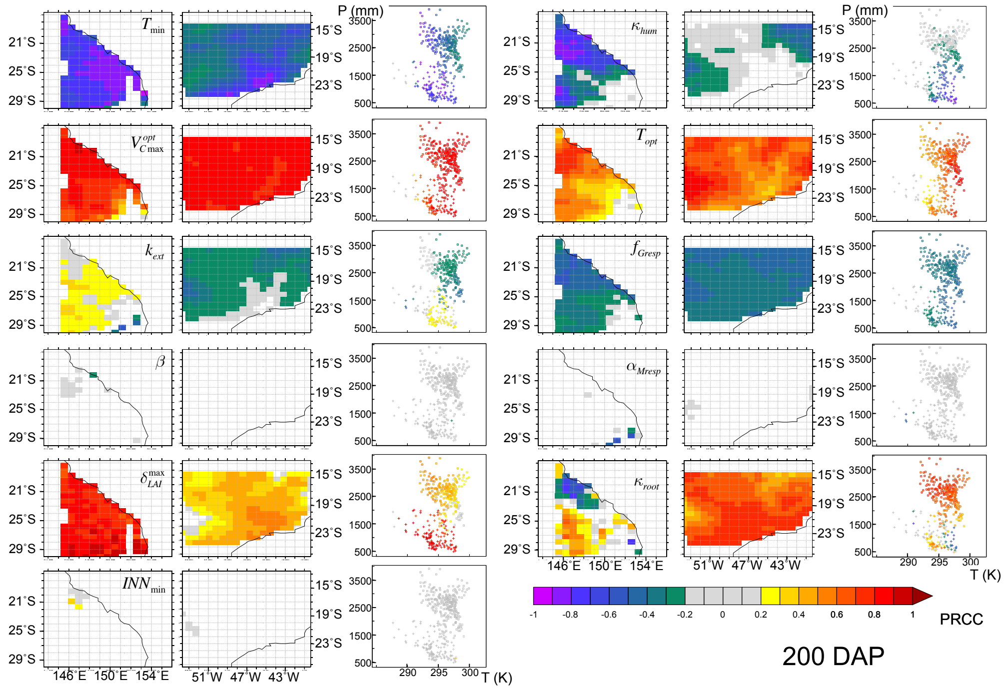


Figure 11: Sensitivity of *ORCHIDEE-STICS* to its main parameters at 200 days after planting, as measured with Partial Ranked Correlation Coefficients (PRCC). The color indicates the strength of the relation between the parameter and the harvestable biomass, which is represented spatially (columns 1,2,4,5) and in a (Temperature, Precipitation) referential (columns 3,6).



P (mm)
T_{min}
κ_{hum}
$V_{C\max}^{opt}$
T_{opt}
k_{ext}
f_{Gresp}
β
α_{Mresp}
δ_{LAI}^{max}
κ_{root}
INN_{min}
21°S
25°S
29°S
15°S
19°S
23°S
3500
2500
1500
500
146°E 150°E 154°E
51°W 47°W 43°W
290 295 300
T (K)
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
PRCC
200 DAP

Figure 12: Sensitivity of *ORCHIDEE-STICS* to its main parameters at 350 days after planting, as measured with Partial Ranked Correlation Coefficients (PRCC). The color indicates the strength of the relation between the parameter and the harvestable biomass, which is represented spatially (columns 1,2,4,5) and in a (Temperature, Precipitation) referential (columns 3,6).

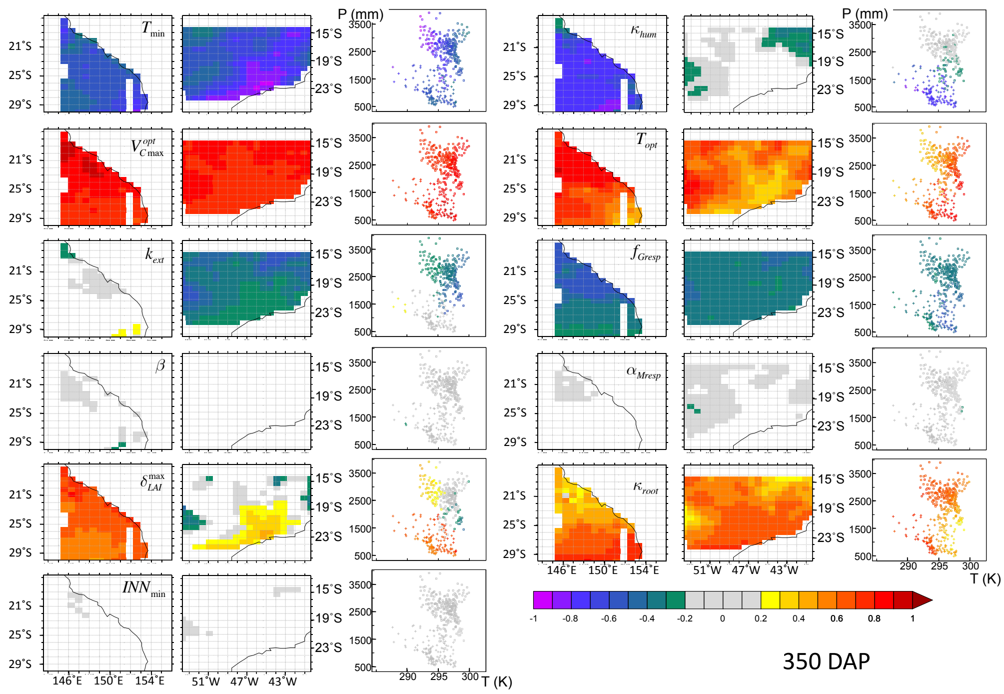


Table 1: Description of climate and management for the sites used in this study in Australia (Ayr, Ingham, Grafton), Brazil (Piracicaba) and La Runion (Colimaons, Tirano).

	Planting and harvest dates		Mean annual precipitation	Average temperature	irrigation	Fertilization
Ayr	4/19/1991	8/13/1992	964	23.4	irrigated	no
Ayr	4/22/1992	8/13/1993	560	23.6	irrigated	yes
Grafton	9/28/1994	9/19/1995	768	19.6	irrigated	yes
Ingham	7/23/1992	10/21/1993	1294	24.2	irrigated	yes
Piracicaba	10/29/2004	9/26/2005	1230	21.6	irrigated	
Colimacons	8/3/1994	12/1/1995	989.5	19	rainfed	yes
Tirano	11/26/1998	11/26/1999	813	22.34	irrigated	yes

Table 2: List of parameters from STICS and ORCHIDEE included in each step of the analysis with their ranges of variation.

			expert judgment based ranges		Uncertainty analysis distribution	Observations constrained ranges	
STICS							
Water budget	absolute value for stomatic closure potential	psisto	5	15			
	Absolute value for start of reduction in cell expansion	psiturg	1	5			
Initial conditions	Table of initial humidity levels in 5 soil horizons for fine soil, % weighted	Hinitf1	11	22			
		Hinitf2	11	22			
		Hinitf3	10	21			
	Table of initial quantities of nitrogen in the 5 siol horizons for fine soil	Ninitf1	0	30			
		Ninitf2	0	30			
		Ninitf3	0	30			
Biomass conversion	Relative age of fruit when rate of growth is maximum	afpf	0.15	0.5			
	Maximum number of set fruits per inflorescence and by degree.day	afruitpot	0.0015	0.2			
	Maximum daily allocation of assimilates towards fruits	allocamx	0.63	0.86			
	Rate of maximum growth as a proportion of maximum fruit weight	bfpf	1	10			
	Radiative effect on conversion efficiency	coefb	0.0015	0.0815			
	Duration of growth of a fruit from setting to physiological maturity	dureefruit	2850	3000			
	Maximum growth efficiency during juvenile phase	efcroijuv	1.7	2.3			
	Maximum growth efficiency during grain filling phase	efcroirepro	2	6			
	Maximum growth efficiency during vegetative phase	efcroiveg	3.2	6			
	Number of age groups of fruits for fruit growth	nboite	12	25			
	Maximum weight of a grain (% water)	pgrainmaxi	1200	2000			
	Fraction of senescent biomass	ratiosen	0	1			
	Quantity of biomass exploited during the cycle	remobil	0.728	0.92			
	Development range between DRP and NOU stages	sdrpnou	552.5	747.5			
	Threshold to calculate trophic stress on LAI	splaimin	0	0.3			
	Time between emergence and senescence	stlevsenms	400	800			

	(degree.day)						
	Cumulated development units allowing germination	stpltger	50	200			
	Optimum temperature for growth in biomass	teopt	15	34.4			
	Optimum temperature for growth in biomass if plateau	teoptbis	35	50			
	Rate of increase in the nitrogen harvest index	vitirazo	0.0085	0.0115			
Development stages	Cumulated development units between AMF and LAX	stamflax	1000	2100			
	Cumulated development units between LEV and AMF	stlevamf	50	400			
	Cumulated development units between LEV and DRP	stlevdrp	1000	1740			
	Maximum threshold temperature for development	tdmax	28	40			
	Minimum threshold temperature for development	tdmin	10	14			
Radiation interception	Coefficient of extinction of PAR in plant cover	extin	0.424	0.699			
Foliage	Compensation between number of stems and density of plants	adens	-1	-0.2			
	Minimum density as from which there is competition between plants for leaf growth	bdens	2	10			
	Maximum rate of production of leaf surface area	**dlaimax**	**0.0002**	**0.0015**	**uniform**	**0.00022**	**0.0011**
	Coefficient of sink strength of vegetative organs	sbv	127.5	172.5			
	Maximum temperature for growth	tcmax	35	42			
	Minimum temperature for growth	tcmin	10	14			
	Stress threshold from which there is an effect on the LAI	tutressmin	0	1			
Mineralization	Organic nitrogen content in moisture soil horizon	Norg	0.05	0.2			
	Reference temperature for soil mineralization	tref	15	27			
Roots	**Growth rate of root front**	**croirac**	**0**	**0.2**	**uniform**	**0.07**	**0.092**
	Depth of tillage	zlabour	17	23			
	Depth at which root density is reduced by half compared with surface	zpente	24	110			
	Maximum depth of root profile	zprlim	111	140			
Soil	Thickness of third soil horizon	epc3	5	60			
Water/Nitrogen stress	Nitrogen absorption rate by the plant's roots	absodrp	0	1			
	Minimum INN value possible for the crop	**INNmin**	**0**	**0.5**	**uniform**	**0.3**	**0.3**

ORCHIDEE							
Allocation		f_fruit	0.05	0.5			
	Maximum LAI per PFT	lai_max	3	9			
	Average critical age for leaves	leaf_age_crit	30	200			
	Upper bounds for leaf allocation	max_lto_lsr	0.25	0.5			
	Lower bounds for leaf allocation	min_lto_lsr	0.05	0.24			
	Root allocation	R0	0.05	0.5			
	Sapwood allocation	S0	0.05	0.5			
Photosynthesis	**Extinction coefficient**	**ext_coef**	**0.5**	**0.9**	**uniform**	**0.5**	**0.72**
	Slope of relationship between assimilation and stomatal conductance	**gsslope**	**7**	**11**	**beta(2,2)**	7.7	**9.5**
	Temperature at which photosynthesis is maximal	tphoto_max	30	45			
	Temperature at which photosynthesis is minimal	**tphoto_min_c**	**12**	**19**	**uniform**	12	**16.7**
	Temperature at which photosynthesis is optimal	**tphoto_opt**	**24**	**36**	**uniform**	24	**36**
	Maximum carboxylation rate	**vcmax_opt**	**40**	**100**	**beta(2,2)**	64	**81.3**
Respiration	**Fraction of biomass available for growth respiration**	**frac_growthresp**	**0.2**	**0.5**	**beta(2,2)**	**0.23**	**0.3**
	Slope of the relationship between temperature and maintenance respiration	**maint_resp_slope1**	**0.08**	**0.16**	**beta(2,2)**	**0.11**	**0.12**
Water budget	**Root profile to determine soil moisture content available to plants**	**humcste**	**0.8**	**7.2**	**uniform**	**3.2**	**4.1**

Table 3: Uncertainty associated with *STICS*, *ORCHIDEE*, or *ORCHIDEE+STICS* parameters uncertainties expressed as percentage of the reference harvested biomass for each site and for each of the two uncertainty analysis.

		Total Uncertainty (% of observed value)	ORCHIDEE Uncertainty (% of observed value)	STICS Uncertainty (% of observed value)
Expert-based parameters' uncertainties	Ayr 91-92	35.11	20.43	20.73
	Ayr 92-93	27.21	25.26	9.31
	Ingham 92-93	38.60	31.42	21.04
	Grafton 94-95	26.05	23.92	14.07
	Piracicaba 04-05	25.49	23.36	14.00
	Colimacons 94-95	41.21	41.87	18.61
	Tirano 98-99	44.26	36.80	30.61
Optimization-based parameters' uncertainties	Ayr 91-92	31.20	14.01	25.64
	Ayr 92-93	15.84	15.60	4.58
	Ingham 92-93	21.66	22.35	9.19
	Grafton 94-95	16.84	15.25	9.81
	Piracicaba 04-05	14.67	14.80	5.84
	Colimacons 94-95	21.31	20.01	10.28
	Tirano 98-99	22.26	18.06	15.03