



**ÉCOLE CENTRALE DES ARTS  
ET MANUFACTURES  
“ÉCOLE CENTRALE PARIS”**

**Dissertation**  
**presented by: Qiongli WU**  
**for the degree of Doctor of Applied Mathematics**

Speciality: Applied Mathematics  
Laboratory: Mathématiques Appliquées aux Systèmes (MAS)

## **Sensitivity Analysis for Functional Structural Plant Modelling**

Date of defense :19/04/2012

Composition of jury:

<b>M. Stefano TARANTOLA</b>	(President, Reviewer)
<b>M. Gerhard BUCK-SORLIN</b>	(Reviewer)
<b>M. Enrico ZIO</b>	
<b>Ms. Amélie MATHIEU</b>	
<b>M. Paul-Henry COURNÈDE</b>	(Supervisor)

N° ordre : 2012ECAP0021



# Sensitivity Analysis for Functional Structural Plant Modelling

Global sensitivity analysis has a key role to play in the design and parameterization of functional-structural plant growth models (FSPM) which combine the description of plant structural development (organogenesis and geometry) and functional growth (biomass accumulation and allocation). Models of this type generally describe many interacting processes, count a large number of parameters, and their computational cost can be important. The general objective of this thesis is to develop a proper methodology for the sensitivity analysis of functional structural plant models and to investigate how sensitivity analysis can help for the design and parameterization of such models as well as providing insights for the understanding of underlying biological processes. Our contribution can be summarized in two parts: from the methodology point of view, we first improved the performance of the existing Sobol's method to compute sensitivity indices in terms of computational efficiency, with a better control of the estimation error for Monte Carlo simulation, and we also designed a proper strategy of analysis for complex biophysical systems; from the application point of view, we implemented our strategy for 3 FSPMs with different levels of complexity, and analyzed the results from different perspectives (model parameterization, model diagnosis).

**Keywords:** Sensitivity analysis, FSPM, SRC, Sobol's method, Non-linearity assessment, Error estimation, GreenLab, NEMA



# Analyse de Sensibilité pour la Modélisation Structure-Fonction des Plantes

L'analyse de sensibilité globale a un rôle clé à jouer dans la conception et la paramétrisation des modèles structure-fonction de la croissance des plantes (FSPM). Ceux-ci combinent la description du développement structurel des plantes (organogénèse et géométrie) et de leur croissance fonctionnelle (accumulation de biomasse et allocation). Les modèles de ce type décrivent généralement de nombreux processus en interaction, comptent un grand nombre de paramètres et leur coût de calcul peut être important. L'objectif de cette thèse est de développer une méthodologie appropriée pour l'analyse de sensibilité des modèles structure-fonction des plantes et d'étudier comment l'analyse de sensibilité peut aider à la conception et la paramétrisation de ces modèles, ainsi qu'à l'analyse et la compréhension des processus biologiques en jeu. Notre contribution peut être vue en deux parties : du point de vue méthodologique et du point de vue de l'application des méthodes aux modèles. D'un point de vue méthodologique, nous avons tout d'abord amélioré les performances de la méthode de Sobol pour le calcul des indices de sensibilité en termes d'efficacité de calcul, avec un meilleur contrôle de l'erreur d'estimation par les simulations de Monte Carlo. Nous avons également conçu une stratégie d'analyse adaptée aux systèmes biophysiques complexes. Du point de vue applicatif, nous avons implémenté notre stratégie pour 3 FSPMs avec des niveaux de complexité différents, et nous avons analysé les résultats selon différents aspects, paramétrisation et diagnostic de modèles.

**Mots-clés :** Analyse de sensibilité, FSPM, SRC, Sobol, Indice de linéarité, Estimation de l'erreur, GreenLab, NEMA



# CONTENTS

<b>Liste of figures</b> . . . . .	11
<b>Liste of tables</b> . . . . .	13
<b>1. Introduction</b> . . . . .	15
1.1 General background . . . . .	15
1.2 Important issues in SA of FSPMs . . . . .	21
1.3 Objectives of this work . . . . .	24
1.4 Organization of the dissertation . . . . .	25
<b>Part I Preliminaries</b> . . . . .	27
<b>2. FSPM</b> . . . . .	29
2.1 General concepts about FSPM . . . . .	29
2.2 Functional-structural plant model design steps . . . . .	31
2.3 Limitations and challenges of FSPMs . . . . .	34
2.4 Introduction to several FSPMs . . . . .	35
2.4.1 GreenLab model for maize . . . . .	35
2.4.2 GreenLab model for poplar tree . . . . .	37
2.4.3 NEMA model for wheat . . . . .	40
2.5 Concluding remarks . . . . .	41
<b>3. Sensitivity analysis for FSPM</b> . . . . .	45
3.1 Uncertainty analysis and sensitivity analysis . . . . .	45
3.2 Model definition . . . . .	46
3.3 Model inputs . . . . .	47
3.4 Purpose of sensitivity analysis . . . . .	48
3.5 SA methods . . . . .	50
3.5.1 Classification of methods . . . . .	50
3.5.2 Description of methods . . . . .	51
3.6 Basic steps of SA for FSPM . . . . .	56
3.7 Concluding remarks . . . . .	58

<b>Part II</b>	<b>Algorithm and methodology design</b>	59
<b>4.</b>	<b>An Efficient Computational Method for Global Sensitivity Analysis . . .</b>	61
4.1	Computation of sensitivity indices . . . . .	61
4.1.1	General concepts of Sobol's sensitivity analysis . . . . .	61
4.1.2	Sobol's computing method and H-S improvement . . . . .	63
4.1.3	A new method to compute Sobol's indices . . . . .	65
4.2	Error estimation for Sobol's method . . . . .	68
4.2.1	Standard error, probable error . . . . .	68
4.2.2	For Sobol's formula . . . . .	69
	Error estimation for $S_i$ . . . . .	70
	error estimation for $ST_i$ . . . . .	72
4.3	Computational Tests . . . . .	72
4.3.1	Analytical functions . . . . .	72
4.3.2	error estimation test . . . . .	73
4.3.3	Comparison of H-S and proposed method . . . . .	73
4.4	Discussion . . . . .	74
<b>5.</b>	<b>Strategy design . . . . .</b>	77
5.1	Introduction . . . . .	77
5.2	SRC method and non-linearity assessment . . . . .	79
5.3	A methodology for complex biophysical models . . . . .	82
5.4	Processing for functional output case . . . . .	88
5.5	Concluding remarks and discussions . . . . .	91
<b>6.</b>	<b>Software platform: PyGMAIion . . . . .</b>	93
6.1	Background and objectives . . . . .	93
6.2	PyGMAIion . . . . .	94
6.3	Implementation in PyGMAIion . . . . .	95
6.4	Sensitivity analysis in PyGMAIion . . . . .	98
<b>Part III</b>	<b>Applications and simulations</b>	101
<b>7.</b>	<b>Methodology practice: application to FSPMs . . . . .</b>	103
7.1	GreenLab maize . . . . .	103
7.1.1	Local sensitivity analysis and its normalized version . . . . .	103
7.1.2	Standardized Regression coefficients . . . . .	105
7.1.3	Sobol's method . . . . .	107
7.1.4	Discussion . . . . .	111
7.2	Poplar tree model . . . . .	112
7.3	Model of C-N dynamics (NEMA) . . . . .	115
7.3.1	Output AreaGreenTotal . . . . .	116

Contents	9
7.3.2 Output Production . . . . .	127
7.3.3 Output DMgrains . . . . .	132
7.3.4 Output Ngrains . . . . .	135
7.3.5 Output RootNuptake . . . . .	140
7.3.6 Comprehensive comparison . . . . .	145
7.4 Conclusions and discussion . . . . .	147
<b>Part IV Conclusions and perspectives</b>	149
<b>8. Conclusions and perspectives</b> . . . . .	151
8.1 Contributions . . . . .	151
8.2 Discussion and perspectives . . . . .	157
8.3 To conclude . . . . .	160
<b>Appendix</b>	163
<b>A. Approximation of the moments of functions of random variables</b> . . . . .	165
<b>Bibliography</b> . . . . .	169
<b>Acknowledgment</b> . . . . .	179
<b>Publications</b> . . . . .	181



# LIST OF FIGURES

1.1	Methodology: Sketch of the various SA techniques available . . . . .	22
2.1	Poplar tree: topology after 30 time steps . . . . .	38
2.2	NEMA: Overview scheme . . . . .	42
5.1	Non-linearity assessment: GreenLab Maize . . . . .	80
5.2	Non-linearity assessment: Poplar tree . . . . .	81
5.3	Non-linearity assessment: NEMA . . . . .	81
5.4	Methodology: Mesh grid 3 dimension parameter space . . . . .	84
5.5	Methodology: $ST_{\Omega_i}^g$ for one module decomposition scheme . . . . .	88
5.6	Methodology: $S_i$ for 35 parameters . . . . .	89
5.7	Methodology: Total variance of output . . . . .	90
6.1	PyGMAIion: Sensitivity analysis configuration . . . . .	99
7.1	GreenLab Maize: Biomass allocation at each GC . . . . .	104
7.2	GreenLab maize: Local sensitivity analysis . . . . .	105
7.3	GreenLab maize: SRCs for the outputs $q$ . . . . .	107
7.4	GreenLab maize: $\sum(\gamma_{x_i})^2$ and $R_Y^2$ . . . . .	107
7.5	GreenLab maize: $S_i$ and $ST_i$ for $\mu$ . . . . .	108
7.6	GreenLab maize: linearity indexes, $S_p$ and $\mu$ fixed . . . . .	109
7.7	GreenLab maize: Sobol's method for SA, $S_p$ and $\mu$ fixed . . . . .	110
7.8	GreenLab maize: $ST_i$ and $S_i$ for $\alpha_b$ . . . . .	110
7.9	GreenLab maize: Interactions between $\alpha_b$ to the other parameters . . . . .	111
7.10	Poplar tree: Oscillation of Q/D . . . . .	112
7.11	Poplar tree: Number of phytomers of physiological age 4 . . . . .	113
7.12	Poplar tree: Evolution of linearity index . . . . .	113
7.13	Poplar tree: Sensitivity indexes for all the model parameters . . . . .	114
7.14	Poplar tree: Sensitivity indices when fixing $\mu$ and $S_p$ . . . . .	115
7.15	NEMA: Output AreaGreenTotal, Linearity . . . . .	117
7.16	NEMA: Output AreaGreenTotal, $S_{\Omega_i}^g$ . . . . .	118
7.17	NEMA: Output AreaGreenTotal, $ST_{\Omega_i}^g - S_{\Omega_i}^g$ . . . . .	119
7.18	NEMA: Output AreaGreenTotal, $S_{\Omega_{ij}}^g$ . . . . .	119
7.19	NEMA: Output AreaGreenTotal, Selected factors $S_i$ . . . . .	124
7.20	NEMA: Output AreaGreenTotal, Selected factors $ST_i - S_i$ . . . . .	126

7.21	NEMA: Output Production, Linearity . . . . .	128
7.22	NEMA: Output Production, $S_{\Omega_i}^g$ . . . . .	128
7.23	NEMA: Output Production, $ST_{\Omega_i}^g - S_{\Omega_i}^g$ . . . . .	129
7.24	NEMA: Output Production, $S_{\Omega_{ij}}^g$ . . . . .	129
7.25	NEMA: Output Production, Selected factors $S_i$ . . . . .	130
7.26	NEMA: Output Production, Selected factors $ST_i - S_i$ . . . . .	131
7.27	NEMA: Output DMgrains, Linearity . . . . .	132
7.28	NEMA: Output DMgrains, $S_{\Omega_i}^g$ . . . . .	133
7.29	NEMA: Output DMgrains, Selected factors $S_i$ . . . . .	134
7.30	NEMA: Output Ngrains, Linearity . . . . .	135
7.31	NEMA: Output Ngrains, $S_{\Omega_i}^g$ . . . . .	136
7.32	NEMA: Output Ngrains, $ST_{\Omega_i}^g - S_{\Omega_i}^g$ . . . . .	136
7.33	NEMA: Output Ngrains, $S_{\Omega_{ij}}^g$ . . . . .	137
7.34	NEMA: Output Ngrains, Selected factors $S_i$ . . . . .	138
7.35	NEMA: Output Ngrains, Selected factors $ST_i - S_i$ . . . . .	139
7.36	NEMA: Output RootNuptake, Linearity . . . . .	140
7.37	NEMA: Output RootNuptake, $S_{\Omega_i}^g$ . . . . .	141
7.38	NEMA: Output RootNuptake, $ST_{\Omega_i}^g - S_{\Omega_i}^g$ . . . . .	141
7.39	NEMA: Output RootNuptake, $S_{\Omega_{ij}}^g$ . . . . .	142
7.40	NEMA: Output RootNuptake, Selected factors $S_i$ . . . . .	143
7.41	NEMA: Output RootNuptake, Selected factors $ST_i - S_i$ . . . . .	144
8.1	Brief illustration of improved Sobol's estimator. . . . .	152

# LIST OF TABLES

2.1	GreenLab Maize: Parameter Distributions . . . . .	37
2.2	GreenLab poplar tree: Parameter Distributions . . . . .	40
2.3	NEMA: Model parameters . . . . .	43
4.1	Sobol's Estimator improvement . . . . .	66
4.2	H-S Sobol's method for $S_i$ . . . . .	74
4.3	H-S Sobol's method for $ST_i$ . . . . .	75
4.4	Efficiency comparison of H-S and new Sobol's method for $S_i$ . . . . .	75
4.5	Efficiency comparison of H-S and new Sobol's method for $ST_i$ . . . . .	76
7.1	GreenLab maize: SRC index and $R^2$ at selected typical GC . . . . .	106
7.2	NEMA: List of tables and figure describing the results of SA . . . . .	116
7.3	NEMA: Output AreaGreenTotal, Module DMfluxes . . . . .	120
7.4	NEMA: Output AreaGreenTotal, Module Nfluxes . . . . .	121
7.5	NEMA: Output AreaGreenTotal, Module Photosynthesis . . . . .	121
7.6	NEMA: Output AreaGreenTotal, Module RootNuptake . . . . .	122
7.7	NEMA: Output AreaGreenTotal, Module Tissuedeath . . . . .	122
7.8	NEMA: Output AreaGreenTotal, Overall parameter selection analysis . . . . .	122
7.9	NEMA: Output AreaGreenTotal, Intra-Module Interaction analysis . . . . .	123
7.10	NEMA: Output AreaGreenTotal, Inter-Module Interaction . . . . .	123
7.11	NEMA: Output AreaGreenTotal, $S_{ij}$ for selected factors . . . . .	125
7.12	NEMA: Output Production, Selected factors list . . . . .	130
7.13	NEMA: Output Production, Overall parameter selection analysis . . . . .	130
7.14	NEMA: Output DMgrains, Selected factors . . . . .	133
7.15	NEMA: Output DMgrains, Overall parameter selection analysis . . . . .	133
7.16	NEMA: Output Ngrains, Selected factors . . . . .	137
7.17	NEMA: Output Ngrains, Overall parameter selection analysis . . . . .	137
7.18	NEMA: Output Ngrains, Intra-Module Interaction analysis . . . . .	137
7.19	NEMA: Output Ngrains, Inter-Module Interaction . . . . .	138
7.20	NEMA: Output Ngrains, $S_{ij}$ at thermal time after flowering 685.063 °Cd . . . . .	138
7.21	NEMA: Output RootNuptake, Selected factors . . . . .	142
7.22	NEMA: Output RootNuptake, Overall parameter selection analysis . . . . .	142
7.23	NEMA: Output RootNuptake, Intra-Module Interaction analysis . . . . .	143
7.24	NEMA: Output RootNuptake, Inter-Module Interaction . . . . .	143
7.25	NEMA: Selected $S_i$ ranking in overall model for different outputs . . . . .	146



# 1. INTRODUCTION

## 1.1 General background

### Mathematical modelling

Mathematical modelling is a key tool for the analysis of a wide range of real-world phenomena ranging from physics and engineering to chemistry, biology and economics [Weiße, 2009]. A mathematical model is defined by a series of equations, input factors, parameters, and state variables to characterize the processes being investigated. An increasing number of models have been developed in recent decades thanks to the progress of computational and statistical tools.

The purposes of modelling include: 1) integration of knowledge (exceeding the capacity of the human brain), 2) quantitative testing of hypotheses, 3) extrapolation of effects of factors beyond the range of conditions covered experimentally, 4) revealing of knowledge gaps and ‘guiding’ research, and 5) support of practical management decisions. In research environments, modelling commonly serves purposes such as integrating knowledge or the quantitative testing of hypotheses [Vos et al., 2007]. Through a better understanding of phenomena and the prediction and simulation of the impact of decisions, the different models developed often have an application goal like agricultural advice, and may in some cases serve as a tool for decision support for economic and political decision makers.

Modelling of natural phenomena is always facing several sources of uncertainty which should be considered qualitatively or quantitatively for modelers [De Rocquigny et al., 2008]. As we will see, it is particularly true in life sciences. For dynamical models describing the mechanisms of a given phenomenon, we distinguish generally two main sources of uncertainty: uncertainty about the structure of the model describing the phenomenon like the form of the equation  $f(\bullet)$ , and uncertainty in model inputs (parameters and external variable inputs). We do not consider in this thesis the first type of uncertainty although the latter is related to the model structure.

In general, the more the model integrates the relevant knowledge, the better it represents the phenomenon under study for a given level of complexity. We are interested in the statistical methods to identify the most important relevant knowledge to be included in a model in the process of modelling. The parameters carrying out the relevant knowledge are then used in the mathematical formulation to describe the phenomena under investigation. Then some logical questions arise: what subset of

parameters is more important according to our research aims? How to minimize the number of parameters to summarize the representation of the phenomena? How do the interactions between parameters or subsets of parameters work?...etc. All these questions can be answered by uncertainty analysis and sensitivity analysis. The uncertainty of a factor is from either the factor in question is not very well known for the true value, or it is subject to inherent variability [Wallach et al., 2002]. In the following, the uncertainty of a factor will be described by a probability distribution that can represent the possible values of the factor according to the scope of analysis. Uncertain parameters and inputs contribute to the variance of model outputs.

For us, a model designates the mathematical equations about the phenomenon. Considering a dynamic model represented by the following mathematical equation:  $Y(t) = f(\mathbf{Z}, \theta, t)$ , where  $\mathbf{Z}$  is the vector of input variables of the model,  $\theta$  is the vector of uncertain parameters and  $Y(t)$  is the model output at time  $t$ , for  $t \in 1, 2, \dots, T$ , the function  $f(\bullet)$  describes the phenomenon studied and it is either deterministic or stochastic. Our work in this thesis only concentrates on deterministic models and stochastic models will not be discussed in detail but kept in future possible work.

In the following, we will consider both input variables and uncertain parameters as potential sources of uncertainty and call them uncertain input ‘factors’ denoted by  $\mathbf{X}$  when we talk about sensitivity analysis. Then the dynamic model equation is written:  $Y(t) = f(\mathbf{X}, t)$ .

### **Functional-structural plant models (FSPM)**

Plant modelling has become a key research activity, with the objective of developing applications in agriculture, forestry and environmental sciences. The use of an interdisciplinary approach is necessary to advance research in plant growth modelling. It involves a variety of disciplines including botany, biology, ecology, mathematics and information technology, as such to handle this complexity for the plant growth modellers is a very challenging aspect. Due to the growth of computer resources and the sharing of experiences between biologists, mathematicians and computer scientists, the development of plant growth models has progressed enormously during the last two decades [Fourcaud et al., 2008]. Plant growth modelling represents a necessary tool for understanding plant growth and developing predictive tools for decision making. For this purpose, models have been developed to simulate biomass production and distribution among organs, in interaction with the environment [De Reffye et al., 2008].

Initially process-based models (PBM) were developed separately from structural (or: architectural or morphological) plant models (SPM). Combining PBMs and SPMs into functional-structural plant models (FSPM) or virtual plants has become feasible particularly thanks to the progress in modelling and computational science [Sievänen et al., 2000b]. This adds a dimension to classical crop growth modelling [Vos et al., 2010]. FSPM are particularly suited to analyse problems in which the spatial structure

of the system is an essential factor contributing to the behaviour of the system of study [Vos et al., 2007].

Functional structural plant models (FSPMs) can be defined as models that combine the descriptions of both metabolic (physiological) processes and the structure development of a plant. They usually contain the following components 1) Presentation of the plant structure in terms of basic units, 2) Rules of morphological development and 3) Models of metabolic processes that drive plant growth. The main emphasis in these applications has been individual plants [Sievänen et al., 2009]. Since FSPMs usually describe the evolution with time of the state variables characterizing plant growth, they are generally dynamic models.

Due to the detailed description of the plant structure in FSPMs generally at organ level, and sometimes, of the local environment of each organ [Chelle, 2005], the FSPMs tend to require a large number of parameters and input data. Owing to the large amount of information they contain about the plant and the number of processes they aim at describing, they also tend to be computationally heavy. Moreover, the complicated and interacting biophysical processes governing plant growth bring a large amount of uncertainty into FSPMs: field surveys for collecting the necessary data for the development of models are difficult and expensive. In consequence, input data (environmental factors) and experimental data from which model parameters are estimated are also characterized by great uncertainty. Finally, modelling complex processes, model parameters estimation and input data collection all contribute to model uncertainty [Monod et al., 2006], [Wallach et al., 2002].

Good modelling practice requires that the modeler provides an evaluation of the confidence in the model. Uncertainty analysis (UA) and Sensitivity analyses (SA) offer valid tools for this evaluation. This thesis is mainly focused on sensitivity analysis, but since in practice, uncertainty and sensitivity analyses are most often run in tandem, the implementation of an uncertainty analysis and the implementation of a sensitivity analysis are very closely connected from both a conceptual and computational point of view [Helton et al., 2006b] and customarily lead to an iterative revision of the model structure. So when we study sensitivity analysis, we will generally mention uncertainty analysis.

### **Sensitivity analysis for FSPM**

A possible definition of sensitivity analysis is the following: ‘The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input’ [Saltelli et al., 2004].

Sensitivity Analysis (SA) has the role of ordering by importance the strength and relevance of the inputs in determining the variations of the output variables of interest. Such information may provide some help for model assessment:

- Measurement of model adequacy (e.g. does the model fit observation?)
- Knowledge of model relevance (e.g. is the model-based inference robust?)
- Identification of critical regions in the inputs' space (e.g. which combination of factors corresponds to the highest risk?)
- Detection of interactions between factors
- Priorities for research and experimentations
- Simplification of model structure [Saltelli et al., 2004].

Notice that the model output for which sensitivity analysis is performed is classically a scalar value. So SA are usually performed separately at each calculation time step for FSPMs. In the case of dynamic crop models, simulations are usually computed at a daily time step and so is the sequential implementation of sensitivity analysis at each simulation date with one index per parameter per simulation date.

Sensitivity analysis has an important role to play in functional-structural plant growth modelling by assessing the different sources of uncertainty. In this recent field of research in plant biology, models are not yet stable [Sievänen et al., 2000a]. FSPMs aim to describe the plant structural development (organogenesis and geometry), the functional growth (biomass accumulation and allocation) and the complex interactions between both. The complexity of the underlying biological processes, especially the interactions between functioning and structure [Mathieu et al., 2009], usually makes it very difficult to identify the key physiological processes described by the model. The same problem exists for parameter estimation for which we need to discriminate the parameters with different levels of importance so as to deduce the proper method for estimation processing. Likewise, in the process of experiment design, if we have some guide information about the priority of parameters we need to estimate from experiments, the cost of experiments can be more effectively arranged by more frequent measurements and more accurate study of those that contribute to the output variables, and vice versa.

Typically, for the problem of parameterization, there are two groups of parameters in the model: the observed ones that can be directly measured by experimental observations, and the hidden ones, that can not be measured directly and have to be estimated by model reversion [Cournède et al., 2011]. For the observed parameters, we may need to be clear about the level of experimental data accuracy, so that for those parameters that mostly contribute to the variability of the output, more attention should be paid to. Regarding the hidden parameters, there is also a proper balance to be found between the number of parameters used to describe the biophysical processes and the complexity of their estimation, which is always a bottleneck in the modelling process. Thanks to sensitivity analysis, we can rank the parameters by their significances to the model output. According to the SA results, we can fix the

least influential parameters, and we should pay more attention to those who play important roles in the outputs' variances. In both cases, the sensitivity analysis may help to optimize the trade-off between experimental cost and accuracy [Wu and Cournède, 2009]. With the higher order SA indices and the methods of grouping together parameters, SA can help computing the level of interactions between parameters or group of parameters, in order to know the interactions' contributions to the variance of the output, hence to know the main processes under investigation.

The most common classifications of SA methods are either distinguished between quantitative and qualitative methods or between local and global techniques.

- Qualitative methods aim at selecting the 'most important' subset of parameters, while quantitative techniques can be designed to give information on the amount of variance explained by each factor.
- In local approaches (known as one-at-a-time, OAT), the effect of a single factor's variation is estimated while keeping all the others fixed at their average values. However they cannot include the effect of the shape of the density functions of the inputs, and they are not model-independent.
- Global approaches estimate the effect on the output of a factor keeping all the others varying. Generally, global approaches use model-independent methods while not requiring assumptions of additivity or linearity. As a drawback, they are usually computationally expensive [Cariboni et al., 2007].

The Standardized Regression Coefficients (SRC) can be viewed as an interesting trade-off between local and global methods, regarding the advantages and shortcomings of both: the accuracy of the analysis and the computing cost. It is based on the linear approximation of the model and Monte Carlo simulations. SRC method takes into account the shape of the probability distribution of every factor. The other important index produced by SRC is the model coefficient of determination,  $R^2$ , which represents the fraction of the output variance explained by the linear regression model itself. A side result of the model coefficient of determination ( $R^2$ ) is that it provides an indicator of the degree of non-linearity of the model representing the level of interaction between parameters and how this interaction contributes to the variance of the output. When  $R^2 = 1$ , the system is linear and the SRCs can totally explain the variance of the output affected by each factor. Even when models are moderately non-linear (i.e.  $> 0.9$ ), the SRCs can provide valid qualitative information. When  $R^2$  gets small, the SRCs are no longer reliable sensitivity representations. To be more direct, we consider this model coefficient of determination as a linearity index, and use it to assess the non-linearity of models.

Sampling-based approaches to uncertainty and sensitivity analysis are both effective and widely used [Helton et al., 2006a]. One important category of it are 'Variance

based' methods. The basic concept for this kind of method is to decompose the output variance into the contributions imputable to each input factor. The most widely used methods are the FAST (Fourier Amplitude Sensitivity Test, see [Cukier et al., 1973, 1978; Koda et al., 1979]), and Sobol's method, see [Sobol, 1993]. FAST method decomposes the output variance  $V(Y)$  by means of spectral analysis. Sobol's method is based on the same decomposition of variance, which is achieved by Monte Carlo methods in place of spectral analysis. In this thesis, Sobol's method that we widely investigate plays a key role [Sobol, 1993], since the different types of sensitivity indices that it estimates can fulfill different objectives of sensitivity analysis: factor prioritization, factor fixing, variance cutting or factor mapping [Saltelli et al., 2004]. It is a very informative method but potentially computationally expensive [Helton et al., 2006a].

For a given factor  $X_i$ , the value of first-order Sobol's index  $S_i$  indicates whether a factor is mainly influent, while an important difference between  $ST_i$  (Total order effect) and  $S_i$  flags an important role of interactions for that factor regarding the output  $Y$ . If this is the case, inspection of the second order index  $S_{ij}$  for all  $i \neq j$  will allow us to identify which factor  $X_i$  interacts with <sup>1</sup>. In fact, beside the first-order effects, Sobol's method also aims at determining the levels of interaction between parameters [Wu and Cournède, 2010]. In [Saltelli and Tarantola, 2002], the authors also devised a strategy for sensitivity analysis that could work for correlated input factors, based on the first-order and total-order indices from variance decomposition.

Basically, SA for models has to be performed in an orderly fashion. In practice the development of SA often proceeds in a loop, since the modelers may not have enough knowledge about the attributes related to the decisions in SA, including the range of parameters' uncertainty, lack of experimental data etc. The steps outlined below are more or less in a logical and chronological order, but there are numerous reasons to deviate from the sequence that is presented.

1. The first step is to establish the goal of our sensitivity analysis and consequently to define the output functions that answer the questions. The aims of SA can be 'Factor Prioritization', 'Factor Fixing', 'Variance Cutting' and 'Factor Mapping'. In most of the references, when we say SA for screening, it intends to help identifying parameters that are not important, and could be fixed to their mean values. What we mentioned in the definition of mathematical model 'the output of the model' corresponds to the state variables of the model which plays very important role in the model.
2. Next we need to decide which input factors should be included in our analysis, that is to say, to define the parameter space for sensitivity analysis based on the objective issue of the first step. At this level, trigger parameters can be defined, allowing one to sample across model structures, hypotheses, etc. After deciding the parameter space, we need to choose a distribution function for each of the input factors.

---

<sup>1</sup> Subscript  $i$  and  $j$  stand for the factors with uncertainty and under sensitivity analysis in the model

3. Afterwards, the task is to choose sensitivity methods or to design the strategy if it must be the combination of more than one SA method. Two classes of methods exist: local methods and global ones.
4. Then we start the Monte Carlo simulation sampling of the input factors for the analysis, afterwards we evaluate the model on the generated sample and produce the output, which contains N (sampling number) output values.
5. Lastly, we analyse the model outputs using the estimators provided by the chosen methods and draw our conclusions based on the analysis result.

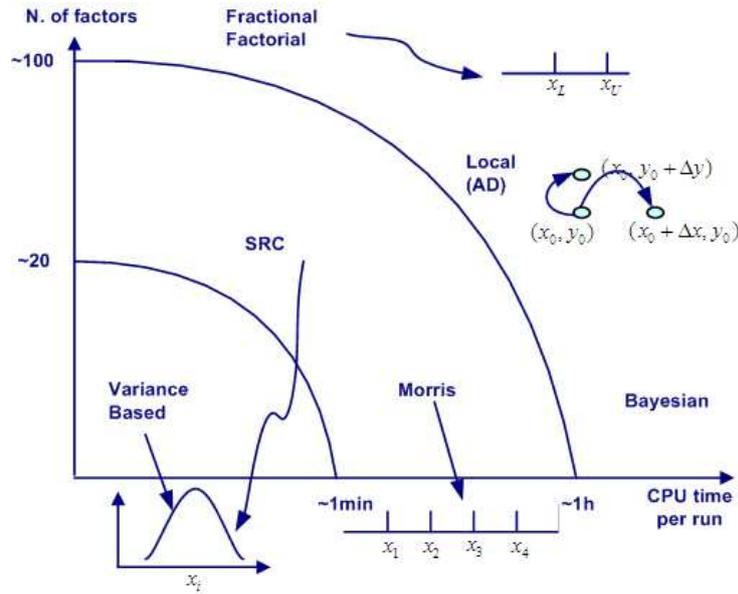
## 1.2 Important issues in SA of FSPMs

If sensitivity analysis is quite usual in crop and plant growth models, it had long been restricted to local sensitivity analysis or to analysis of variance for linear models. In recent years, global sensitivity analysis has brought increasing interest to assess the relative importance of parameters in ecological models [Cariboni et al., 2007] or crop models [Makowski et al., 2006]. In [Pathak et al., 2007], the author investigates whether global sensitivity analysis would provide better information on the importance of model parameters than the simpler and commonly used local sensitivity analysis method. [Makowski et al., 2006] uses global sensitivity analysis method like both the variance-based method and extend-FAST to fulfill the aim of model simplification by reducing the number of parameters. Our interest is to investigate the proper methodology of sensitivity analysis for FSPMs, which are generally more complex than crop models and for which it may not be possible to apply classical methods in a straight forward way.

### Computational issue

As mentioned in the previous section, FSPMs tend to require a large number of parameters and/or input data and owing to the large amount of information they contain about the plant to be modeled, they also tend to be computationally heavy. In [Cariboni et al., 2007], the author pointed out that the choice of the most suitable technique for sensitivity analysis depends on the number of factors of the model and on the CPU time required to run it as shown in fig.1.1.

However, it is of crucial importance to locate the interactions between parameters. SA can help computing the level of interactions between parameters with the higher order SA index, so that to evaluate how this interaction contributes to the variance of the output. So on the one hand, we want to use a global sensitivity analysis method like Sobol's to locate the quantitative interaction information for the model and on the other hand since FSPMs usually have large number of parameters and the model evaluation is computationally heavy, the implementation of the strategy faces a great challenge for the computing cost issue. So it deserves our effort to improve the computing efficiency of the method itself.



**Fig. 1.1 :** [Cariboni et al., 2007] Sketch of the various techniques available and their use as a function of computational cost of the model and dimensionality of the input space. AD: automated differentiation, SRC: Standard regression coefficient, Morris: [Morris, 1991]

Specifically for the Sobol's method, computational methods to evaluate Sobol's indices sensitivity rely on Monte-Carlo sampling and re-sampling [Sobol, 1993], [Homma and Saltelli, 1996]. For a  $k$  dimensional factor of model uncertainty, the  $k$  first-order effects and the 'k' total-order effects are rather expensive to estimate, needing a number of model evaluations strictly depending upon  $k$  [Saltelli et al., 2010]. Therefore, it is crucial to not only devise efficient computing techniques, in order to make best use of model evaluations [Saltelli, 2002], but also to have a good control of the estimation accuracy with respect to the number of samples. Error estimation is of crucial interest to check whether the SA computing has properly converged. Moreover, it can be used to give confidence bounds of the result. Previous work as in [Homma and Saltelli, 1996] gave interesting results about error estimation, but the conclusions are based on some restrictive assumptions.

## Strategy design

A good sensitivity analysis practice does not only needs well designed SA estimators (as discussed in the previous paragraph about computational issue) but also needs good understanding of how to comprehensively use more than one methods to make them be complementary to each other since different methods tackle different issues of interest. This is what we consider as 'strategy design'.

Though pointed out in [Saltelli et al., 2004], one property of an ideal sensitivity analysis method is that it should be 'model independent', which means a method

should work regardless of the attributes of the model itself like additive, linear, etc. However, the strategy design is necessary in our work of sensitivity analysis for FSPMs. More work needs to be done for exploring how global sensitivity analysis can help in the parameterization of FSPM, by quantifying the driving forces of the phenomena described by the models and the relative importance of the described biophysical processes regarding the outputs of interest.

Complex biological models are usually characterized by several interacting processes with sub-models describing each of them. Most FSPMs are such models. It is interesting to evaluate the importance of the sub-models (usually ‘function’ modules corresponding to the biophysical processes they describe) by sensitivity analysis. For this objective, in practice we need to firstly classify the parameters into different biological function modules according to the biologist modeller’s expert knowledge, then to check the joint sensitivity effects of the groups of parameters that belong to those modules. This is how ‘module-by-module’ analysis for complex biophysical system is put forward. The strategy design should be divided into several steps for which we choose different SA methods to fulfill different requirements.

The choice of a proper sensitivity analysis method to fulfill different aims of different sub-steps of the analysis faces the same SA general issues as mentioned before. However, module-by-module analysis requires us to make the combination of more than one SA method in order to make best use of each method’s advantages and to make them complementary to each other.

An early attempt for the ‘module-by-module analysis’ strategy was proposed by [Ruget et al., 2002]. The authors practice a variance-based analysis for the crop model STICS, with the objective of choosing the main parameters to be estimated. Analysis was made in two steps: first, within each meta-process (module), the most important parameters are identified; then sensitivity to the identified parameters is calculated taking into account all meta-processes together. The main factors addressed concern the interaction with the environment, which is of crucial interest. However, it was based on regression techniques, so strict requirements have to be fulfilled for the model functions: the model has to be linear, additive, or for surface response method, the model has to be a continuous system. For FSPMs, such hypothesis are not always satisfied.

As such, according to the past references [Pagano and Ratto, 2007], [Ruget et al., 2002], several more points need to be improved for a better application strategy for this module-by-module analysis:

- ‘A certain number of parameters (according to empirical information) are selected from each module for the inter-module analysis’. It is risky to rely on such empirical information that may be misleading afterwards, especially regarding the parameter space issue. There may be some parameters missed at this step, like the ones that have important effect on the output, for example,

through interactions with the others. And since each module mostly tend to have different importance in the model, if we decide empirically the number of parameters selected for each module, it may cause that for some modules we select not enough parameters and for some modules we select too many. This decision directly affects the importance evaluation in the final step. So we need a quantitative standard to choose the proper number of parameters from each module: even though in the internal analysis of each module, the results we get should be given in a unified framework to be comparable. In these regards, it corresponds to keeping the same sampling space while doing the Monte Carlo simulations.

- To consider parameters only in one module, while fixing the other parameters to their mean values, the SA indices obtained this way can not stand for the importance of the parameters in the complete space, but to a surface formed by the fixed values of the parameters in the space. Plus, by fixing the parameters in the other modules, the interactions between parameters from different modules will be eliminated even though it may prove to be important.

### 1.3 Objectives of this work

Considering the contradiction between the computing cost issue and the necessity of interest using Sobol's method for the quantitative information about sensitivity of models especially regarding the interaction information, our work aims at improving a computing method inspired by [Homma and Saltelli, 1996] so that best use of the model evaluations can be made. We also aim at deriving an estimator of the error of sensitivity indices evaluation with respect to the sampling size for this generic type of computational methods so that a better control of the Monte Carlo simulation convergence can be achieved.

Plants are known to be complex systems with a strong level of interactions and competitions, and the aim of FSPMs is to describe and understand this complexity. As such, non-linearity is expected to play a key role in the study. So our first objective for strategy design is about this issue: to evaluate the non-linearity of the model by determination coefficient in the SRCs method application, and to check how it works as a preliminary step to provide us a general scheme for the next steps of the sensitivity analysis studies.

To make all steps of the 'module-by-module analysis' more quantitatively precise, we also aim at exploring an effective simulation design to help the sensitivity analysis for complex models with several logically distinct but biological functioning interacting modules, like the NEMA model.

We also aim at applying the developed strategy of sensitivity analysis to 3 FSPMs with different levels of complexity, and infer in each case what information can be

drawn from this analysis. The 3 FSPMs are firstly a simple source-sink model of maize growth, which is used to specifically study the process of carbon (C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations, secondly the GreenLab model of tree growth (applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis [Mathieu, 2006], which describes tree structural plasticity in response to trophic competition, lastly a functional-structural model, NEMA [Bertheloot et al., 2011a], describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering. This model is more mechanistic but also more complex than the two previous ones.

## 1.4 Organization of the dissertation

The dissertation is consist of four parts:

Part I presents the preliminaries of the thesis. Chapter 2 first gives an overview about FSPM and modelling techniques, especially about the attributes of FSPM that are important to be considered for application of SA, then the description of the 3 FSPMs with different complexity for the comprehensive methodology investigation. Chapter 3 introduces the basic concepts, the methods and scheme design of sensitivity analysis. Definitions and equations about all the indices of SA applied in our simulation are given.

Part II constitutes the main part of the thesis and introduces the methodological aspects we developed. The efficient computational method based on some improvement to make best use of the model evaluation in the numerical implementation technique is described in Chapter 4. In this chapter, we present the error estimation to control the convergence of Monte Carlo simulation in this algorithm, following which numerical tests are given. Chapter 5 introduces the strategy we proposed to study complex biophysical systems, mainly about the ‘module-by-module’ analysis scheme design and several important points to be paid attention to. To complete the numerical implementation issue, we present a new platform PyGMAIion in which we have implemented the algorithms relating to all our application practices in Chapter 6.

Part III presents the simulation results. It illustrates all the applications and results corresponding to the algorithm and strategy design presented in Part II. Chapter 7 gives out the result for the three FSPMs presented in Chapter 2. Some conclusions relating to model parameterization and model diagnosis are given.

Part IV summarizes the conclusions of our work and gives some perspectives of this work.

Some technical material is provided to the Appendix. It includes basic computations for the expectation and variance of functions of one and two dimensional random variables.



**Part I**

**PRELIMINARIES**



## 2. FUNCTIONAL-STRUCTURAL PLANT MODELLING

This chapter first gives an overview about the background of FSPM in section.2.1 and modelling techniques relating to FSPM in section.2.2 so that the interest of sensitivity analysis for model design can be clear. In section.2.3 we introduce some limitations and challenges in functional structural plant growth modelling, especially about the problems that can be potentially resolved by SA. Lastly, section.2.4 describes some specific FSPMs to which we applied sensitivity analysis in this thesis.

### 2.1 General concepts about FSPM

Plants and plant populations can be considered as complex systems, in a mathematical sense. Complex systems consist of high numbers of heterogeneous entities between which multi-scale interactions generate holistic behaviours (*i.e.* some emergent properties of the system cannot be deduced from the independent studies of its components) [Ricard, 2003].

One of the most interesting research in plant modelling has been the construction of integrated models representing both the function and structure of plants ([Room et al., 1996], [De Reffye et al., 1997], [Kurth and Sloboda, 1997], [Fournier and Andrieu, 1998], [Lacointe, 2000], [Godin et al., 2004]). These models are often termed functional structural plant models (FSPM). The general approach behind most of these models is to represent the plant as a relatively large number of interconnected components (such as leaves and internodes) and to separately model the various physical, chemical and physiological processes (such as light interception, photosynthesis and nutrient transport) that occur within and between these components ([Perttunen et al., 1998], [Sievänen et al., 2000b], [Lacointe, 2000], [Sinoquet and Le Roux, 2000]).

Initially process-based models (PBM) were developed separately from structural (or: architectural or morphological) plant models (SPM). Combining PBMs and SPM into functional-structural plant models (FSPM) or virtual plants has become feasible particularly thanks to the progress in modelling and computational science [Sievänen et al., 2000b]. This adds a dimension to classical crop growth modelling [Vos et al., 2010]. FSPM are particularly suited to analyse problems in which the spatial structure of the system is an essential factor contributing to the behaviour of the system of study [Vos et al., 2007].

Functional structural plant models (FSPMs) can be defined as models that combine the descriptions of both metabolic (physiological) processes and the structure development of a plant. They usually contain the following components 1) Presentation of the plant structure in terms of basic units, 2) Rules of morphological development and 3) Models of metabolic processes that drive plant growth. The main emphasis in these applications has been individual plants [Sievänen et al., 2009]. Though there are static FSPMs which can be useful to answer certain research question, a large number of FSPMs usually describe the evolution with time of the state variables characterizing plant growth, they are generally dynamic models.

Most FSPMs include the same basic processes, simulated at each time step [Letort, 2008]:

- initiation of new architectural units
- biomass production
- biomass partitioning

GreenLab [de Reffye and Hu, 2003] is a functional-structural model simulating the processes of biomass production and allocation into organs at the whole-plant scale. Organogenesis is driven by formal grammars determining the topological rules of organ initiation, production and organization [Cournède et al., 2006], [Loi and Cournède, 2008]. The state variable of this automaton is the differentiation state of apical meristems, called their physiological age [Barthélémy and Caraglio, 2007]. Three versions of the model can be considered: deterministic (*GL1*) [Yan et al., 2004], stochastic (*GL2*) [Kang et al., 2008] or mechanistic, *i.e.* deterministic with feedback of photosynthesis on organogenesis (*GL3*) [Mathieu, 2006] [Mathieu et al., 2009]. New versions are still being developed: *GL4* [Pallas et al., 2011] and *GL5* [de Reffye et al., 2012]. Biomass production is computed at each time step (growth cycle) depending on the plant total foliar surface and taking into account the effects of self-shading of leaves. Biomass is allocated to expanding organs regardless of their position (common pool of biomass) according to a source-sink model [Warren-Wilson, 1967], [Wardlaw, 1990].

The work in [Mathieu, 2006] has brought significant advances allowing realistic simulations of branched plants with GreenLab. Trees are considered as self-regulating systems with several physiological and developmental processes being influenced by their internal trophic state. It allows reproducing the tree architectural plasticity in response to environmental or ontogenetic changes (e.g. progressive appearance of higher branching orders in branches at different growth stages). The model also generates cyclic patterns as an emergent property, similarly to what can be observed on real plants (e.g. rhythmic appearance of fruits) [Mathieu et al., 2008].

Besides GreenLab, many other FSPMs have been developed [Allen et al., 2005], [Evers et al., 2007], [Wernecke et al., 2007]. For instance, in [Bertheloot et al., 2011a],

the author describes a functional-structural model of nitrogen economy for wheat after flowering, NEMA, which links nitrogen fluxes to physiological activities. Inputs of nitrogen fertilizers are fundamental to get high-yielding crops and a production of high quality with the required protein content. This required a proper understanding of root N uptake regulation and of N determinism on yield and production. Complex interactions exist between root N uptake, N remobilization to grains, and photosynthesis, the regulatory mechanisms of which remain far from clear. In NEMA, the Nitrogen content of each photosynthetic organ and its remobilization follow RubisCO turnover, which depends on intercepted light and a mobile nitrogen pool. This pool is enriched by root uptake and nitrogen release from vegetative organs, and is depleted by grain uptake and protein synthesis in vegetative organs; it also accounts for the negative feedback from circulating nitrogen on root uptake, which is formalized following HATS and LATS activities. Organ Nitrogen content and intercepted light determine dry matter production via photosynthesis, which is distributed between organs according to their respective demands.

## 2.2 Functional-structural plant model design steps

Modelling covers an ever increasing range of disciplines, even in communities not necessarily used to strong quantitative or model-building backgrounds. These trends imply a need for wider awareness of what constitutes good model development practice: the modelling process has to be conducted in an orderly fashion. Good modelling practice involves different steps in model development. Descriptions of Good Modelling Practice (GMP) were articulated for example in the Good Modelling Practice Handbook [Van Waveren et al., 1999], which developed a checklist for deterministic, numerical models. It was applied for models in water management [Scholten et al., 2001], [Refsgaard and Henriksen, 2004]. [Blocken and Gualtieri, 2006] outlined ten steps underpinning best practice model development to support natural resource management. All these guidelines for model development can be generalized for a universal modelling sketch.

As far as FSPMs are concerned, in [Vos et al., 2007], the author provided a simplified version of good modelling practice for FSPMs and specified the steps specially for FSPMs. These steps include the conceptual modelling, data collection, model implementation, model verification and evaluation, sensitivity analysis and scenario studies.

In the following, we recall the different modelling steps discussed in [Vos et al., 2007], with a particular stress on the specific points of FSPMs and sensitivity analysis on which this thesis focuses. A clear view about the model design scheme is important for us to locate the work in this thesis in the model design procedure, in order to know how sensitivity analysis can be implemented in this procedure and how it relates to the other steps of modelling.

In practice a complete model development procedure often includes a cyclic series of activities, roughly three types: development of concepts, experimentation field work and modelling.

### 1. The conceptual model

- Definition of the the modelling purpose. The application fields of FSPMs are currently mostly restricted to research and teaching. In research environments, modelling commonly serves purposes such as integration of knowledge or the quantitative test of hypotheses. For example, a developer or scientific user may put a stress on the ability of the model to show what processes dominate a system behaviour, but the model user is more likely to be concerned with the prediction of the function of the model.
- Specification of the modelling context: scope and resources. This step answers the questions about ‘what has to be included’. For FSPMs, at this stage important decisions have to be made on which aspects of function and structure the model needs to explain. In other words, what processes need to be described in a mechanistic way? The identification of a system structure, state variables and inputs should be decided in this step.
- Definition of equations for the system. This step answers the questions about ‘how to describe’ the scope defined in the previous step. An explanation of the desired functions as emergent from the behaviour of the relevant components should be provided. Prior science-based theoretical knowledge is needed to find the model structure described by the relations between the variable in the model.

In FSPM the conceptual model includes specifically:

- Recognition of the important components of which a plant consists. Information on plant composition and topology and qualitative information on their changes over time should be clear. For each plant species of interest such concepts are the basis of architectural modelling.
- A choice of the basic unit of plant modelling.
- The physiological functions to be included in the model, for instance: photosynthesis, respiration, carbon allocation and sink-source interactions transport of water, nutrients and signals in the plant structure.
- The assumed relationship between environmental variables (e.g. temperature) and plant development (progression in phenological stages) or growth processes.
- The time step of relevance to the purpose of the model.
- The construction of a diagram, showing all the important components of the modelled system, their interrelations, the flows of material and the flows of information, the external driving forces and the processes they affect.

2. **Mathematical analysis.** Once data requirements are satisfied, the modeller can give values to inputs, parameters and run model simulations by model implementation. Once the model is technically run, the work is to make diagnostic checking by mathematical analysis to answer the question ‘is the model built right?’. The aspects that need to be checked are: the model behaviour, the limits and the model stability, etc. Sensitivity analysis can be applied here to check if the model has the problem of ‘over parameterization’ by identifying the most ‘non-influential’ parameters. SA can also check whether the variance of the model output is out of range, by fixing which group of the minimum number of parameters can achieve the best variance reduction of the output. In section.3.4, the two objectives of sensitivity analysis are ‘Factor fixing’ and ‘Variance cutting’.
3. **Experimentation, collection and analysis of data.** The deliverables of the conceptual modelling phase include at least a list of parameters and external inputs that are needed to construct a functional-structural model. Protocols need to be made specifying how unknown parameters will be measured in experiments, or from which data they can be estimated for hidden parameters (or those who are too difficult to measure). Theoretically, this step is mathematically complex. We need to study the model identifiability: from a given set of experimental data, is it possible to estimate the set of unknown parameters? It is possible to answer this question by using preliminary empirical virtual data from a given set of parameters and by trying to retrieve the parameters by an estimation method.
4. **Parametric identification.** After the structure and parameter list of the model is decided, we need to do parameter estimation from the experimental data.
  - **Choice of estimation performance criteria and technique:** The parameter estimation criteria (hardly ever a single criterion) reflect the desired properties of the estimates. Generally, the criteria of parameter estimation should be: computationally as simple as possible, robust, efficient, numerically well-conditioned with good statistical properties.
5. **Model validation, scenario studies.** Model validation seeks to answer the questions ‘does the model built achieve the modelling objective?’ This step corresponds to the qualitative validation. The answer to this question is commonly obtained by comparing model results with data from the real system. Such tests of the model performance will gain in value if independent data are included from conditions that differ from the ones for which the model was derived, for example from a different agro-ecological zone. Validation of a model under a wide range of conditions using independent data sets is perhaps not practiced as widely as desirable mostly due to the difficulty of getting the proper experimental data sets, while it is of utmost importance if we want to have reliable models.

6. Model evaluation and model selection. If we consider the previous step as a qualitative validation to check if the model achieves its objective, then this step concentrates on the quantitative validation of the objective. Uncertainty analysis provides an evaluation of the model robustness and test of the predictive capacity. Besides, to compute some information criteria to compare models for model selection is also necessary. Model selection assures that our final model is the most optimized one corresponding to the initial modelling objective.

In this modelling scheme, we put sensitivity analysis before parameterization. Actually, according to different sensitivity analysis aims, it can be performed after parameterization, also can be performed again in model validation for model diagnosis. Generally, sensitivity analysis plays a very important role in model development process. We will specify its roles in Chapter.3.

## 2.3 Limitations and challenges of FSPMs

The application fields of FSPMs are currently restricted to research and teaching [Le Roux et al., 2001]. They also have an important role for the other disciplines with a framework: biomechanical studies of light interception, of root growth, of plant-environment interactions in heterogeneous environments [Sievänen et al., 2000a].

Due to the detailed description of the plant structure in FSPMs generally at organ level, and sometimes, of the local environment of each organ, the FSPMs tend to require a large number of parameters and input data. Owing to the large amount of information they contain about the plant and the number of process they aim at describing, they also tend to be computationally heavy. Moreover, the complicated and interacting biophysical processes governing plant growth bring a large amount of uncertainty into FSPMs: field surveys for collecting data necessary for the development of models are generally difficult and expensive, though there are now techniques available for automated/semi-automated data acquisition like digitizing or high-throughput phenotyping. In consequence, input data (environmental factors) and experimental data from which model parameters are estimated are also characterized by great uncertainty. Finally, modelling complex processes, model parameters estimation and input data collection all contribute to model uncertainty [Monod et al., 2006], [Wallach et al., 2002]. Good modelling practice requires that the modeler provides an evaluation of the confidence in the model. Uncertainty analysis (UA) and Sensitivity analyses (SA) offer valid tools for this evaluation. Actually, sensitivity analysis provides a possible resolution for most challenges we are facing in functional-structural modelling, see section.3.4.

## 2.4 Introduction to several FSPMs

In our work of sensitivity analysis for FSPMs in this thesis, we mainly applied our analysis to 3 FSPMs with different levels of complexity, and infer in each case what information can be drawn from this analysis. The first model is a simple source-sink model of maize growth, GreenLab (description and parameterization can be found in [Ma et al., 2008]). It is used to specifically study the process of carbon (C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations. The second model is the GreenLab model of tree growth (GL3, applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis [Mathieu et al., 2008], which describes tree structural plasticity in response to trophic competition. Lastly, we consider a functional-structural model, NEMA [Bertheloot et al., 2011a], describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering. This model has the specificity to integrate physiological processes governing N economy within plants: root N uptake is modeled with high affinity transport systems (HATS) and low affinity transport systems (LATS), and N is distributed between plant organs according to the turnover of the proteins associated to the photosynthetic apparatus. C assimilation is predicted from the N content of each photosynthetic organ. Consequently, this model is more mechanistic but also more complex than the two previous ones.

In the following of this chapter, we describe more specifically the models concerned in our work. The description about the general modelling architectures, main model equations, and parameters with the corresponding value range we used for SA are given respectively for each model here. The comprehensive details of the model can be further checked in the references given.

### 2.4.1 GreenLab model for maize

GreenLab is a functional-structural model that simulates plant development, growth and morphological plasticity. The model simulates individual organ production and expansion as a function of the growth cycle (GC). For maize, the growth cycle corresponds to the phyllochron (thermal time in degree days between the appearances of two consecutive leaves on the main stem)[Ma et al., 2008].

Plant morphogenesis depends on biomass production and allocation to expanding organs or competing sinks. Biomass production per plant at growth cycle is simplified according to the following mathematical equation:

$$q(i) = E(i) \cdot \mu \cdot Sp \cdot [1 - \exp(-\frac{\lambda}{Sp}S(i))] \quad (2.1)$$

where  $E(i)$  is an environmental function at growth cycle  $i$  (generally related to the Photosynthetically Active Radiation),  $\mu$  is a conversion efficiency,  $\lambda$  is analogous

to the extinction coefficient of Beer-Lambert's Law,  $S_p$  is a characteristic surface of leaves,  $S(i)$  is the photosynthesis leaf surface area at GC  $i$ . Here  $\lambda$  is set to 0.7.

Organs receive an incremental allocation of biomass that is proportional to their relative sink strengths. The relative sink strength for each type of organ is defined as a function of its age in terms of GCs:

$$p_o(j) = P_o f_o(j) \quad (2.2)$$

where  $o$  denotes organ type ( $b$ : leaf blade;  $s$ : sheath;  $e$ : internode;  $f$ : cob;  $m$ : tassel).  $P_o$  is the sink strength associated to organ type  $o$ . For leaf blade,  $P_b = 1$  is set as a normalized reference. The relative sink strength for the first six short internodes is  $K_e P_e$ , with  $K_e$  an empirical coefficient.  $f_o(j)$  is an organ type-specific function of sink variation. A normalization constraint

$$\sum_{j=0}^{T_o-1} f_o(j) = 1 \quad (2.2A)$$

is set, with  $T_o$  being the maximum expansion duration for organ  $o$ .

In the course of organ development, its relative sink strength is assumed to vary according to a beta function  $f_o$  given by:

$$\begin{aligned} f_o(j) &= \begin{cases} g_o(j)/M_o & (0 \leq j \leq T_o - 1) \\ 0 & (j \geq T_o) \end{cases} \\ g_o(j) &= (j + 0.5)^{\alpha_o} (T_o - j - 0.5)^{\beta_o} \\ M_o &= \sum_{j=0}^{T_o-1} g_o(j) \end{aligned} \quad (2.3)$$

The parameters  $\alpha_o$  and  $\beta_o$  vary with organ type. This function is flexible to describe the shape of the sink variation and parameters can be estimated by inverse methods.

At a given GC  $i$ ,  $d(i)$  is computed as the sum of the demands of all expanding organs [Guo et al., 2006]:

$$\begin{aligned} d(i) &= \sum_{o=b,s,e} P_o \cdot \sum_{j=\max(0, i-T_{ext}+1)}^{\min(i, T_o-1)} f_o(j) \\ &+ P_f \cdot f_f(i - 15) + P_m \cdot f_m(i - 21) \end{aligned} \quad (2.4)$$

The biomass allocated to the compartment of organs of type  $o$  ( $o = b, s, e$ ) at  $GCi$  is denoted by  $q_o(i)$  and given by:

$$q_o(i) = \sum_{k=\max(0, i-T_{ext}+1)}^{\min(T_o-1, i)} p_o(k) \frac{q(i)}{d(i)} \quad (2.5)$$

where  $T_{ext}$  stands for the GC at which organogenesis ceases for the whole plant. Here  $T_{ext} = 21$ . Note that for the cob, the demand for biomass  $p_f(k)$  is 0 before  $GC$  15, and tassel only expands during two GCs: the 21<sup>st</sup> and 22<sup>nd</sup>. The plant life span is 33  $GC$ s.

Another variable of interest is the accumulated biomass value for each kind of organ. Let  $Q_o(i)$  denote the total mass of organs of type  $o$  at  $GC$   $i$ :

$$Q_o(i) = \sum_{n=0}^i q_o(n) \quad (2.6)$$

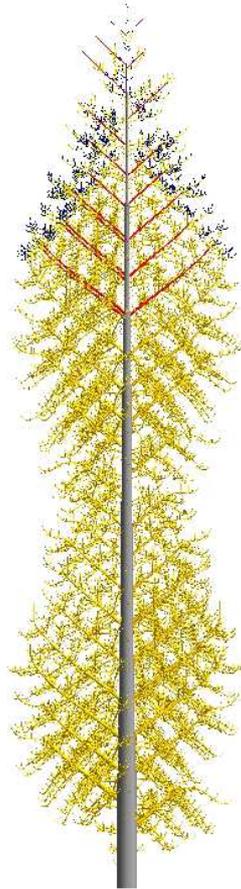
We assume each uncertain input parameter has a uniform distribution, and we use the data from [Guo et al., 2006; Ma et al., 2008, 2007] to set the mean value and variance of all the parameters as listed in Table.2.1.

**Tab. 2.1** : GreenLab Maize: Distribution characteristics for uncertain parameters

Symbol	Definition	Mean value	Uniform distribution
$P_s$	Sink sheath	0.93	[0.837, 1.023]
$P_e$	Sink internode	1.63	[1.467, 1.793]
$P_f$	Sink cob	162.18	[145.962, 178.398]
$P_m$	Sink tassel	1.33	[1.197, 1.463]
$\alpha_b$	Beta coefficient	2.65	[2.385, 2.915]
$\beta_b$	Beta coefficient	2.35	[2.115, 2.585]
$\alpha_s$	Beta coefficient	3.5	[3.15, 3.85]
$\beta_s$	Beta coefficient	1.5	[1.35, 1.65]
$\alpha_e$	Beta coefficient	4.4	[3.96, 4.84]
$\beta_e$	Beta coefficient	0.6	[0.54, 0.66]
$\alpha_f$	Beta coefficient	2.9	[2.61, 3.19]
$\beta_f$	Beta coefficient	2.1	[1.89, 2.31]
$Sp$	Empirical coefficient	0.23	[0.207, 0.253]
$\mu$	Light conversion efficiency	0.0046	[0.00414, 0.00506]

## 2.4.2 GreenLab model for poplar tree

As we mentioned before, GreenLab is a functional-structural plant model that simulates plant development, growth and morphology [Cournède et al., 2006]. The plant growth is discretized with a time step corresponding to the architectural growth cycle, that is to say the time necessary to set in place new growth units (e.g. one year for a tree in a temperate climate). Regarding plant architecture, trees are decomposed into



**Fig. 2.1** : Plant topology after 30 time steps for poplar tree. The yellow organs stand for dead organs. Each physiological age is represented by a different color, respectively grey, red, yellow and purple for physiological ages 1, 2, 3 and 4

elementary units called phytomers that are gathered into categories according to morphological properties [Barthélémy and Caraglio, 2007]. These categories are indexed by a variable called physiological age, from 1 for the trunk to  $P_m$  for the small twigs.  $P_m$  is taken as 4 in our test case, which is generally sufficient to describe complex trees. A time step starts with the appearance of the new organs on the plant, and the tree architecture remains constant during the whole time step.

Interactions between plant organogenesis and functional mechanisms have been implemented by linking the number of new organs to the ratio of available biomass to plant demand. The main equations of the model are given in the following of the section but a complete description can be read in [Mathieu et al., 2009]. Note however that we restrict the study to a theoretical tree (with strong similarities in terms of behaviour with the poplar tree) and simplify the eco-physiological sub-models for the sake of clarity.

For the sensitivity analysis, the variable of interest is the amount of biomass produced at each time step denoted by  $n$ . For this study, it is computed with a simple empirical equation adapted from the Beer-Lambert law:

$$Q(n) = E_{PAR}(n)\mu S_p(1 - \exp^{-\kappa \frac{S_f(n)}{S_p}}) \quad (2.7)$$

$E_{PAR}(n)$  denotes the amount of photosynthetical active radiation received by the plant during the whole time step  $n$ ,  $\mu = 0.33$  is an efficiency simulating the conversion of light energy into biomass,  $S_p$  is an empirical coefficient linked to the projected ground surface,  $\kappa$  denotes the extinction coefficient of the Beer-Lambert law.  $S_f(n)$  is the whole leaf surface area of the tree. It is the sum of each individual leaf surface area. Leaves are assumed to have a constant leaf mass per area  $e = 0.03g.cm^{-2}$ , allowing to deduce their surfaces from their masses.

The amount of biomass  $Q(n)$  will be used for secondary growth and growth of new growth units and organs at the next time step. The allocation model is a proportional one, which means that the biomass allocated to an organ is proportional to its sink strength divided by the plant demand, that is the sum of the organ sink strengths. The biomass used for the secondary growth is proportional to the number of leaves, in accordance with the pipe model theory [Shinozaki et al., 1964].

The number of functional buds, *i.e.* the ones that will give birth to new growth units, and the number of phytomers on each branch depends on the ratio of available biomass to demand, see [Mathieu et al., 2009] for details. This key variable is correlated to the amount of biomass allocated to new organs denoted by  $Q_B(n)$ . Finally, the tree leaf surface area is given by the equation:

$$S_f(n) = \frac{1}{e} \sum_{k=1}^{P_m} p_b^k \frac{S_B^k Q_B(n)}{d_B^k \cdot D_{fb}(n)} \quad (2.8)$$

$p_b^k$  and  $S_B^k$  denote respectively the sink strength of buds and leaves of physiological age  $k$ ,  $d_B^k$  denote the demand of the corresponding growth unit and  $D_{fb}(n)$  is the demand of these buds. For this case study,  $S_B = 1$ ,  $S_I = 1$  and  $S_L = 0.1$  denote respectively the sink strengths for leaves, internodes and layers. They have the same values for each physiological age.

The plant behaviour is the result of the interactions between organogenesis and photosynthesis. When there is enough available biomass, the ratio of biomass to demand is high and a lot of organs will appear in the tree. As the demand is proportional to the number of organs, the increase in the number of leaves induces an increase in plant demand and consequently a decrease in the ratio of biomass to demand. The number of new organs will decrease, inducing a low demand and a high ratio of biomass to demand, and so on. For some combination of the parameters, rhythms may appear in plant biomass production and topology [Mathieu et al., 2008]. We choose such a case study for this thesis. Fig.2.1 shows the plant topology after 30 years.

**Tab. 2.2 :** GreenLab poplar tree: Distribution characteristics for uncertain parameters

Symbol	Definition	Mean value	Uniform distribution
$(S_B)0$	Sink blade PA=1	1	[0.9, 1.1]
$(S_B)1$	Sink blade PA=2	1	[0.9, 1.1]
$(S_B)2$	Sink blade PA=3	1	[0.9, 1.1]
$(S_B)3$	Sink blade PA=4	1	[0.9, 1.1]
$(S_I)0$	Sink Internode PA=1	1	[0.9, 1.1]
$(S_I)1$	Sink Internode PA=2	1	[0.9, 1.1]
$(S_I)2$	Sink Internode PA=3	1	[0.9, 1.1]
$(S_I)3$	Sink Internode PA=4	1	[0.9, 1.1]
$(S_L)0$	Sink Layer PA=1	0.1	[0.09, 0.11]
$(S_L)1$	Sink Layer PA=2	0.1	[0.09, 0.11]
$(S_L)2$	Sink Layer PA=3	0.1	[0.09, 0.11]
$(S_L)3$	Sink Layer PA=4	0.1	[0.09, 0.11]
$\mu$	Resistance blade	3	[2.7, 3.3]
$Sp$	Empirical coefficient	1.7	[1.53, 1.87]

### 2.4.3 NEMA model for wheat

Optimizing fertilizers' use in agriculture requires being able to simulate nitrogen use by plants. Most crop models rely on the idea that a target nitrogen concentration in tissues can be defined and that plants behave depending on the difference between the target and the actual nitrogen concentrations. However, such a target has never been precisely defined. Inputs of nitrogen fertilizers are fundamental to get high-yielding crops and a production of high quality with the required protein content. This required a proper understanding of root N uptake regulation and of N determinism on yield and production. Complex interactions exist between root N uptake, N remobilization to grains, and photosynthesis, whose regulatory mechanisms remain far from clear.

NEMA, a functional-structural model of N economy within individual plants [Bertheloot et al., 2011a], is developed for wheat after flowering to simulate Nitrogen content

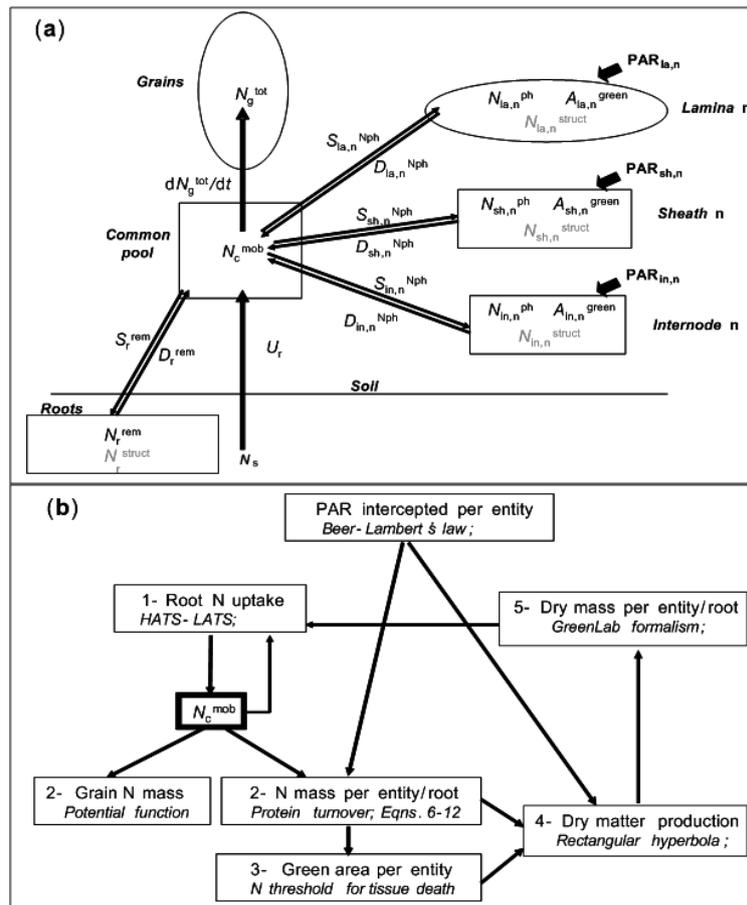
of each photosynthetic organ and its remobilization following RubisCO turnover. The turnover depends on light intercepted and a mobile nitrogen pool, which is enriched by root uptake and nitrogen release from vegetative organs and depleted by grain uptake and protein synthesis in vegetative organs. It also accounts for the negative feedback of circulating nitrogen on root uptake, which is formalized following High Affinity Transport System (HATS) and Low Affinity Transport System (LATS) activities. Organ nitrogen content and light intercepted determine dry matter production by photosynthesis, which is distributed between organs according to their respective demand. This model has the specificity to integrate physiological processes governing N economy within plants: root N uptake is modeled following the HATS and LATS, and N is distributed between plant organs according to the turnover of the proteins associated to the photosynthetic apparatus. C assimilation is predicted from the N content of each photosynthetic organ.

The NEMA scheme is presented in fig.2.2 and all the parameters involved in the parameters sensitivity analysis are listed in tab.2.3.

- Organogenesis ADEL-wheat [Fournier et al., 2003] [Bertheloot et al., 2011a].
- Distribution of nitrogen (N) : photosynthetic N, mobile N, structural N [Triboi and Triboi-Blondel, 2002], [Hafsi et al., 2000], [Bertheloot et al., 2008].
- Distribution of carbon (C) :(interaction with structural N, photosynthetic N) ->GreenLab [Kang et al., 2008].
- Root Absorption [Drouet and Pagès, 2007].
- Photosynthesis computed from incident radiations, surfaces of each leaf and its content in photosynthetic N [Evers et al., 2010].

## 2.5 Concluding remarks

- FSPMs aim to describe: Plant structural development (organogenesis and geometry), functional growth (biomass accumulation and allocation), and the complex interactions between both. The complexity of the underlying biological processes, especially the interactions between functioning and structure [Vos et al., 2010], usually brings a lot of difficulties for experiment and modelling.
- Modelling usually follows principal steps. Generally, it includes development of concepts, experimentation and modelling.
- Sensitivity analysis plays important role to tackle the challenges FSPMs faces, like parameterization and model verification, and it may inspire new experimentation or adjustment of the model structure.



**Fig. 2.2 :** NEMA: Overview of the model of N economy within wheat culms after flowering (*Triticum aestivum* L.)

**Tab. 2.3 :** NEMA: Model parameters: their symbols, definitions, and units. Organs included: grain(g), root(r), and 5 entities(tp): lamina, sheath, internode, peduncle, chaff.

Symbol	Definition	Unit
<i>PAR interception(fixed value)</i>		
$\theta_{La}$	Angle between the vertical and the vector normal to the lamina plan	radians
$k_{vertical}$	PAR extinction coefficient for vertical entities	$m^2m^{-2}$
<i>Root N uptake (5 parameters)</i>		
$\beta_C, \beta_N$	Coefficient for C and N availability effect on root N uptake	dimensionless
$k_{r,1}$	Constant of the Michaelis-Menten function reflecting HATS activity	$gm^{-3}$
$k_{r,2}$	Rate constant of the linear function reflecting LATS activity	$gm^{-3} \circ Cd^{-1}$
$(S_r^{Mrem})_{min}$	Minimum threshold of dry matter influx into roots to sustain root N uptake	$gday^{-1}$
$U_{r,max}$	Theoretical maximum root N uptake at saturating soil N concentration	$gm^{-3} \circ Cd^{-1}$
<i>N fluxes (28 parameters)</i> 1 for grain; 2 for root; 5 for each entity, 25 total for entities		
$\delta_r^N, \delta_{tp}^N$	Relative degradation rates of remobilizable N for roots, entities tp	$\circ Cd^{-1}$
$\gamma$	Relative rate of potential grain N filling during cell division	$\circ Cd^{-1}$
$k_{tp,1}, k_{tp,2}$	Michaelis-Menten constants defining photosynthetic N synthesis associated to xylem influx for entities tp	$gg^{-1}, Jm^{-2}d^{-1}$
$p_r, p_{tp,i}$	Proportion coefficient for N influx following dry mass influx into roots, entities tp	dimensionless
$\sigma_{tp,i}^{Nph}$	Relative rate of photosynthetic N synthesis associated to xylem influx for entities tp	$gg^{-1} \circ Cd^{-1}$
<i>Tissue death and photosynthesis (15 parameters)</i> 3 for each entity, 15 total for entities		
$d_{tp}$	Proportion of maximum specific N mass at which tissues die for entities tp	dimensionless
$\omega_{tp}$	Proportion coefficient linking photosynthesis at saturating PAR and N mass per unit photosynthetic area	$d^{-1}$
$\epsilon_{tp}$	Photosynthetic efficiency	$gJ^{-1}$
<i>Dry matter fluxes (34 parameters)</i> 4 for grain; 5 for root; 5 for each entity, 25 total for entities		
$\alpha_g, \beta_g, \alpha_r, \beta_r, \alpha_{tp}, \beta_{tp}$	Two parameters determining the shape of the Beta function for grains, roots, entities tp.	dimensionless
$\delta_r^M, \delta_{tp}^M$	Relative degradation rates of remobilizable dry mass for roots, entities tp	$\circ Cd^{-1}$
$\sigma_g^M, \sigma_r^M, \sigma_{tp,i}^M$	Relative sink strength of grains, roots, entities tp	dimensionless
$tt_g^{Macc}, tt_r^{Macc}, tt_{tp}^{Macc}$	Duration during which grains, roots, entities tp can accumulate dry mass	$\circ Cd$



## 3. SENSITIVITY ANALYSIS FOR FSPM

In this chapter we introduce the basic concepts, the methods and basic steps of sensitivity analysis. Definitions and equations of all algorithms of the SA applied in our simulation are also given. Since uncertainty analysis (UA) and sensitivity analysis (SA) always come in tandem, we first introduce the general concept of UA and SA in section.3.1. The form of universal model definition is given in section.3.2, as one important part of SA, issues about input factors are illustrated in section.3.3. To mark the importance for the aims of SA, section.3.4 explicits this topic and specifies this issue in our work for FSPM, then we introduce the mathematical description of the main methods we used in our work: SRC and Sobol's, especially several meaningful Sobol's indices in section.3.5. Lastly, the basic steps of SA for FSPM is give in section.3.6.

### 3.1 Uncertainty analysis and sensitivity analysis

Most mathematical problems met in social, economic or natural sciences entail the use of mathematical models. A mathematical model is defined by a series of equations, input variables, parameters, and state variables to characterize the process being investigated. The input is subject to many sources of uncertainty including errors of measurement, absence of information and poor or partial understanding of the driving forces and mechanisms. This uncertainty imposes a limit on our confidence in the response or model output, which is generally too complex for an easy appreciation of the relationship between input factors and output. The understanding of how the model behaves in response to changes in its inputs, is of fundamental importance to ensure a correct use of the models.

Good modelling practice requires that the modeler provides an evaluation of the confidence in the model, possibly assessing the uncertainties associated with the modelling process and with the outcome of the model itself. Uncertainty and sensitivity analysis offer valid tools for characterizing the uncertainty associated with a model. The goal of uncertainty analysis is to answer the question 'what is the uncertainty in  $y(x)$  given the uncertainty in  $x$ ?' and the goal of sensitivity analysis is to answer the question 'how important are the individual elements of  $x$  with respect to the uncertainty in  $y(x)$ ?' [Saltelli et al., 2006].

A possible definition of sensitivity analysis is the following: 'The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned

to different sources of uncertainty in the model input'[Saltelli et al., 2004]. Such information given by SA may provide some help for model assessment:

- Measurement of model adequacy (e.g. does the model fit observation?)
- Knowledge of model relevance (e.g. is the model-based inference robust?)
- Identification of critical regions in the input space (e.g. which combination of factors corresponds to the highest risk?)
- Detection of interactions between factors
- Priorities for research and experimentations
- Simplification of model structure [Saltelli et al., 2004].

Thus the definition of sensitivity analysis involves models, model input and model output. The following sections will present how these parts relate to the nature and purpose of the model, as well as to the set-up of the uncertainty and sensitivity analysis.

### 3.2 Model definition

Before starting sensitivity analysis, we give a definition of what is a 'model' in our practice. A model is represented by a mapping  $f$  (a deterministic or stochastic function) which relates the inputs domain to the output space:

$$Y = f(X_1, X_2, \dots, X_k) \quad (3.1)$$

The input factors  $(X_1, X_2, \dots, X_k)$  are supposed to be random variables described by identified probability distributions which reflect the uncertain knowledge of the system under analysis,  $k$  is the number of factors involved, so we define the  $k$  dimensional parameter vector  $\mathbf{X}$  as one point in the space. Note that the model output  $Y$  under sensitivity analysis is always a scalar value. In the case of dynamic model output, the output is a vector for one state variable. We shall consider each scalar output value of the vector in turn. As mentioned in Chapter.2, most of the FSPMs are dynamic models, the description of the state variables of plant growth evolving by time is the most important way of model presentation. So SA usually are performed separately at each calculation time step for FSPMs.

Consider a dynamic model represented by the following mathematical equation:

$$Y(t) = f(\mathbf{Z}, \theta, t) \quad (3.2)$$

where  $\mathbf{Z}$  is the vector of input variables of the model,  $\theta$  is the vector of uncertain parameters and  $Y(t)$  is the model output at time  $t$ , for  $t \in 1, 2, \dots, T$ , the function

$f(\bullet)$  is the phenomenon studied and it is either deterministic or stochastic. The stochastic aspect of the model will not be discussed in this thesis. However, it is possible to take into account when making repetitions [Ginot et al., 2006], [Lurette et al., 2009].

Consider only the input variables and / or uncertain parameters of the model which we call factors, we note that  $\mathbf{X}$ , the model equation is written:

$$Y(t) = f(\mathbf{X}, t) \quad (3.3)$$

Eqn.3.3 corresponds to all our models of interest in this thesis. For FSPMs, simulations are usually computed at a daily time step and the sequential implementation of sensitivity analysis at each simulation date with one index per parameter per simulation date. On one hand, it is one of the points we are interested in: the SA indices evolving by time for factors can imply the stage evolution relating to plant growth activities. On the other hand, when SA is performed separately at each time step for FSPMs, it can result in several hundreds of sensitivity indices. It is not easy to identify the most important parameters based on such a large number of values. Moreover, this technique has the disadvantage of introducing a high level of redundancy because of the strong correlations between responses from one time step to the next. Likewise, intuitively, to screen one parameter, the requirement should be that the sensitivity response curve is always near zero at all time steps. However, we rarely get this kind of curve from our analysis. Therefore we devised a methodology that can deal with the screening of parameters in this dynamic context, which will be discussed in Chapter.5.

For an easier and simpler way of presentation, next in this chapter, for all the mathematic description about sensitivity analysis algorithm, we still stick to the concept that the output is a scalar as in eqn.3.1.

### 3.3 Model inputs

As we mentioned, we call the input variables and/or uncertain parameters as the model input factors for SA. What constitutes an input for the analysis depends upon how the analysis is set up. The inputs are those factors that are allowed to vary in order to study their effect on the output. An obvious consequence is that the modeller will remain ignorant of the importance of those factors which have been kept out of the scope. This is of course a hazard for the modeller, as a factor deemed non-influential and kept fixed could haunt the results of the analysis at a later stage. Therefore, to choose a proper input factors group should be dealt with as carefully as possible. For us, we have gone an ‘expert review process’ for choosing the proper input factors group. Two objectives were focused: the factors involved in the analysis should not be too many, or else the analysis will be too complex with heavy computing cost; the factors should neither be insufficient, or else the risk of losing important information will come true.

Assuming that we have observations and parameters which are estimated from data now, we might consider the model ‘true’ and run an uncertainty analysis by propagating the uncertainty in the parameters through the model, all the way to the model output. One way of doing this is through Monte Carlo analysis, in which we generate the samples of input according to the random distribution functions of the input parameters, and derive from the propagations for the model output distribution.

Definition of the distributions  $D_{X_1}, D_{X_2}, \dots, D_{X_k}$  that characterize the epistemic uncertainty in factors  $X_1, X_2, \dots, X_k$  is also an important part of a sampling-based uncertainty and sensitivity analysis. These distributions determine the input uncertainty that affect output  $Y$ .

Most of the time, the distributions of input factors are not easy to get. A possible analysis strategy is to perform an initial exploratory analysis with a rather crude definition for  $D_{X_1}, D_{X_2}, \dots, D_{X_k}$  and use sensitivity analysis to identify the most important analysis inputs. Then, resources can be concentrated on characterizing the uncertainty in these inputs and a second presentation of decision-aiding analysis can be carried out with these improved uncertainty characterizations [Helton et al., 2006b].

### 3.4 Purpose of sensitivity analysis

A few heuristic settings for SA, each corresponding to a specific stage or need of the modelling process, are suggested in [Saltelli and Tarantola, 2002] and [Saltelli et al., 2004]. As for all the scientific experiments, the first step is to be clear about the aim of the analysis which can lead to different practice strategies.

Generally, four aims can be classified for SA: Factor Prioritization (FP), Factor Fixing (FF), Factor Mapping (FM) and Variance Cutting (VC) [Saltelli et al., 2008].

- *Factor Prioritization*: It is the identification of the most important factors. If they are fixed to their true values, it would lead to the greatest reduction of the output variance. This setting can be used for the research prioritization, as it allows to identify those factors which deserve the most investment of research attention in order to control the output variance to the minimum.
- *Factor Fixing*: It is the screening non-influential factors in the model, i.e. identifying those factors that can be fixed at any given value in their domains without significantly reducing the output variance. This setting is useful for model simplification to resolve the problem of over-parameterization, or when the modeller has prior beliefs about the importance of some input factors, as it can help in proving or disproving a given model assumption.
- *Variance Cutting*: The reduction of the output variance to a lower threshold is performed by simultaneously fixing the smallest number of input factors. This

setting could be useful when SA is part of a risk assessment study when a regulatory authority was to find the width of the impact estimate distribution too wide. The variance cutting and factor prioritization settings are very similar to each other, as they both aim at reducing the output variance. However, in the case of factor prioritization the scope is to identify the most influential factors one by one, while in the variance cutting setting the objective is to reduce the output variance down to a pre-established level by fixing the smallest subset of factors at once [Saltelli et al., 2004].

- *Factor Mapping*: Finally, the aim of SA can be to study which values of the input factors lead to model realizations in a given range of the output space, e.g. above or below an assigned threshold. For example, the analyst wishes to divide the realizations of the Monte Carlo simulation into two groups, e.g. by categorizing them as acceptable or non-acceptable. This setting can be carried out using the Smirnov test and the approach is known as Regionalized Sensitivity Analysis [Hornberger and Spear, 1981].

FSPMs aim to describe plant structural development (organogenesis and geometry), functional growth (biomass accumulation and allocation) and the complex interactions between both. sensitivity analysis provides a possible resolution for most challenges we are facing in functional-structural modelling:

- Knowledge Integration. Main integration is about structure and function and the different time scales into one consistent modelling framework. For knowledge integration, SA can help for identifying important phases, key variables, important processes regarding some outputs, interaction between processes, between parameters. It is corresponding to the Factor Prioritization: by identifying the most important factors and interactions between them for the output under analysis, key biophysical processes relating to those factors can be marked; by identifying the most important function module (group of factors) and the interaction between modules, the important interactions between processes can be found.
- Link between models and the real world. It mainly relates to model parameterization from experimental data which is one of the most important steps in model design. This challenge is corresponding to Factor Prioritization and Factor Fixing: SA gives some guiding information about the priority of parameters we need to get from an experiment, the cost of an experiment can be more effectively arranged by adding more frequent measurements and more accurate study to the ones with more contribution to the output variables, and vice versa; to simplify the complexity of parameter estimation, we need to cut down the number of factors for this estimation, and one intuitive idea is to fix the non-influential factors identified by SA.

## 3.5 SA methods

### 3.5.1 Classification of methods

There are a lot of classification standards for sensitivity analysis methods: according to the way how the method works, there are local and global methods [Cacuci, 2003], [Cacuci et al., 2005], [Grievank, 2000]; according to the theoretical background, there are sampling-based methods [Helton et al., 2006a] and emulators-based methods [Oakley and O’Hagan, 2004]; according to the purpose of SA, there are qualitative screening methods like Morris method [Morris, 1991], [Campolongo et al., 2007] [Campolongo and Braddock, 1999] and quantitative ranking methods like Sobol’s method [Sobol, 1993].

The most common classifications of SA methods distinguish between quantitative and qualitative methods and between local and global techniques.

- Qualitative methods are aimed at screening, while quantitative techniques can be designed to give information on the amount of variance explained by each factor. In general, the choice of which method to use is driven by the computing cost, as local and qualitative methods are computationally less expensive [Saltelli et al., 2006].
- In local approaches (known as one-at-a-time, OAT), the effect of a single factor’s variation is estimated while keeping all the others fixed at their average values. Yet they cannot include the effect of the shape of the density functions of the inputs, and they are not model-independent.
- Global approaches estimate the effect on the output of a factor keeping all the others varying. Generally, global approaches use model independent methods while not requiring assumptions of additivity or linearity. As a drawback, they are usually computationally expensive [Cariboni et al., 2007].

Sampling-based approaches to uncertainty and sensitivity analysis are both effective and widely used [Helton et al., 2006a], of which an important category of methods are ‘Variance based’ methods. The basic concept for this kind of method is to decompose the output variance into the contributions imputable to each input factor. The most widely used are the FAST (Fourier Amplitude Sensitivity Test, see [Cukier et al., 1973, 1978; Koda et al., 1979]), and Sobol’s methods, see [Sobol, 1993]. FAST method decomposes the output variance  $V(Y)$  by means of spectral analysis. Sobol’s method is based on the same decomposition of variance, which is achieved by Monte Carlo methods in place of spectral analysis. Since it is based on variance decomposition, the different types of sensitivity indices that it estimates can fulfill different objectives of sensitivity analysis: factor prioritization, factor fixing, variance cutting or factor mapping [Cariboni et al., 2007]. It is a very informative method but potentially computationally expensive [Helton et al., 2006a]. Besides the first-order effects,

Sobol's method also aims at determining the levels of interaction between parameters [Wu and Cournède, 2010]. Thus in our algorithm research and practice work in this thesis, we mostly focused on Sobol's method.

### 3.5.2 Description of methods

The simplest and most intuitive way to obtain a local sensitivity index is to compute derivatives (see [Grievank, 2000; Tomovic and Vukobratovic, 1972; Varma et al., 1999]). The sensitivity of the output  $Y$  to a perturbation of an input factor  $X_i$  is estimated at a given value  $X_i^*$ , as

$$Y'_{X_i} = \left. \frac{\partial Y}{\partial X_i} \right|_{X_i=x_i^*} \quad (3.4)$$

In situations where  $Y$  and  $X_i$  have different range of uncertainties, a more balanced measure can be obtained normalizing the derivatives by the factors' standard deviations:

$$S_{X_i}^\sigma = \left. \frac{\sigma_{X_i}}{\sigma_Y} \frac{\partial Y}{\partial X_i} \right|_{X_i=x_i^*} \quad (3.5)$$

The estimation of these OAT methods can be easily implemented, but they are informative only if the model is linear/quasi-linear or if the range of uncertainty of the input factors is small [Cariboni et al., 2007].

The Standardized Regression Coefficients (SRC) method is based on the linear approximation of the model and Monte Carlo simulations. SRC method can demonstrate the shape of the probability distribution of every factor. One important index produced by SRC is the model coefficient of determination  $R^2$ , which represents the fraction of the output variance explained by the linear regression model itself. A side result of the model coefficient of determination ( $R^2$ ) is that it provides an indicator of the degree of non-linearity of the model. When  $R^2 = 1$ , the system is linear and the SRCs can totally explain the variance of the output affected by each factor. Even when models are moderately non-linear (i.e.  $> 0.9$ ), the SRCs can provide valid qualitative information. When getting small, the SRCs are no longer reliable sensitivity representations. In our practice, we call this model coefficient of determination linearity index, and use it to assess the non-linearity of the model.

The SRCs are got based on model linear regression. When we make linear regression for model  $Y = f(\mathbf{X})$ , the result can be:

$$Y = b_0 + \sum_i^k b_{X_i} X_i \quad (3.6)$$

in which  $k$  is the number of the factors considered in SA,  $X_i$  denotes the factors considered in SA, and  $b_{X_i}$  are the regression indices corresponding to each factor. To

change these dimensioned coefficients values to the standardized ones, the SRC index for  $X_i$  denoted by  $\gamma_{X_i}$  is given by:

$$\gamma_{X_i} = \frac{\sigma_{X_i}}{\sigma_Y} b_{X_i} \quad (3.7)$$

in which  $\sigma_{X_i}$  and  $\sigma_Y$  denote respectively the standard deviation in Monte Carlo simulation for factor  $X_i$  and output  $Y$ . Note that if the factors are independent and the true model is linear, the summation of the square value of  $\gamma_{X_i}$  equals to 1. Then the following equation holds (see [Draper and Smith, 1981]):

$$\sum_i (\gamma_{X_i})^2 = 1 \quad (3.8)$$

$\sum_i (\gamma_{X_i})^2$  can therefore be considered as an index of the model linearity. This linearity of the system is described by the so called ‘model coefficient of determination’, which is computed as:

$$R_Y^2 = \frac{\sum_{d=1}^N (Y^* - \mu_Y)^2}{\sum_{d=1}^N (Y - \mu_Y)^2} \quad (3.9)$$

where  $\mu_Y$  is the mean value of  $Y$  in the Monte Carlo simulation,  $Y^*$  is the fitted value in eqn.(3.6),  $Y$  is the actual value at each run, and  $N$  is the total number of runs.

The basic idea of Sobol’s method (see [Sobol, 1993]) is to decompose the function of eqn.(3.1) into terms of increasing dimensionality:

$$\begin{aligned} f(X_1, \dots, X_k) &= f_0 + \sum_{i=1}^k f_i(X_i) \\ &+ \sum_{1 \leq i < l \leq k} f_{il}(X_i, X_l) \\ &+ \dots + f_{1,2,\dots,k}(X_1, \dots, X_k) \end{aligned} \quad (3.10)$$

If the input factors are mutually independent then there exists a unique decomposition of eqn.(3.10), such that all the summands are mutually orthogonal. The variance of the output variable  $Y$  can thus be decomposed into:

$$V(Y) = \sum_{i=1}^k V_i + \sum_{1 \leq i < l \leq k} V_{il} + \dots + V_{1,2,\dots,k} \quad (3.11)$$

Where  $V_i$ ,  $V_{il}$ ,  $V_{1,2,\dots,k}$  denote the variance of  $f_i$ ,  $f_{il}$ ,  $f_{1,2,\dots,k}$  respectively.

Imagine if we fix factor  $X_i$  at its midpoint  $x_i^*$ , how much would this change the variance of  $Y$ ? We indicate the conditioned variance as:

$$V_{\mathbf{X}_{-i}}(Y|X_i = x_i^*) \quad (3.12)$$

where the variance is taken over parameter space  $\mathbf{X}_{-i}$ : a  $k - 1$  dimensional vector of all factors but  $X_i$ .

Comparing the variance  $V(Y)$  in eqn.3.11, by fixing input factors based on eqn.3.12, we say that the smaller the conditional variance becomes, the more important  $X_i$  is, but it would make much difference where  $X_i$  is fixed. Further more, for non-linear models, fixing a factor might actually increase the variance instead of reducing it, also depending on where it is fixed.

The solution is to average eqn.3.12 over all possible values of  $X_i$ , which can be written:

$$E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i)) \quad (3.13)$$

From algebra theory, eqn.3.13 can be included in the equation as following:

$$E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i)) + V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) = V(Y) \quad (3.14)$$

Where  $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$  is called the *main effect* of  $X_i$  on  $Y$ , and  $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$  the *residual*. Hence a small  $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$  or a large  $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$  will imply that  $X_i$  is an important factor.

In this approach the first-order sensitivity index  $S_i$  for factor  $X_i$  defined in eqn.(3.1) is given by:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)} = \frac{V_i(E_{-i}(Y|X_i))}{V(Y)} \quad (3.15)$$

where  $E$  and  $V$  indicate, respectively, the mean and variance operators,  $i$  indicates all possible values of factor  $X_i$ ,  $-i$  indicates all factors but  $X_i$ . The inner expectation is taken at a generic point in the space of variable  $X_i$ , while the outer variance is over all possible values of this generic point. According to eqn.3.11 and eqn.3.15, we can easily refer that  $S_i$  is a value always between 0 and 1. A high value of  $S_i$  signals a significant factor, but not vice versa. We will discuss this question again after all the main indices are introduced. Usually,  $S_i$  is used for 'Factor Priorization' in SA, the variables with high value of  $S_i$  should have more priority for modelling investment.

Intuitively, the higher order sensitivity indexes  $S_{i_1, \dots, i_s}$  deduced from eqn.3.11 should be:

$$S_{i_1, \dots, i_s} = \frac{V_{i_1, \dots, i_s}}{V(Y)}$$

and we can also easily get the characteristic of all the sensitivity indices including all the orders:

$$1 = \sum_{i=1}^k S_i + \sum_{1 \leq i < l \leq k} S_{il} + \dots + S_{1,2, \dots, k} \quad (3.16)$$

It is interesting to write equations giving directly the high order indices of interest (as in eqn.3.15 for  $S_i$ ).

Take second order for example, given two generic factors  $X_i$  and  $X_j$  ( $i \neq j$ ), the following result holds:

$$V(E(Y|X_i, X_j)) = V_i + V_j + V_{ij} \quad (3.17)$$

in which:

$$V_i = V(E(Y|X_i)) \quad (3.18)$$

$$V_j = V(E(Y|X_j)) \quad (3.19)$$

$$V_{ij} = V(E(Y|X_i, X_j)) - V_i - V_j \quad (3.20)$$

We have dropped the indices of both the  $E$  and  $V$  operators. Indeed we do not need them if we accept the convention that the argument conditioning the inner operator,  $X_i$  and  $X_j$  in this case, is also the set over which we apply the outer operator, i.e. the variance is taken over  $X_i$  and  $X_j$ , and the average  $E$  must be taken over all but  $(X_i, X_j)$ . The term  $V_{ij}$  is the interaction term between factors  $X_i$  and  $X_j$ . So the second order index  $S_{i,j}$  is the index that can describe the portion of the output variance represented by  $V_{ij}$ :

$$S_{ij} = \frac{V(E(Y|X_i, X_j)) - V_i - V_j}{V(Y)} = \frac{V(E(Y|X_i, X_j))}{V(Y)} - S_i - S_j \quad (3.21)$$

Eqn.3.21 gives us a clue that to get the second order index, we must first get the first order index of each factor involved. The higher order indices follow the same idea: to compute any higher order index the requirement of all the lower order indices is necessary. Though in [Saltelli, 2002], a strategy is proposed to make best use of model simulations so that it would not be so computing cost heavy to get the higher order indices. Generally, the higher order indices we want to get, the more model simulations we need, thus the heavier computing cost this brings. In Chapter.5, we will illustrate a strategy of combining different lower order (typically) indices to avoid this kind of high order index computing but still get the decomposition of the portions of variances including them when necessary.

The higher order indices especially the second order ones, provide us with a way to evaluate the interaction between variables, from which we can deduce the biophysical interaction processes related to these variables.

Besides first order and high order indices, the total effect index is also frequently used for ‘factor screening’ [Saltelli et al., 2008] (that is to say to identify the least influential parameters). To introduce it, we first consider another index called  $S_{-i}$  based on the definition in eqn.3.15 for first order index  $S_i$ :

$$S_{-i} = \frac{V(E(Y|\mathbf{X}_{-i}))}{V(Y)} = \frac{V(E(Y|X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_k))}{V(Y)} \quad (3.22)$$

By analogy with the second order indices, eqn.3.22 should include all terms of any order that do not include factor  $X_i$ . As for eqn.3.16, the sum of all possible sensitivity

terms must be 1, the difference  $1 - S_{-i}$  must be made up of all terms of any order that include  $X_i$ , we call it total order index, and denote it by  $ST_i$ :

$$ST_i = 1 - \frac{V(E(Y|\mathbf{X}_{-i}))}{V(Y)} \quad (3.23)$$

Again, according to the algebraic decomposition of  $V(Y)$  in eqn.3.14, we also have:

$$ST_i = \frac{E_{-i}(V_i(Y|\mathbf{X}_{-i}))}{V(Y)} \quad (3.24)$$

Now we go back to the previous remark namely that ‘A high  $S_i$  index signals an important variable, but not vice versa.’ To process ‘Factor Fixing’,  $S_i \approx 0$  is a necessary but insufficient condition for fixing factor  $X_i$ . This factor might be involved in interactions with other factors such that, although its first-order term is zero, there might be a non-zero higher order term. So it is meaningful to have a total order index  $ST_i$  for ‘Factor Fixing’, if  $ST_i = 0$ , then  $X_i$  is non-influent; reversely, if  $X_i$  is non-influent, the value of  $ST_i$  must be zero.  $ST_i = 0$  (practically  $ST_i \approx 0$ ) is a necessary and sufficient condition for  $X_i$  being non-influential.

To sum up, for a given factor  $X_i$ , we know from the value of first-order Sobol’s index  $S_i$  whether a factor is influent at the main effect, while an important difference between  $ST_i$  and  $S_i$  flags an important role of interactions for that factor regarding output  $Y$ . If this is the case, inspection of the second order indices  $S_{ij}$  for all  $i \neq j$  will allow us to identify which factor  $X_i$  interacts with. Thus, in fact, besides the first-order effects, Sobol’s method also aims at determining the levels of interaction between parameters [Wu and Cournède, 2010].

Now we can extend the use of eqn.3.17 to a higher order. And we denote the variance as  $V_{i_1, i_2, \dots, i_s}^c$  in which ‘s’ is the number of factors included in this variance:

$$V_{i_1, i_2, \dots, i_s}^c = V(E(Y|X_{i_1}, X_{i_2}, \dots, X_{i_s})) \quad (3.25)$$

Though we can not directly get the higher Sobol’s index that explains the interactions between parameters as  $S_{i_1, i_2, \dots, i_s}$  from eqn.3.25, still it has its own meaning and it is useful when our objective is to investigate a group of parameters instead of independent parameters.  $V_{i_1, i_2, \dots, i_s}^c$  represents the variance donated by the set and subsets of factors  $\{X_{i_1}, X_{i_2}, \dots, X_{i_s}\}$ : the variance caused by all the factors in the sets and all the interactions between these factors. As such, it evaluates the main effect of the group of factors to the output.

In the case of complex system models that are composed of several sub-models (or modules), we are interested in the sensitivity of these modules to the model output. Define the whole parameter space as a set  $\mathcal{P}$ . Let  $\{\Omega_1, \Omega_2 \dots \Omega_r\}$  be a partition of  $\mathcal{P}$

such that each subset  $\Omega_i = \{X_{i_1}, X_{i_2}, \dots, X_{i_s}\}$ , corresponds to the set of parameters identified as belonging to the  $i$ th module. We denote the first order group index of subset  $\Omega_i$  as  $S_{\Omega_i}^g$ , and  $i \in \{1, 2, \dots, r\}$ , the superscript ‘ $g$ ’ means group index

$$S_{\Omega_i}^g = \frac{V_{i_1, i_2, \dots, i_s}^c}{V(Y)} = \frac{V(E(Y|X_{i_1}, X_{i_2}, \dots, X_{i_s}))}{V(Y)} \quad (3.26)$$

Similarly,  $V_{-i_1, i_2, \dots, i_s}^c$  will indicate the sum of all indices in the complementary set of  $\{i_1, i_2, \dots, i_s\}$ , i.e.  $V_{-i_1, i_2, \dots, i_s}^c = V_{l_1, l_2, \dots, l_{k-s}}^c$ , where  $i_p \neq l_q$  for all  $p \in [1, 2, \dots, s]$ ,  $q \in [1, 2, \dots, k-s]$ . If we define a ‘total effect’ index for subset  $\Omega_i$  as in eqn.3.23 for one factor:

$$ST_{\Omega_i}^g = 1 - \frac{V_{-i_1, i_2, \dots, i_s}^c}{V(Y)} = 1 - \frac{V(E(Y|X_{l_1}, X_{l_2}, \dots, X_{l_{k-s}}))}{V(Y)} \quad (3.27)$$

In the same way, if we define a second order group index for two subsets  $\Omega_i$  and  $\Omega_j$  as in eqn.3.21 for two factors:

$$S_{\Omega_{ij}}^g = \frac{V(E(Y|\Omega_i, \Omega_j))}{V(Y)} - S_{\Omega_i}^g - S_{\Omega_j}^g \quad (3.28)$$

The total effect index  $ST_{\Omega_i}^g$  for subset  $\Omega_i$  means all the effects brought by the factors set  $\{X_{i_1}, X_{i_2}, \dots, X_{i_s}\}$ : in the right side of eqn.3.16, it is the sum of all the terms of any order that include at least one factor from the set, including the interactions of all the subsets of these factors with the other ones that are not in  $\Omega_i$ .

In this sense, the difference  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  means the interactions of the factors set with the ones not belonging to the set. The second order group index  $S_{\Omega_{ij}}^g$  specifically indicates the interactions between two subsets of factors. For a complex model in which subset of factors means the sub-module of the model,  $S_{\Omega_{ij}}^g$  explains the interaction between two modules.

The group indices given here  $S_{\Omega_i}^g$ ,  $ST_{\Omega_i}^g$ ,  $S_{\Omega_{ij}}^g$  and the difference  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  are useful when the object we consider is a set of factors i.e. factors from one module in the whole model. We will introduce a full methodology based on these indices in Chapter.5 for module by module analysis and illustrate this practice in Chapter.7 for the NEMA model.

### 3.6 Basic steps of SA for FSPM

Basically, SA of any model has to proceed in an orderly fashion. In practice the development of SA often proceeds in a loop of activities, because the modelers do not yet have enough knowledge about the attributes related to the decisions in SA,

including the range of parameters with uncertainty, lack of experimental data etc. The steps outlined below are more or less in a logical and chronological order, but there are numerous reasons to deviate from the sequence that is presented.

1. It is the first step to establish the goal of sensitivity analysis and consequently to define the form of the output function that answers the question(s). As for FSPMs in our practice, ‘Factor prioritization’ for modelling decision and ‘Factor fixing’ for model simplification are our two main aims of SA. Most of the time, what we called here ‘the output of the model’ is actually the state variables of the model which play a very important role in the model. In our case, we mostly chose the total biomass production of each growth cycle as the output of interest for the two GreenLab models: maize and poplar tree, for the reason that biomass generation and allocation is the main formalism of GreenLab. What’s more, we also took the ratio of available biomass to total demand for poplar tree, because it is representative of the level of trophic competition inside the plant. Thus it allows simulating as well the ontogenetic changes in plant topology throughout its growth phases (progressive set up of architecture units) as architectural plasticity in response to environmental changes. The interactions between organogenesis and functioning can be detected by looking into this variable with sensitivity analysis. In the case of NEMA model, since it is far more complicated than the two previous models, the strategy we worked out is a little different to the one we used for the first two GreenLab models. We will therefore talk about this strategy called ‘module by module analysis’ in Chapter.5.
2. Next we need to decide which input factors should be included in our analysis, that is to say, to define the parameter space for sensitivity analysis based on the objective issue in the first step. At this level, trigger parameters can be defined, allowing one to sample across model structures, hypotheses, etc. Moreover, in the case of multi-dimensional maps of factors, define the characterizing parameters to represent them in the SA. After we decided the parameter space, it is to choose a distribution function for each of the input factors as mentioned in section.3.3. In our case, the distribution of parameters is a uniform distribution.
3. Afterwards, we need to choose sensitivity methods or to design the strategy if it must be the combination of more than one SA methods. There are basically two classes of SA methods to decide: local methods and global ones. Local methods cannot include the effect of the shape of the density functions of the inputs, and they are not model independent but they are really computationally cheap. Global approaches estimate the effect on the output of a factor keeping all the others varying. Generally, global approaches use model independent methods while not requiring assumptions of additivity or linearity. As a drawback, they are usually computationally expensive [Cariboni et al., 2007]. The Standardized Regression Coefficients (SRCs) can be viewed as an interesting trade-off between the local and global method, regarding the advantages and shortcomings of both: the accuracy of the analysis and the computational cost. The main function of SRCs we can make use of is the linearity index generalized by the determination

coefficient. Alternatively, to fulfill our aim of ‘Factor Prioritization setting’ and ‘Factor Fixing setting’ in our SA for FSPMs, we can choose Sobol’s method to get the index of  $S_i$  and  $ST_i$ , because Sobol’s method can fulfill most the aims of our analysis and can easily get the quantitative parameter interaction evaluation based on the variance decomposition by  $S_{ij}$ .

4. After all these preparations, the following is to start the Monte Carlo simulation sampling of the input factor for the analysis, then to evaluate the model on the generated sample and produce the output, which contains N (sampling number) output values.
5. Lastly, we analyse the model outputs using the estimators provided by the methods we chose before and draw our conclusions based on the analysis result.

### 3.7 Concluding remarks

- Sensitivity analysis is the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input [Saltelli et al., 2004]. SA involves models, model input and model output.
- $Y(t) = f(\mathbf{X}, t)$  corresponds to all our models of interest in this thesis. For FSPMs, simulations are usually computed at a daily time step and the sequential implementation of sensitivity analysis at each simulation date with one index per parameter per simulation date. For an easier and simpler way of presentation, in all the mathematic description about sensitivity analysis algorithm next, we still stick to the concept that the output is a scalar as  $Y = f(X_1, X_2, \dots, X_k)$ ,  $k$  being the number of uncertain factors.
- Most of the time, the distributions of input factors are difficult to obtain. A possible strategy is to perform an initial exploratory analysis with rather crude definition for the distribution of the inputs and use sensitivity analysis to identify the most important inputs; then, resources can be concentrated on characterizing the uncertainty of these inputs and a second sensitivity analysis can be carried out with these improved uncertainty characterizations [Helton et al., 2006b].
- It is important to be clear about the aims at the beginning of SA: Factor Prioritization (FP), Factor Fixing (FF), Factor Mapping (FM) and Variance Cutting (VC). In our practice, FP and FF is the case. Factor fixing is also called ‘factor screening’ in some references, which means after identification of the least influential factors, in the modelling processing that needs us to fix some factors to their nominal value, they could be fixed.
- Determination coefficient  $R^2$  generated by the SRC method, first order index  $S_i$ , second order  $S_{ij}$ , total order index  $ST_i$ , first order group index  $S_{\Omega_i}^g$ , total order index  $ST_{\Omega_i}^g$ , second order group index  $S_{\Omega_{ij}}^g$  given by Sobol’s method are the indices we mainly focused on in our work.

## **Part II**

# **ALGORITHM AND METHODOLOGY DESIGN**



# 4. AN EFFICIENT COMPUTATIONAL METHOD FOR GLOBAL SENSITIVITY ANALYSIS

Sobol's method decomposes the variance of the output of interest into terms due to individual parameters but also to interactions between parameters. We are particularly interested in the later study on interactions between parameters using sobol's method. Such information is crucial for models with potentially high levels of non-linearity and interactions between processes, like plant growth models.

However, the computation of Sobol's indices relies on Monte Carlo sampling and re-sampling, whose costs can be very high when model simulation is also expensive. Especially, for some complicated FSPMs, at organ level, the cost of model simulation can be very heavy [Sievänen et al., 2000a]. Therefore, it is crucial to not only devise efficient computing techniques, in order to make best use of model evaluations [Saltelli, 2002], but also to have a good control of the estimation accuracy with respect to the number of samples.

The objective of this chapter is to study these two aspects. First, we propose a computing method inspired by [Homma and Saltelli, 1996], which slightly improves their use of model evaluations, and then derive an estimator of the error of sensitivity indices evaluation with respect to the sampling size for this generic type of computational methods. Numerical tests are then shown to illustrate the results. The method is applied to the functional-structural models we have introduced in Chapter.2 when Sobol's method is needed. For example, the GreenLab model for tree growth, whose particularity is the strong level of interaction between plant functioning and organogenesis. The simulation result of the new method application to GreenLab tree model will be presented in Chapter.7.

## 4.1 Computation of sensitivity indices

### 4.1.1 General concepts of Sobol's sensitivity analysis

We recall here the basic concepts of Sobol's method [Sobol, 1993] to present the original work about sensitivity analysis indices. The function  $f(\mathbf{X}) \equiv f(X_1, X_2, \dots, X_k)$  under investigation is defined in the  $k - dimensional$  cube  $\mathbf{K}^k$ . If the input factors are

mutually independent then there exists a unique decomposition of  $f(\mathbf{X})$ :

$$\begin{aligned} f(X_1, \dots, X_k) &= f_0 + \sum_{i=1}^k f_i(X_i) \\ &+ \sum_{1 \leq i < l \leq k} f_{il}(X_i, X_l) \\ &+ \dots + f_{1,2,\dots,k}(X_1, \dots, X_k) \end{aligned} \quad (4.1)$$

The basic idea of Sobol's method (see [Sobol, 1993]) is to decompose the function of interest into terms of increasing dimensionality as in eqn.(4.1), such that all the summands are mutually orthogonal. The variance of the output variable  $Y$  can thus be decomposed into:

$$V = \sum_{i=1}^k V_i + \sum_{1 \leq i < l \leq k} V_{il} + \dots + V_{1,2,\dots,k} \quad (4.2)$$

Where  $V_i$ ,  $V_{il}$ ,  $V_{1,2,\dots,k}$  denote the variance of  $f_i$ ,  $f_{il}$ ,  $f_{1,2,\dots,k}$  respectively.

In this approach the first-order sensitivity index for factor  $X_i$  is given by:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)} = \frac{V_i(E_{-i}(Y|X_i))}{V(Y)} \quad (4.3)$$

where  $E$  and  $V$  indicate, respectively, the mean and variance operators,  $i$  indicates all possible values of factor  $X_i$ ,  $-i$  indicates all factors but  $X_i$ . The inner expectation is taken at a generic point in the space of variable  $X_i$ , while the outer variance is over all possible values of this generic point.

The complementary first order index  $S_{-i}$  is given by:

$$S_{-i} = \frac{V(E(Y|\mathbf{X}_{-i}))}{V(Y)} \quad (4.4)$$

The second order sensitivity indexes  $S_{i_1,\dots,i_s}$  are given by:

$$S_{ij} = \frac{V(E(Y|X_i, X_j)) - V_i - V_j}{V(Y)} = \frac{V(E(Y|X_i, X_j))}{V(Y)} - S_i - S_j \quad (4.5)$$

The total order effect  $ST_i$  is instead given by:

$$ST_i = 1 - \frac{V(E(Y|\mathbf{X}_{-i}))}{V(Y)} \quad (4.6)$$

If  $ST_i = 0$ , then  $X_i$  is non-influent, so the index  $ST_i$  is suitable for fixing non-influential factors. Standard Sobol's method was proposed in [Sobol, 1993]. In [Homma and Saltelli, 1996], an improved estimator is presented to compensate the system error, completed in [Saltelli, 2002], and a computationally efficient design is discussed. This method will therefore be called Homma-Saltelli method. We then propose an improvement of this method to promote its convergence characteristics.

### 4.1.2 Sobol's computing method and Homma-Saltelli (H-S) improvement

The standard Sobol's method for SA was put forward in [Sobol, 1993], and the Monte Carlo procedure of Sobol's index was first proposed by [Saltelli et al., 1993]. Let  $\mathbf{U} = (X_{i_1}, X_{i_2}, \dots, X_{i_s})$  and  $\mathbf{V} = (X_{l_1}, X_{l_2}, \dots, X_{l_{k-s}})$  be a partition of the  $k$  random variables corresponding to the input space. If we define:

$$\bar{V}_{\mathbf{u}} = \int \int \int f(\mathbf{u}, \mathbf{v}) f(\mathbf{u}, \mathbf{v}') D_U(u) D_V(v) D_V(v') du dv dv' \quad (4.7)$$

with  $D_U$  the probability distribution of random vector  $\mathbf{U}$  (corresponding to parameter uncertainty) and  $D_V$  the probability distribution of vector  $\mathbf{V}$ . then:

$$V(E(Y|\mathbf{U})) = \bar{V}_{\mathbf{u}} - E^2(Y) \quad (4.8)$$

The same argument, if:

$$\bar{V}_{-\mathbf{u}} = \int \int \int f(\mathbf{u}, \mathbf{v}) f(\mathbf{u}', \mathbf{v}) D_U(u) D_U(u') D_V(v) du du' dv \quad (4.9)$$

then:

$$V(E(Y|-\mathbf{U})) = \bar{V}_{-\mathbf{u}} - E^2(Y) \quad (4.10)$$

We denote  $V(E(Y|-\mathbf{U})) = V(E(Y|\mathbf{V}))$  to note that  $\mathbf{V}$  is the complementary vector of  $\mathbf{U}$ .

We take  $s = 1$  which represents the computing for first order Sobol's index. And as indicated in [Saltelli et al., 2010], 'radial sampling' should be applied to achieve better 'balance design' in the sampling. We describe as following the numerical implementation of Sobol's index computing based on radial sampling. To get the conditional expectation value for model output  $Y$ , we first decide the base sampling dimension  $N$ , then we implement the following steps:

1. Generate a Monte Carlo sampling of dimension  $N$  of the input factors according to their random distributions and form the  $N \times k$  matrix  $U_{N \times k}$  ( $k$  being the dimension of the input space) with each row a set of parameters;  $U_{N \times k}$  is called the 'sampling matrix'

$$U_{N \times k} = \begin{bmatrix} x_{1(1)} & \cdots & x_{i(1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i(2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i(N)} & \cdots & x_{k(N)} \end{bmatrix} \quad (4.11)$$

2. Generate another sampling matrix of dimension  $N \times k$ ,  $W_{N \times k}$ , called the 're-sampling matrix'

$$W_{N \times k} = \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i(N+1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i(N+2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i(2N)} & \cdots & x_{k(2N)} \end{bmatrix} \quad (4.12)$$

3. Define a matrix  $W'_{N \times k}$  formed by all columns of  $W_{N \times k}$ , except the  $i^{th}$  column obtained from the  $i^{th}$  column of  $U_{N \times k}$ ,

$$W'_{N \times k} = \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i(1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i(2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i(N)} & \cdots & x_{k(2N)} \end{bmatrix} \quad (4.13)$$

4. Compute the model output for each set of input parameters from  $U_{N \times k}$  and  $W'_{N \times k}$  (that is to say for each row in  $U_{N \times k}$  and  $W'_{N \times k}$ ), to obtain two column vectors of model outputs of dimension  $N$ :  $\mathbf{y} = f(U_{N \times k})$ ,  $\mathbf{y}'_R = f(W'_{N \times k})$ .
5. The sensitivity indices are hence computed based on scalar products of the above defined vectors of model outputs.

The applicability of the sensitivity estimates  $S_i$  to a large class of functions  $f(\mathbf{X})$  is linked to the possibility of evaluating the multidimensional integral associated with these estimates via Monte Carlo methods. For a given sampling size  $N$  tending to  $\infty$  the following estimate for the mean value of the output is straightforward:

$$\hat{f}_0 = \frac{1}{N} \sum_{j=1}^N \mathbf{y}^{(j)} \quad (4.14)$$

where  $\mathbf{y}^{(j)}$  is the model output for a sample point in the parameter space  $\mathbf{K}^k$ . The hat symbol  $\hat{\cdot}$  will be used to denote estimates.

To list the estimator for the standard Sobol's in [Sobol, 1993], the following notation will be introduced:

$$\bar{V} = \frac{1}{N} \sum_{j=1}^N \mathbf{y}^{(j)2} \quad (4.15)$$

$$\bar{V}_i = \frac{1}{N} \sum_{j=1}^N \mathbf{y}^{(j)} \mathbf{y}'^{(j)}_R \quad (4.16)$$

$$\bar{V}_{-i} = \frac{1}{N} \sum_{j=1}^N \mathbf{y}^{(j)} \mathbf{y}'^{(j)} \quad (4.17)$$

Then we estimate the output variance by:

$$\hat{V} = \bar{V} - \hat{f}_0^2 \quad (4.18)$$

$$V_i(E_{-i}(Y|X_i)) \approx \hat{V}_i = \bar{V}_i - \hat{f}_0^2 \quad (4.19)$$

$$V_{-i}(E_i(Y|X_{-i})) \approx \widehat{V}_{-i} = \overline{V}_{-i} - \widehat{f}_0^2 \quad (4.20)$$

and finally:

$$\widehat{S}_i = \frac{\widehat{V}_i}{\widehat{V}} = \frac{\overline{V}_i - \widehat{f}_0^2}{\widehat{V}} \quad (4.21)$$

$$\widehat{S}_{Ti} = 1 - \frac{\widehat{V}_{-i}}{\widehat{V}} = 1 - \frac{\overline{V}_{-i} - \widehat{f}_0^2}{\widehat{V}} \quad (4.22)$$

As mentioned in [Homma and Saltelli, 1996], to compensate the ‘systematic error’ in standard Sobol’s method, better estimates for the term  $V_i(E_{-i}(Y|X_i))$  is obtained by also computing the output of the ‘re-sampling matrix’  $W_{N \times k}$ , we denote it as  $\mathbf{y}_R = f(W_{N \times k})$ . We define  $\gamma^2$ :

$$\gamma^2 = \frac{1}{N} \sum_{j=1}^N \mathbf{y}^{(j)} \mathbf{y}_R^{(j)} \quad (4.23)$$

then the variance estimator is chosen as:

$$V_i(E_{-i}(Y|X_i)) \approx \widehat{V}_i = \overline{V}_i - \gamma^2 \quad (4.24)$$

For the same reason:

$$\widehat{S}_i = \frac{\widehat{V}_i}{\widehat{V}} = \frac{\overline{V}_i - \gamma^2}{\widehat{V}} \quad (4.25)$$

$$\widehat{S}_{Ti} = 1 - \frac{\widehat{V}_{-i}}{\widehat{V}} = 1 - \frac{\overline{V}_{-i} - \gamma^2}{\widehat{V}} \quad (4.26)$$

In [Saltelli, 2002], a detailed proof of the cheaper computing cost for this method is given, and one conclusion is that by adding the computational cost of  $N$  model runs, we can get the full set of sensitivity indices (of all orders) for half the cost of the standard Sobol’s method [Sobol, 1993]. For the Homma-Saltelli method, to estimate each of the  $k$  first order indices, the computing cost is  $N(k + 2)$  model runs, with  $N$  model runs for  $\mathbf{y}$ ,  $N$  model runs for  $\mathbf{y}_R$  and  $Nk$  model runs for  $\mathbf{y}'_R$ . The same reasoning can be applied for  $\widehat{S}_{Ti}$ , see eqn.4.26. In the next section, we propose to adapt the Homma-Saltelli method in order to slightly improve the use of model evaluations.

### 4.1.3 A new method to compute Sobol’s indices

The attempt for the work is to promote the computing efficiency and convergence of Sobol’s method, and the basic idea is to make the sampling-resampling processing ‘smoother’ by ‘averaging’. In the Homma-Saltelli method, to compute  $E_{-i}(Y|X_i)$ , model outputs are computed for the  $N$  sampling matrix  $U_{N \times k}$ ,  $N$  re-sampling matrix

**Tab. 4.1 :** Sobol's Estimator improvement, H-S: Homma-Saltelli method [Homma and Saltelli, 1996].  $\mathbf{y} = f(U_{N \times k})$ ,  $\mathbf{y}_R = f(W_{N \times k})$ ,  $\mathbf{y}' = f(U'_{N \times k})$ ,  $\mathbf{y}'_R = f(W'_{N \times k})$ .

Variables	H-S	Proposed
$\hat{f}_0$	$\mathbf{y}^{(j)}$	$\frac{1}{2}(\mathbf{y}^{(j)} + \mathbf{y}_R^{(j)})$
$\bar{V}_i$	$\mathbf{y}^{(j)}\mathbf{y}_R^{(j)}$	$\frac{1}{2}(\mathbf{y}^{(j)}\mathbf{y}_R^{(j)} + \mathbf{y}_R^{(j)}\mathbf{y}'^{(j)})$
$\bar{V}_{-i}$	$\mathbf{y}^{(j)}\mathbf{y}'^{(j)}$	$\frac{1}{2}(\mathbf{y}^{(j)}\mathbf{y}'^{(j)} + \mathbf{y}_R^{(j)}\mathbf{y}_R'^{(j)})$
$\bar{V}$	$\mathbf{y}^{(j)2}$	$\frac{1}{2}(\mathbf{y}^{(j)2} + \mathbf{y}_R^{(j)2})$
$\gamma^2$	$\mathbf{y}^{(j)}\mathbf{y}_R^{(j)}$	$\frac{1}{2}(\mathbf{y}^{(j)}\mathbf{y}_R^{(j)} + \mathbf{y}'^{(j)}\mathbf{y}_R'^{(j)})$

$W_{N \times k}$  and for the changed  $i$ th column 're-sampling' matrix ( $W'_{N \times k}$ ). The changed  $i$ th column 'sampling' matrix  $U'_{N \times k}$  can be defined and made full use of:

$$U'_{N \times k} = \begin{bmatrix} x_{1(1)} & \cdots & x_{i(N+1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i(N+2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i(2N)} & \cdots & x_{k(N)} \end{bmatrix} \quad (4.27)$$

We denote  $\mathbf{y}' = f(U'_{N \times k})$ . Alternatively,  $\mathbf{y}'$  can also be used for  $E_{-i}(Y|X_i)$ . Correspondingly, when we average these two ways of computing  $E_{-i}(Y|X_i)$ , then the outer  $V_i(E_{-i}(Y|X_i))$  should be inferior, because by doing the averaging, we will get a more 'balanced' simulation architecture. Thus, we obtain the proposed variant Sobol's estimator as shown in table.4.1.

The computing cost is mentioned as  $N(k+2)$  for the old Sobol's method to compute each of the  $k$  first order indices [Pagano and Ratto, 2007]. For the estimator proposed in this paper, we make full use of the  $2N$  samplings, just with  $Nk$  more model runs, that is to say with the computing cost of  $N(2k+2)$  model runs. In order to fairly compare the efficiency of the two computing methods, we need to ensure the same number of model evaluations. It is thus obtained if the sampling size in H-S method and new method obey  $N_{H-S} = N_{new} * 2(k+1)/(k+2)$ .

Based on this new estimator configuration in table.4.1, we now extend this method to second order index without adding extra model evaluation computing costs.

We rewrite the definition of second order index here:

$$S_{i_1 i_2} = \frac{V(E(Y|X_{i_1}, X_{i_2})) - V_{i_1} - V_{i_2}}{V(Y)} = \frac{V(E(Y|X_{i_1}, X_{i_2}))}{V(Y)} - S_{i_1} - S_{i_2} \quad (4.28)$$

According to eqn.4.7 and eqn.4.8, to compute  $V(E(Y|X_{i_1}, X_{i_2}))$ , we can use the same sampling and re-sampling matrix as in eqn.4.11 and eqn.4.12. Normally, like in eqn.4.13 and eqn.4.27 for the first order index computing, we should compute new

matrices for second order index computing as following:

$$W'_{N \times k, i_1 i_2} = \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i_1(1)} & x_{i_2(1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i_1(2)} & x_{i_2(2)} & \cdots & x_{k(N+2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i_1(N)} & x_{i_2(N)} & \cdots & x_{k(2N)} \end{bmatrix} \quad (4.29)$$

$$U'_{N \times k, i_1 i_2} = \begin{bmatrix} x_{1(1)} & \cdots & x_{i_1(N+1)} & x_{i_2(N+1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i_1(N+2)} & x_{i_2(N+2)} & \cdots & x_{k(2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i_1(2N)} & x_{i_2(2N)} & \cdots & x_{k(N)} \end{bmatrix} \quad (4.30)$$

in which  $i_1$  and  $i_2$  are the identities of the factors involved in the second order index. Considering factor set  $(X_{i_1}, X_{i_2})$  as factor subset  $\mathbf{u}$  with  $s = 2$ , then  $f(W'_{N \times k, i_1 i_2})$  equals to  $f(\mathbf{u}, \mathbf{v}')$  in eqn.4.7. If we use the idea of our proposed new estimator for first order index mentioned before, the estimators should be:

$$\overline{V}_{\mathbf{u}} = \frac{1}{2}(f(U_{N \times k})f(W'_{N \times k, i_1 i_2}) + f(W_{N \times k})f(U'_{N \times k, i_1 i_2})) \quad (4.31)$$

and

$$\overline{V}_{-\mathbf{u}} = \frac{1}{2}(f(U_{N \times k})f(U'_{N \times k, i_1 i_2}) + f(W_{N \times k})f(W'_{N \times k, i_1 i_2})) \quad (4.32)$$

However, if we check the matrix for first order as eqn.4.13 and eqn.4.27 in turn for  $i_1^{th}$  and  $i_2^{th}$  factor respectively:

$$U'_{N \times k, i_1} = \begin{bmatrix} x_{1(1)} & \cdots & x_{i_1(N+1)} & x_{i_2(1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i_1(N+2)} & x_{i_2(2)} & \cdots & x_{k(2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i_1(2N)} & x_{i_2(N)} & \cdots & x_{k(N)} \end{bmatrix} \quad (4.33)$$

$$W'_{N \times k, i_1} = \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i_1(1)} & x_{i_2(N+1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i_1(2)} & x_{i_2(N+2)} & \cdots & x_{k(N+2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i_1(N)} & x_{i_2(2N)} & \cdots & x_{k(2N)} \end{bmatrix} \quad (4.34)$$

$$U'_{N \times k, i_2} = \begin{bmatrix} x_{1(1)} & \cdots & x_{i_1(1)} & x_{i_2(N+1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i_1(2)} & x_{i_2(N+2)} & \cdots & x_{k(2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i_1(N)} & x_{i_2(2N)} & \cdots & x_{k(N)} \end{bmatrix} \quad (4.35)$$

$$W'_{N \times k, i_2} = \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i_1(N+1)} & x_{i_2(1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i_1(N+2)} & x_{i_2(2)} & \cdots & x_{k(N+2)} \\ \vdots & & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i_1(2N)} & x_{i_2(N)} & \cdots & x_{k(2N)} \end{bmatrix} \quad (4.36)$$

We can find that:

$$f(U'_{N \times k, i_1})f(W'_{N \times k, i_2}) = f(U_{N \times k})f(W'_{N \times k, i_1 i_2}) \quad (4.37)$$

$$f(W'_{N \times k, i_1})f(U'_{N \times k, i_2}) = f(W_{N \times k})f(U'_{N \times k, i_1 i_2}) \quad (4.38)$$

$$f(U'_{N \times k, i_1})f(U'_{N \times k, i_2}) = f(U_{N \times k})f(U'_{N \times k, i_1 i_2}) \quad (4.39)$$

$$f(W'_{N \times k, i_1})f(W'_{N \times k, i_2}) = f(W_{N \times k})f(W'_{N \times k, i_1 i_2}) \quad (4.40)$$

The ‘equal’ symbol in all the four equations means according to the Monte Carlo procedure for Sobol’s index, both sides have the same definition of computing given in eqn.4.7 and eqn.4.9. Take eqn.4.37 for example, from the left side, matrix  $W'_{N \times k, i_2}$  has all the other factors ‘re-sampled’ but factor set  $(X_{i_1}, X_{i_2})$  comparing to matrix  $U'_{N \times k, i_1}$ , while from the right side, matrix  $W'_{N \times k, i_1 i_2}$  has all the other factors ‘re-sampled’ but factor set  $(X_{i_1}, X_{i_2})$  compared to matrix  $U_{N \times k}$ . The difference is that from the right side, model simulation of matrix  $W'_{N \times k, i_1 i_2}$  adds extra computing cost when we have finished the first order index computing, while for the left side, all the matrices involved have already got the model evaluations in the first order index computing. So actually, by taking the left sides instead of the right sides in the second order index estimator, we can re-use the full sets of model evaluations necessary for first order index computing to compute the second order index. As we know that the computing cost of Sobol’s mainly comes from the model evaluations, in this ‘re-use’ strategy of numerical implementation,  $2 * C_k^2 * N$  model evaluations can be saved, as such to promote the computing efficiency by large.

The test result of the new method for first order index will be given in section.4.3. Before testing the estimator, we propose a way to evaluate the convergence characteristics of such type of computing methods.

## 4.2 Error estimation for Sobol’s method

### 4.2.1 Standard error, probable error

Assume function  $f(\mathbf{X}) = f(x_1, x_2, \dots, x_k)$  under investigation is defined in the  $k$ -dimensional random variable space, Monte Carlo method computing is applied to get the random distribution of  $f(\mathbf{X})$  with  $N$  sampling points, then standard error is:

$$\sigma(f) = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{n=1}^N f^2(\mathbf{X}_n) - \widehat{f}_0^2} \quad (4.41)$$

with the distribution mean value

$$\hat{f}_0 = \frac{1}{N} \sum_{n=1}^N f(\mathbf{X}_n) \quad (4.42)$$

and its variance

$$\hat{V}(f) = \frac{1}{N} \sum_{n=1}^N f^2(\mathbf{X}_n) - \hat{f}_0^2 \quad (4.43)$$

In [Homma and Saltelli, 1996], the use of the probable error  $\delta$  corresponding to the crude Monte Carlo method is computed like:

$$\delta f = 0.6745 \times \sigma(f) \quad (4.44)$$

with the population  $f_0$  having 50% chance of falling in the interval  $f_0 \pm \delta f_0$ . So, eqn.4.41 is the form we adopt for the error estimate.

Before we start for the probable error estimation of  $S_i$ 's simulation result, we should make sure that eqn.4.41 can be applied to our situation: first get the standard error of the considered variables, then we multiply the factor 0.6745 to get the probable error, based on the prerequisite that the variables should obey Gaussian distributions.

### 4.2.2 For Sobol's formula

Error estimation is of crucial interest to check whether the SA computing has properly converged. Moreover, it can be used to give confidence bound of the result. Previous work as in [Homma and Saltelli, 1996] gave interesting results for error estimation, but the conclusions are based on some restrictive assumptions:

- It assumes that the problem under analysis is scaled before computing the variance, so that  $\hat{f}_0^2$  is small.
- The variance from  $\hat{f}_0^2$  is neglected based on condition 1), so the variances of  $\hat{V}$  and  $\hat{V}_i$  are replaced by  $\bar{V}$  and  $\bar{V}_i$  respectively.

We thus perform a more comprehensive error estimation without considering these two hypotheses.

We take the H-S Sobol's method for this error analysis. To get the error estimation for the new computing method proposed in Section.4.1.3, we simply need to replace the estimator in table.4.1 with the improved ones, and change the denominator from  $N$  to  $2N$ .

**Error estimation for  $S_i$** 

**Result 1:** Approximation of the probable error for the computation of  $S_i$  is given by:

$$\begin{aligned} \delta \widehat{S}_i & \quad (4.45) \\ & \approx 0.6745 \times \sigma \widehat{S}_i \\ & = 0.6745 \times \sqrt{\frac{(\sigma \widehat{V}_i)^2}{\widehat{V}^2} + \frac{\widehat{V}_i^2 (\sigma \widehat{V})^2}{\widehat{V}^4} + 2 \frac{\widehat{V}_i (\sigma \widehat{V}) (\sigma \widehat{V}_i)}{\widehat{V}^3}} \end{aligned}$$

**Proof:** Here we rewrite the formula for Homma-Saltelli method to compute Sobol's indices:

$$\widehat{S}_i = \frac{\widehat{V}_i}{\widehat{V}} = \frac{\overline{V}_i - \gamma^2}{\widehat{V}} = \frac{\overline{V}_i - \gamma^2}{\overline{V} - f_0^2} \quad (4.46)$$

We need to evaluate the standard error of the four items:  $\sigma \overline{V}_i$ ,  $\sigma \overline{V}$ ,  $\sigma \gamma^2$  and  $\sigma f_0^2$ . According to eqn.4.41 for the definition of standard error and eqn.4.14, eqn.4.16, eqn.4.15 eqn.4.23, for the computing of  $f_0^2$ ,  $\overline{V}_i$ ,  $\overline{V}$ , and  $\gamma^2$  respectively, then:

$$\sigma \overline{V}_i = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{j=1}^N [(\mathbf{y}^{(j)} \mathbf{y}_R^{(j)})^2] - \overline{V}_i^2} \quad (4.47)$$

$$\sigma \gamma^2 = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{j=1}^N [(\mathbf{y}^{(j)} \mathbf{y}_R^{(j)})^2] - ((\gamma^2)^2)} \quad (4.48)$$

$$\sigma \overline{V} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{j=1}^N [((\mathbf{y}^{(j)})^2)^2] - \overline{V}^2} \quad (4.49)$$

$$\sigma \widehat{f}_0^2 = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{j=1}^N [(\mathbf{y}^{(j)})^2] - \widehat{f}_0^2} \quad (4.50)$$

According to eqnA.6 in appendix, the variance for  $\widehat{f}_0^2$  is:

$$V_{\widehat{f}_0^2} = (\sigma(\widehat{f}_0^2))^2 = (2 \times \widehat{f}_0 \times \sigma \widehat{f}_0)^2 \quad (4.51)$$

So:

$$\sigma(\widehat{f}_0^2) = (2 \times \widehat{f}_0 \times \sigma \widehat{f}_0) \quad (4.52)$$

and using the eqnA.8 giving the variance of a sum of random variables:

$$\sigma \widehat{V} = \sqrt{V_{\overline{V}} + V_{\widehat{f}_0^2} - 2 \text{cov}(\overline{V}, \widehat{f}_0^2)} \quad (4.53)$$

Since  $\bar{V}$  and  $\hat{f}_0^2$  use the same set of samples, the correlation index  $\rho$  can be approximated to be 1 here, so:

$$\text{cov}(\bar{V}, \hat{f}_0^2) = \rho \sqrt{V_{\bar{V}}} \sqrt{V_{\hat{f}_0^2}} \approx \sqrt{V_{\bar{V}}} \sqrt{V_{\hat{f}_0^2}} = \sigma \bar{V} \sigma(\hat{f}_0^2) \quad (4.54)$$

Then:

$$\begin{aligned} \sigma \hat{V} &\approx \sqrt{(\sigma \bar{V})^2 + (\sigma(\hat{f}_0^2))^2 - 2\sigma \bar{V} \sigma(\hat{f}_0^2)} \\ &\approx \left| \sigma \bar{V} - \sigma(\hat{f}_0^2) \right| \end{aligned}$$

so:

$$\begin{aligned} \sigma \hat{V} &\approx \left| \sigma \bar{V} - \sigma(\hat{f}_0^2) \right| \\ &\approx \left| \sqrt{\frac{1}{N} \left( \frac{1}{N} \sum_{j=1}^N [((\mathbf{y}^{(j)})^2)^2] - \bar{V}^2 \right)} \right. \\ &\quad \left. - 2 \times \hat{f}_0 \times \sigma \hat{f}_0 \right| \end{aligned}$$

Likewise, for  $\sigma S_i$  with  $\hat{V}_i = \bar{V}_i - \gamma^2$ , and correlation coefficient of  $\bar{V}_i$  and  $\gamma^2$  set to 1:

$$\sigma \hat{V}_i \approx \left| \sigma \bar{V}_i - \sigma \gamma^2 \right| \quad (4.55)$$

We then use eqnA.10 given in appendix to estimate the variables of  $\frac{\hat{V}_i}{\hat{V}}$ , but we have the same problem to get the correlation index of  $\hat{V}_i$  and  $\hat{V}$ , and it is not 1 anymore, because they don't share the same sample. However, to make sure that we will get the 'upper bound' of the real variance, we consider the least favorable case, obtained for a correlation of -1, that is to say:  $\text{cov}(\hat{V}_i, \hat{V}) \approx -\sigma \hat{V} \sigma \hat{V}_i$  and then:

$$\sigma \hat{S}_i = \sqrt{\frac{(\sigma \hat{V}_i)^2}{\hat{V}^2} + \frac{\hat{V}_i^2 (\sigma \hat{V})^2}{\hat{V}^4} + 2 \frac{\hat{V}_i (\sigma \hat{V}) (\sigma \hat{V}_i)}{\hat{V}^3}} \quad (4.56)$$

according to eqn.4.44 about the relationship between probable error and standard error, So:

$$\begin{aligned} \delta \hat{S}_i & \quad (4.57) \\ &\approx 0.6745 \times \sigma \hat{S}_i \\ &= 0.6745 \times \sqrt{\frac{(\sigma \hat{V}_i)^2}{\hat{V}^2} + \frac{\hat{V}_i^2 (\sigma \hat{V})^2}{\hat{V}^4} + 2 \frac{\hat{V}_i (\sigma \hat{V}) (\sigma \hat{V}_i)}{\hat{V}^3}} \end{aligned}$$

**error estimation for  $ST_i$** 

**Result 2:** Approximation of the probable error for the computation of  $ST_i$  is given by:

$$\begin{aligned} \delta\widehat{ST}_i & \quad (4.58) \\ & \approx 0.6745 \times \sigma\widehat{ST}_i \\ & = 0.6745 \times \sqrt{\frac{(\sigma\widehat{V}_{-i})^2}{\widehat{V}^2} + \frac{\widehat{V}_{-i}^2(\sigma\widehat{V})^2}{\widehat{V}^4} + 2\frac{\widehat{V}_{-i}(\sigma\widehat{V})(\sigma\widehat{V}_{-i})}{\widehat{V}^3}} \end{aligned}$$

**Proof:** As we check eqn.4.26, we can get that the error estimation for  $ST_i$  is similar as for  $S_i$ , the difference is changing  $\sigma\widehat{V}_i$  to  $\sigma\widehat{V}_{-i}$  and changing  $\widehat{V}_i$  to  $\widehat{V}_{-i}$  in eqn.errSi, and we can easily get that:

$$\sigma\overline{V}_{-i} = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{j=1}^N [(\mathbf{y}^{(j)} \mathbf{y}'^{(j)})^2] - \overline{V}_{-i}^2} \quad (4.59)$$

so as in eqn.4.55:

$$\sigma\widehat{V}_{-i} = |\sigma\overline{V}_{-i} - \sigma\gamma^2| \quad (4.60)$$

finally:

$$\begin{aligned} \delta\widehat{ST}_i & \quad (4.61) \\ & \approx 0.6745 \times \sigma\widehat{ST}_i \\ & = 0.6745 \times \sqrt{\frac{(\sigma\widehat{V}_{-i})^2}{\widehat{V}^2} + \frac{\widehat{V}_{-i}^2(\sigma\widehat{V})^2}{\widehat{V}^4} + 2\frac{\widehat{V}_{-i}(\sigma\widehat{V})(\sigma\widehat{V}_{-i})}{\widehat{V}^3}} \end{aligned}$$

### 4.3 Computational Tests

#### 4.3.1 Analytical functions

To demonstrate the performance of the error estimation and the efficiency of the proposed method, an artificial analytical model with 3 input variables (Ishigami Function) is considered, as in [Ishigami and Homma, 1990]:

$$f(X_1, X_2, X_3) = \sin X_1 + a \sin^2 X_2 + b X_3^4 \sin X_1 \quad (4.62)$$

where its input probability density functions (pdf) is as follows:

$$p_i(X_i) = \begin{cases} \frac{1}{2\pi}, & -\pi \leq X_i \leq \pi \\ 0, & X_i < -\pi \text{ or } X_i > \pi \end{cases} \quad i = 1, 2, 3$$

from eqn.4.2, the total variance  $V$  and partial variances  $V_i$  can be obtained analytically as:

$$V = \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2} \quad (4.63)$$

$$V_1 = \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + \frac{1}{2} \quad (4.64)$$

$$V_2 = \frac{a^2}{8} \quad (4.65)$$

$$V_3 = 0 \quad (4.66)$$

In our test case, the constants in eqn.4.63, 4.64, 4.65, 4.66 are given the values  $a = 7$  and  $b = 0.1$ . We use the ‘Mersenne Twister random number generator’ to get the random sampling for the 3 input parameters.

### 4.3.2 error estimation test

The purpose of this test is to examine how the error estimation works in the Monte Carlo computation of sensitivity with the comparison between the ones computed analytically from eqn.4.45, eqn.4.58 respectively for  $\delta\widehat{S}_i$ ,  $\delta\widehat{ST}_i$  and those obtained by repeating the computation of SA indices for a number of times (with different seeds for the MT generator) are denoted by  $\delta\widehat{S}_i^*$ ,  $\delta\widehat{ST}_i^*$ . NbR will denote the number of repetitions. All the data shown here are calculated using combined data from all runs. We use the Homma-Saltelli Sobol’s estimator in this test case in order to compare the error estimation with those obtained in the previous works in [Homma and Saltelli, 1996], denoted by  $\delta\widehat{S}_{i_{ref}}$ ,  $\delta\widehat{ST}_{i_{ref}}$ .

Since the set of all  $S_i$  plus the set of all  $ST_i$  give a fairly good description of the model sensitivities, here we only give these two sets of indices for the test.

As shown in table.4.2 and table.4.3, we can see that the theoretical upper bound error estimation  $\delta\widehat{S}_i$ ,  $\delta\widehat{ST}_i$  is very effective compared to the one we computed by numerical simulations  $\delta\widehat{S}_i^*$ ,  $\delta\widehat{ST}_i^*$ , while it is much more accurate than the previous work  $\delta\widehat{S}_{i_{ref}}$ ,  $\delta\widehat{ST}_{i_{ref}}$ . So the proposed error estimation can be a good indication to judge the level of accuracy of the SA index computation, according to which we can find a proper number of sampling (NbS) for SA.

### 4.3.3 Comparison of H-S and proposed method

We compare the Homma-Saltelli (H-S) method and the proposed new one in table.4.4 and table.4.5, with the same model evaluation computing cost corresponding to different numbers of samplings as explained in Section.4.1.3. As we mentioned in section.4.1.3,  $N_{H-S} = N_{new} * 2(k+1)/(k+2)$ , here  $k = 3$  for our test case, when  $N_{new} = 1000$ , then  $N_{H-S} = 1600$ . We can see that, both for  $S_i$  or  $ST_i$ , our proposed new method can provide more stable results, that is to say, it can get the result with less variance, as also confirmed by our standard error estimation. This result agrees with our expectation that the complimentary characteristic of the sampling matrix and re-sampling matrix can help to get a more ‘balanced’ result. Other types of estimators for Sobol indices exist in the literature. In a recent paper [Saltelli et al., 2010],

**Tab. 4.2 :** H-S Sobol's method for  $S_i$ , same Number of Runs(NbR=100), different Numbers of Samplings(NbS)

NbS=10						
Variables	$S_i(\text{exact})$	$\widehat{S}_i$	Bias	$\delta\widehat{S}_i$	$\widehat{\delta S_{i_{ref}}}$	$\delta\widehat{S}_i^*$
$X_1$	0.3138	0.332128893	0.018328893	0.420126891	0.64922391	0.394116208
$X_2$	0.4424	0.576970598	0.134570598	0.441915034	0.799264515	0.459467686
$X_3$	0	-0.047977424	0.047977424	0.368651048	0.40482392	0.329106545
NbS=100						
Variables	$S_i(\text{exact})$	$\widehat{S}_i$	Bias	$\delta\widehat{S}_i$	$\widehat{\delta S_{i_{ref}}}$	$\delta\widehat{S}_i^*$
$X_1$	0.3138	0.317898441	0.00409844	0.113155712	0.17496605	0.098588362
$X_2$	0.4424	0.44332627	0.00092627	0.110788108	0.19688452	0.072259899
$X_3$	0	0.014475517	0.014475517	0.126772756	0.129683521	0.082991208
NbS=1000						
Variables	$S_i(\text{exact})$	$\widehat{S}_i$	Bias	$\delta\widehat{S}_i$	$\widehat{\delta S_{i_{ref}}}$	$\delta\widehat{S}_i^*$
$X_1$	0.3138	0.31704138	0.00324138	0.035478791	0.056020475	0.029357167
$X_2$	0.4424	0.44426211	0.00186211	0.034121285	0.062915907	0.029415413
$X_3$	0	0.003420994	0.003420994	0.040333006	0.040555193	0.02417304

the authors compared different types of estimators for  $ST_i$ , and Jansen's [Jansen, 1999] estimator is shown to be the most efficient. It would thus be interesting to adapt the strategy proposed in this paper to other estimators to check whether the same convergence characteristics can be obtained.

#### 4.4 Discussion

With the objective of an efficient computational method for sensitivity analysis of functional-structural tree growth models, we proposed a new estimator based on Homma-Saltelli method to compute Sobol's indices. This new estimator can be considered as an effort to improve the efficiency of SA methods for models.

Another problem brought by this sampling-based computing strategy is to get results as accurate as possible but with as few samples as possible. We generally lack benchmarks to control the convergence of computing methods. Most of the time we do not have the analytical result for the sensitivity indices. Therefore, we derived a theoretical analysis of the error estimation for the sensitivity analysis for the class of Sobol's estimators (it can be applied to all the three Sobol's estimators mentioned in this paper). An analytical test function is used to test the error estimation, and we obtained that the error estimation in this paper gives out a better 'upper bound' than the previous works related to this problem. This error estimation directly relates to the variance of the result, so it can also be used for checking the confidence interval, which is usually difficult to attain.

The computation method for Sobol indices as well as the error estimation can also

**Tab. 4.3 :** H-S Sobol's method for  $ST_i$ , same Number of Runs(NbR=100), different Numbers of Samplings(NbS)

NbS=10						
Variables	$ST_i(\text{exact})$	$\widehat{ST}_i$	Bias	$\delta\widehat{ST}_i$	$\delta\widehat{ST}_{i_{ref}}$	$\delta\widehat{ST}_i^*$
$X_1$	0.5574	0.471006778	0.086393222	0.587139075	0.773255643	0.452078804
$X_2$	0.4442	0.433857546	0.010342454	0.470375601	0.80001505	0.419140296
$X_3$	0.241	0.09090012	0.15009988	0.518931875	1.04536509	0.45891867
NbS=100						
Variables	$ST_i(\text{exact})$	$\widehat{ST}_i$	Bias	$\delta\widehat{ST}_i$	$\delta\widehat{ST}_{i_{ref}}$	$\delta\widehat{ST}_i^*$
$X_1$	0.5574	0.54219821	0.01520179	0.123465946	0.21247326	0.11183428
$X_2$	0.4442	0.457708634	0.013508634	0.126828544	0.23284288	0.08433098
$X_3$	0.241	0.238775324	0.002224676	0.1248745	0.26201749	0.114110644
NbS=1000						
Variables	$ST_i(\text{exact})$	$\widehat{ST}_i$	Bias	$\delta\widehat{ST}_i$	$\delta\widehat{ST}_{i_{ref}}$	$\delta\widehat{ST}_i^*$
$X_1$	0.5574	0.55231685	0.00508315	0.038581076	0.067597875	0.037170484
$X_2$	0.4442	0.43980209	0.00439791	0.040345109	0.076681373	0.028618586
$X_3$	0.241	0.23869646	0.00230354	0.035262641	0.084598942	0.034681808

**Tab. 4.4 :** Efficiency comparison of H-S and our new Sobol's method for  $S_i$ , with the same number of model evaluations, H-S method with Number of Samplings(NbS=1600), Number of Runs(NbR=100), new method with Number of Samplings(NbS=1000), Number of Runs(NbR=100)

H-S						
Variables	$S_i(\text{exact})$	$\widehat{S}_i$	Bias	$\delta\widehat{S}_i$	$\delta\widehat{S}_{i_{ref}}$	$\delta\widehat{S}_i^*$
$X_1$	0.3138	0.31533298	0.00153298	0.02799958	0.044172429	0.022979037
$X_2$	0.4424	0.44665802	0.00425802	0.027065163	0.049959473	0.020148217
$X_3$	0	-0.000991375	0.000991375	0.031977602	0.031926896	0.018688355
New						
Variables	$S_i(\text{exact})$	$\widehat{S}_i$	Bias	$\delta\widehat{S}_i$	$\delta\widehat{S}_{i_{ref}}$	$\delta\widehat{S}_i^*$
$X_1$	0.3138	0.31298693	0.00081307	0.024947723	0.051779454	0.014254
$X_2$	0.4424	0.44600325	0.00360325	0.024195279	0.062443818	0.014995118
$X_3$	0	-0.000171562	0.000171562	0.028633187	0.02861765	0.01049768

**Tab. 4.5 :** Efficiency comparison of H-S and our new Sobol's method for  $ST_i$ , with the same number of model evaluations, H-S method with Number of Samplings(NbS=1600), Number of Runs(NbR=100), new method with Number of Samplings(NbS=1000), Number of Runs(NbR=100)

H-S						
Variables	$ST_i(\text{exact})$	$\widehat{ST}_i$	Bias	$\delta\widehat{ST}_i$	$\delta\widehat{ST}_{i_{ref}}$	$\delta\widehat{ST}_i^*$
$X_1$	0.5574	0.55433338	0.00306662	0.030524063	0.053367661	0.028061632
$X_2$	0.4442	0.44145391	0.00274609	0.031846525	0.060492716	0.024130443
$X_3$	0.241	0.24447647	0.00347647	0.027831009	0.066898161	0.02334635
New						
Variables	$ST_i(\text{exact})$	$\widehat{ST}_i$	Bias	$\delta\widehat{ST}_i$	$\delta\widehat{ST}_{i_{ref}}$	$\delta\widehat{ST}_i^*$
$X_1$	0.5574	0.55543245	0.00196755	0.027455945	0.065578754	0.018760671
$X_2$	0.4442	0.44419771	2.29E-06	0.028548756	0.076191397	0.01669605
$X_3$	0.241	0.23800909	0.00299091	0.024822554	0.089599272	0.022789275

be performed with other types of random number generator which are proved to also have an important influence on the efficiency of the algorithms [Saltelli et al., 2010], therefore, same tests were performed with quasi-random generator [Saltelli et al., 2010] leading to the same conclusions.

## 5. STRATEGY DESIGN

In the previous chapter, we present our work on the computational issue of Sobol's method, which is the first important issue in our thesis. In this chapter, we focus on the second issue we proposed at the beginning of this thesis: the strategy design of SA for FSPMs. We first give a general view of this topic in section.5.1. We mainly introduce the strategy design in two aspects: non-linearity assessment from SRC method in section.5.2 and the more general methodology for full SA of complex biophysical models in section.5.3.

### 5.1 Introduction

A good sensitivity analysis practice does not only needs well designed SA estimators (as in the previous chapter) but also needs good understanding of how to comprehensively use more than one methods to make them be complementary to each other since different methods tackle different issues of interest. This is what we consider as 'strategy design'.

Though pointed out in [Saltelli et al., 2004], one property of an ideal sensitivity analysis method is that it should be 'model independent', which means a method should work regardless of the attributes of the model itself like additive, linear, etc. However, the strategy design is necessary in our work of sensitivity analysis for FSPMs. More work needs to be done for exploring how global sensitivity analysis can help in the parameterization of FSPM, by quantifying the driving forces of the phenomena described by the models and the relative importance of the described biophysical processes regarding the outputs of interest.

As we have mentioned in the performance steps of SA in section.3.6, the strategy design includes:

- To choose a proper sensitivity analysis method to fulfill different aims.
- To combine more than one SA method in order to make best use of each method's advantages and to make them complementary to each other.

For the first type of issue, in Chapter.3, we have already recalled that:

- Basically we can choose between local methods and global methods. Local methods cannot include the effect of the shape of the density functions of the inputs,

and they are not model independent but they are really computationally cheap. Global approaches estimate the effect on the output of a factor keeping all the others varying. Generally, global approaches use model independent methods and do not require assumptions of additivity or linearity. As a drawback, they are usually computationally expensive [Cariboni et al., 2007]. So it depends on our knowledge about the properties of the model and the computing cost we can afford.

- The Standardized Regression Coefficients (SRCs) can be viewed as an interesting trade off between local and global method, regarding the advantages and shortcomings of both: the accuracy of the analysis and the computing cost. The determination coefficient computed by SRCs can be used in non-linearity assessment for models.
- Alternatively, to fulfill our aim of ‘Factor Prioritization setting’ and ‘Factor Fixing setting’ in our SA for FSPMs, we can choose Sobol’s method to get the first order index of  $S_i$  and total order index  $ST_i$ . Moreover, with Sobol’s method, we can get the quantitative parameter interaction evaluation based on the variance decomposition by  $S_{ij}$ . For a given factor  $X_i$ , the value of first-order Sobol’s index  $S_i$  indicates that whether a factor is mainly influent, while an important difference between  $ST_i$  (Total order effect) and  $S_i$  flags an important role of interactions for that factor in output  $y(t)$ . In this case, inspection of the second order index  $S_{ij}$  for all  $i \neq j$  will allow us to identify which factor  $X_i$  interacts with.
- The group indices given in section.3.5.2:  $S_{\Omega_i}^g$ ,  $ST_{\Omega_i}^g$  are useful when our objective is to analyze the effect of a subset of factor (usually factors from one module or sub-model in the whole model) to the output variance, and the difference  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  means the interactions of the set of factors with those that do not belong to the set.  $S_{\Omega_{ij}}^g$  specifically indicate the interactions between two subsets of factors.

Complex biological models are usually characterized by several interacting processes with sub-models describing each of them. Most FSPMs are such models. It is interesting to evaluate the importance of the sub-models (usually ‘function’ modules corresponding to the biophysical processes they describe) by sensitivity analysis. For this objective, in practice we need to firstly classify the parameters into different biological function modules according to the biologist modeller’s expert knowledge, then to check the joint sensitivity effects of the groups of parameters that belong to those modules. This is how ‘module-by-module’ analysis for complex biophysical system is put forward. The strategy design should be divided into several steps for which we choose different SA methods to fulfill different requirements.

The choice of a proper sensitivity analysis method to fulfill different aims of different sub-steps of the analysis faces the same SA general issues as mentioned before.

However, module-by-module analysis requires us to make the combination of more than one SA method in order to make best use of each method's advantages and to make them complementary to each other. We will propose a module by module analysis for complex FSPMs inspired by the strategy in [Ruguet et al., 2002] in section.5.3.

## 5.2 Standardized Regression Coefficients (SRC) method and non-linearity assessment

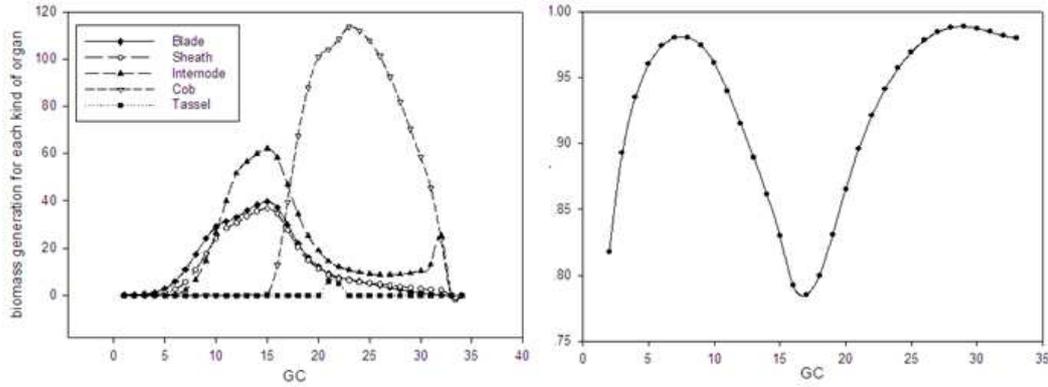
Plants are known as complex systems with a strong level of interactions and compensations, and one of the aims of FSPMs is to describe and understand this complexity. As such, non-linearity is expected to play a key role in the study. The knowledge of the intrinsic non-linearity of the model and of its dynamic evolution throughout plant growth is very useful for the study of model behavior and properties, to underline the occurrence of particular biological phenomena or to improve the statistical analysis when confronting models to experimental data (e.g. statistical properties of estimators or numerical methods to compute the propagation of errors [Julier et al., 2000]).

As detailed in Chapter.3, the Standardized Regression Coefficients method is based on a global linear approximation of the model. It provides the model coefficient of determination  $R^2$ , which represents the fraction of the output variance explained by the linear regression model itself. When  $R^2 = 1$ , the system is linear and the SRCs can totally explain the variance of the output affected by each factor. Even when models are moderately non-linear (i.e.  $R^2 > 0.9$ ), the SRCs can provide valid qualitative information. When it gets small, the SRCs are no longer reliable sensitivity representations. To be more direct, we call this model coefficient of determination linearity index, and use it to assess the non-linearity of the model.

The objective of this section is thus to explore the level of linearity of the 3 FSPMs introduced in Chapter.2, which have different levels of complexity, and infer in each case what information can be drawn from this analysis.

We recall here generally the three FSPMs involved in this work. Firstly a simple source-sink model of maize growth, which is used to specifically study the process of carbon(C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations, secondly the GreenLab model of tree growth (applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis [Mathieu et al., 2008], which describes tree structural plasticity in response to trophic competition, lastly a functional-structural model, NEMA [Bertheloot et al., 2011a], describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering.

In this study, we follow dynamically the linearity index throughout plant growth. In the models of the GreenLab type, the output of interest for the SA is the biomass



**Fig. 5.1 :** GreenLab Maize (a)At each GC, biomass allocation per organ type (b)Evolution of the linearity index

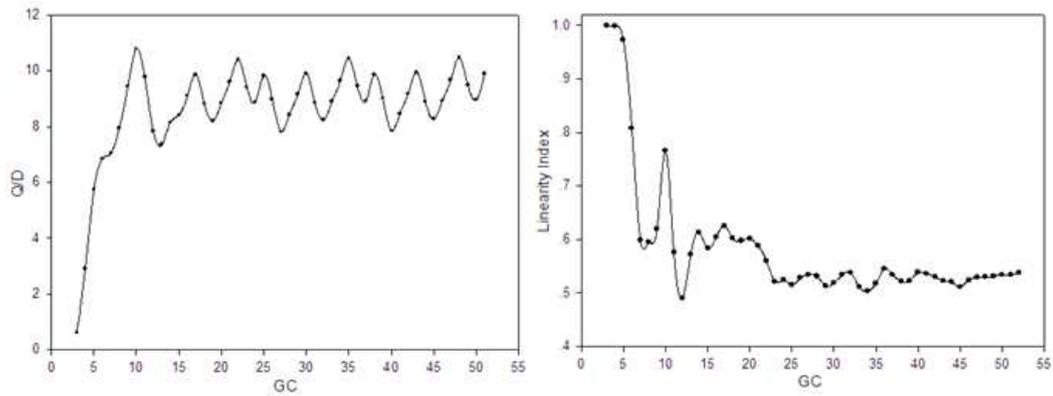
production at each GC, and in the NEMA model, grain dry mass, grain N mass and root N uptake are considered.

The linearity analysis for the biomass production for GreenLab Maize is given in fig.5.1.b. A non-linear period is denoted by the minimum around GC17. It is a key step in terms of biophysical processes corresponding to the transition between two allocation phases, the first one corresponding mostly to leaf area increase and the second one to grain filling (as illustrated by fig.5.1.a showing the biomass allocation to each type of organ).

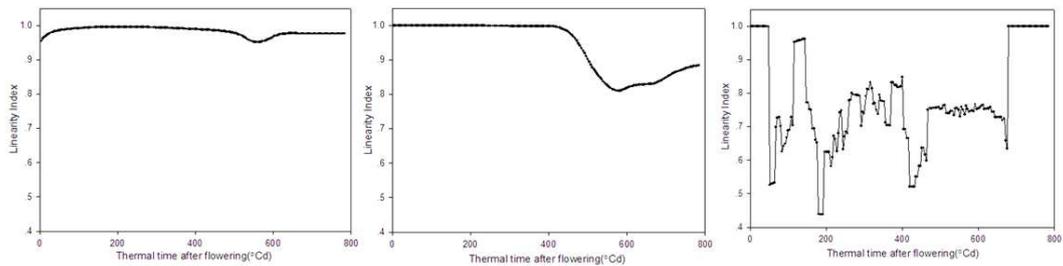
For the poplar tree model with retroaction of functioning on organogenesis, fig.5.2.a shows the evolution of the ratio for biomass production to organs' demand which is the key variable controlling tree organogenesis. The linearity index is shown in fig.5.2.b. The initial states show high linearity indices, but these decrease rapidly with the increasing influence of the trophic competition on organogenesis, particularly the appearance of the first branch at growth cycle 5. The linearity index stabilizes at around 0.5, which is quite low, showing the importance of retroaction in the dynamic system.

For the model of C-N dynamics (NEMA), the evolutions of the linearity indices for N mass of the grains, their dry mass and the rate of root N uptake are shown in fig.5.3. If the model is highly linear for the grain dry mass fig.5.3.a, it is not the case for grain Nitrogen mass fig.5.3.b. For N uptake by root fig.5.3.c, the evolution of the linearity index is hierarchical with a low level of linearity. It seems reasonable when we consider the complex driving forces for N uptake in the model, involving the function of the nitrate concentration in the soil modulated by positive and negative feedbacks of respectively C and N in the plant, on root activity.

Considering the dynamic evolution of the linearity index throughout plant growth may reveal phases during which linearity is very high, probably showing some stable



**Fig. 5.2 :** Poplar tree model with retroaction of functioning on organogenesis: (a) Evolution of ratio of available biomass to organs' demand (Q/D) (b) Evolution of linearity index



**Fig. 5.3 :** NEMA: Evolution of the linearity index for different outputs (a) Dry mass of the grains (b) Nitrogen mass of the grains (c) Root N uptake

behavior for which the experimental process for parameter estimation could be lighter. On the contrary, the analysis may show evidence of some moments when a strong non-linearity occurs. Such moments may be characteristic of very specific biological phenomena during plant growth as well as the high level of interactions between parameters, either known by the modeler or unknown, in which case they should probably be investigated more. For non-linear phases, a special care should be taken of with respect to the frequency and details of experimental measures.

For example, since the parametric estimation of GreenLab by model inversion relies on multi-stage observations, the information given by the linearity analysis can be taken advantage of. For maize, it is important to have detailed and frequent observations between cycles 14 and 20, while the measurements can be lighter after cycle 20. For the poplar tree model, the high level of non-linearity is coupled with the difficulty to get regular observation data (due to the time scale and the investment level of experiments). The linearity analysis should thus be coupled with a full SA to help us define a proper strategy for parameter estimation: for example, in two steps as proposed by [Letort et al., 2008a] first consider the observed topology as fixed to get the functional parameters, and then estimate the parameters driving the retroaction of functioning on organogenesis.

For complex bio-physical models like NEMA, we expect a strong genetic determinant of model parameters. One of the interests of assessing the level of model linearity is that non-linear phases are characterized by a high level of interactions between parameters. When model parameters are genetic, understanding and quantifying this interaction is crucial in the objective of using plant models as an intermediate to develop a predictive capacity from genotype to phenotype [Hammer et al., 2006; Buck-Sorlin and Bachmann, 2000] and design ideotypes. If a parameter has little interaction with others, we can directly concentrate on this trait for the design of ideotypes [Qi et al., 2010]. If the interaction is strong, it is more complex. If the parameters are strongly genetically related (determination by the same genes), the model parameterization should be improved to take into account this fundamental interaction. If they are not genetically related, breeding strategy should rely on multi-dimensional optimization to handle the interacting processes. The next objective of our study is thus to use Sobol's method [Sobol, 1993] to explore the interactions between processes and parameters for this model.

### 5.3 A methodology for complex biophysical models

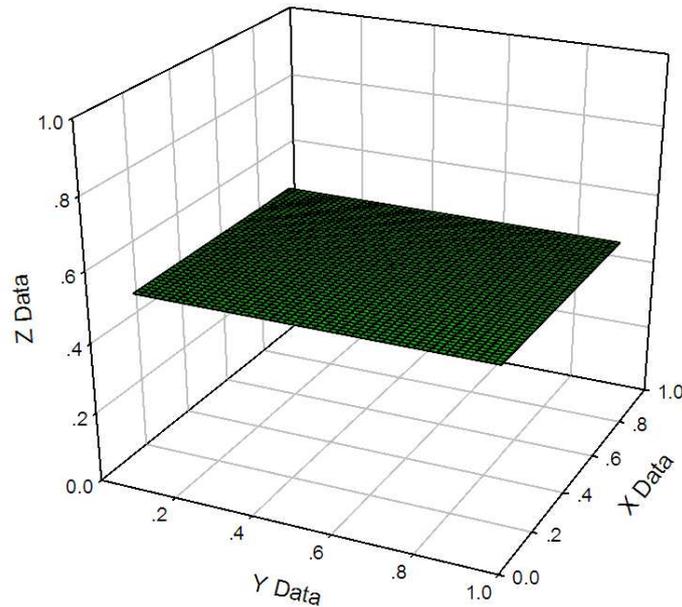
Complex biological models are usually characterized by sub-models describing several interacting processes respectively. Most FSPMs are such models. It is interesting to evaluate the importance of the sub-models (usually 'function' modules corresponding to the biophysical processes they describe) by sensitivity analysis. For the objective, in practice we need to firstly classify the parameters into different biological function modules according to the biologist modeller's expert knowledge, then to check the joint

sensitivity effects of the groups of parameters that belong to those modules. This is how ‘module-by-module’ analysis for complex biophysical system is put forward. The strategy design should be divided into several steps for which we choose different SA methods to fulfill different requirements.

Our proposed methodology for module by module analysis is inspired by [Ruget et al., 2002], in which analysis was made in two steps: first, within each meta-process (module), the most important parameters are identified; then sensitivity to identified parameters is calculated taking into account together all meta-processes. A certain number of parameters (according to empirical information) are selected from each module for the analysis. The influence of parameters inside their own modules was studied through response surface methodology, in order to give sensitivity values linking outputs and parameters and to choose the most representative parameter of each module to use in the intermodule analysis. To do the internal analysis of each module, they sampled the parameters selected for this module according to their distributions while fixing the parameters in all the other modules to their mean values, then the parameter samples are used for the regression surface functions to resolve the indices. Afterwards, they did the internal modules analysis: using response surface method with all the parameters selected from each module in the previous step.

However, there are several drawbacks for this strategy:

- There are a lot of limitations for using the regression method. First, strict requirements have to be fulfilled for the model functions like linear, additive, or for surface response, the model has to be a continuous system, for FSPMs this requirement can not always be satisfied. Besides, only second order ‘interaction’ has been found in [Ruget et al., 2002]. The higher order regression for getting the group interactions from more than two parameters is also a difficult issue for the surface response method.
- Usually, ‘a certain number of parameters (according to empirical information) are selected from each module for the inter-module analysis’. It is risky to rely on such empirical information that may be misleading afterwards, especially regarding the parameter space issue. There may be some parameters missed at this step, like the ones that have important effect on the output, for example, through interactions with the others. And since each module mostly tends to have different importance in the model, if we decide empirically the number of parameters selected for each module, it may cause that for some modules we select not enough parameters and for some modules we select too many. This decision directly affects the importance evaluation in the final step. So we need a quantitative standard to choose the proper number of parameters from each module, even though in the internal analysis of each module, the results we get should be given in a unified framework to be comparable. With regard to this, it corresponds to keeping the same sampling space while doing the Monte Carlo simulations.



**Fig. 5.4 :** Mesh grid 3 dimension parameter space. All the samplings for X and Y obey uniform distribution in  $[0,1]$ , while fixating Z to its mean value 0.5.

- To consider parameters only in one module, while fixing the other parameters to their mean values, the SA indices obtained this way can not stand for the importance of the parameters in the complete space, but to a surface formed by the fixed values of the parameters in the space. Take the model with the 3 dimension parameter space for example, as shown in fig.5.4, all the samplings for X and Y follow a uniform distribution in  $[0,1]$ , while fixing Z to its mean value 0.5, the sampling parameter space is not the cube anymore, it is restricted to the surface given by the function  $Z = 0.5$ . In addition, by fixing the parameters in the other modules, the interactions between parameters from different modules will be eliminated even though it may prove to be important.

To make all steps of this ‘module by module analysis’ more quantitatively precise, we worked out a strategy specifically suitable to the characteristics of FSPMs. It is a procedure that combines both SRC and Sobol’s method, so that both the advantages of SRC’s computing efficiency and Sobol’s quantitative analysis can be used.

**Step1: Non-linearity study with SRC** As we mentioned in section.3.5.2 we first adopt the SRC for our analysis. This step does not concern any module division issues, because we took into account all the parameters with the uncertainties in the analysis. Since it is based on linear regression, computing cost is not a problem for us to get the first approximate overview of the SA for all the parameters. It is necessary for us to know the information of the basic non-linearity evolving by time step for the

model, so that we can make an analysis decision qualified by the SRC result for the next step. And it also helps for checking the basic ranking of the parameters. Though SRC does not necessarily give out the precisely ‘correct’ ranking of the parameters especially for the system with low linearity, normally the first overview given by SRC could be a general background information for the model.

**Step2: Group analysis** Our objective at this step is to compute the sensitivity indices of the total set of parameters corresponding to each module of the model. First we need to make a partition of the parameters with uncertainty, so that each subset in the partition corresponds to one biological function module of our model. Then we compute the group indices:  $S_{\Omega_i}^g$ ,  $ST_{\Omega_i}^g$ ,  $S_{\Omega_{ij}}^g$  and the difference  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  as defined in section.3.5.2. As far as numerical computing is concerned, we use the similar matrix definition from section.4.1.3, except when we configure the matrix  $U'_{N \times k}$  and  $W'_{N \times k}$ , we exchange the corresponding multi-columns of  $W_{N \times k}$  and  $U_{N \times k}$  with the column identity of  $i_1, i_2, \dots, i_s$ , where the set of parameter identities  $\{i_1, i_2, \dots, i_s\}$  includes all the parameters belong to the module under analysis. We rewrite the matrix here:

$$\begin{aligned}
 U_{N \times k} &= \begin{bmatrix} x_{1(1)} & \cdots & x_{i(1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i(2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i(N)} & \cdots & x_{k(N)} \end{bmatrix}, \\
 W_{N \times k} &= \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i(N+1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i(N+2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i(2N)} & \cdots & x_{k(2N)} \end{bmatrix}, \\
 U'_{N \times k} &= \begin{bmatrix} x_{1(1)} & \cdots & x_{i_1(N+1)} & x_{i_2(N+1)} & \cdots & x_{i_s(N+1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i_1(N+2)} & x_{i_2(N+2)} & \cdots & x_{i_s(N+2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & & \vdots & & \\ x_{1(N)} & \cdots & x_{i_1(2N)} & x_{i_2(2N)} & \cdots & x_{i_s(2N)} & \cdots & x_{k(N)} \end{bmatrix}, \\
 W'_{N \times k} &= \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i_1(1)} & x_{i_2(1)} & \cdots & x_{i_s(1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i_1(2)} & x_{i_2(2)} & \cdots & x_{i_s(2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & & \vdots & & \\ x_{1(2N)} & \cdots & x_{i_1(N)} & x_{i_2(N)} & \cdots & x_{i_s(N)} & \cdots & x_{k(2N)} \end{bmatrix}.
 \end{aligned}$$

Based on those four matrices and the method in section.4.1.3, we obtained the  $S_{\Omega_i}^g$ ,  $ST_{\Omega_i}^g$  for each module and  $S_{\Omega_{ij}}^g$  for each pair of modules. To distinguish from the sensitivity index of one parameter, we use a superscript ‘g’ to identify the index of one group (one module) in this step. The computing cost for this step is not so heavy because there are 5 modules which is much fewer than the number of parameters, with the sampling number  $N$ , the model evaluations we need here is  $2N(k+1)$  with  $k=5$ . There are several small steps that relate to how to use the different SA index generated by Sobol’s method to complete our analysis aims:

- Ranking module’s importance given by  $S_{\Omega_i}^g$ . Different function modules may

dominate different main biological processes at different stages. This group analysis can help us to know the main biological process in plant growth stages. By this step, the general scheme of the sensitivity distribution among modules is given. It is necessary to know the percentage of the first order sensitivity each module takes from the whole model, because it is the background information of the next step which provides us a quantitative clue for screening parameters.

- Identifying the intermodule interactions by computing the difference between  $ST_{\Omega_i}^g$  and  $S_{\Omega_i}^g$  for each module. It helps us to locate where the main interactions exist, and to which extend a module can be studied independently from the other modules.
- Identifying the interactions from pairs of modules by  $S_{\Omega_{ij}}^g$ . It is the specified information about the interactions inter-modules, so that it indicate us the interactions between biological processes.

**Step3: Internal module analysis** The main aim of this step is to identify the most sensitive parameters in each specific module quantitatively. On one hand we need to consider each module separately, on the other hand, it is necessary for us to keep one common analysis framework for all the modules, so that the quantitative comparison can be possible between modules. So here we choose to make the analysis with the same sampling points in the same sampling space to make the variance for each module analysis to be the same. We stick to the improved Sobol's mentioned in section.4.1.3 to make the analysis. We rewrite the four matrices used here to make the basis of the two analysis basis clear:

$$\begin{aligned}
 U_{N \times k} &= \begin{bmatrix} x_{1(1)} & \cdots & x_{i(1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i(2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i(N)} & \cdots & x_{k(N)} \end{bmatrix} \\
 W_{N \times k} &= \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i(N+1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i(N+2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i(2N)} & \cdots & x_{k(2N)} \end{bmatrix} \\
 U'_{N \times k} &= \begin{bmatrix} x_{1(1)} & \cdots & x_{i_m(N+1)} & \cdots & x_{k(1)} \\ x_{1(2)} & \cdots & x_{i_m(N+2)} & \cdots & x_{k(2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(N)} & \cdots & x_{i_m(2N)} & \cdots & x_{k(N)} \end{bmatrix} \\
 W'_{N \times k} &= \begin{bmatrix} x_{1(N+1)} & \cdots & x_{i_m(N+1)} & \cdots & x_{k(N+1)} \\ x_{1(N+2)} & \cdots & x_{i_m(N+2)} & \cdots & x_{k(N+2)} \\ \vdots & & & \ddots & \vdots \\ x_{1(2N)} & \cdots & x_{i_m(2N)} & \cdots & x_{k(2N)} \end{bmatrix}
 \end{aligned}$$

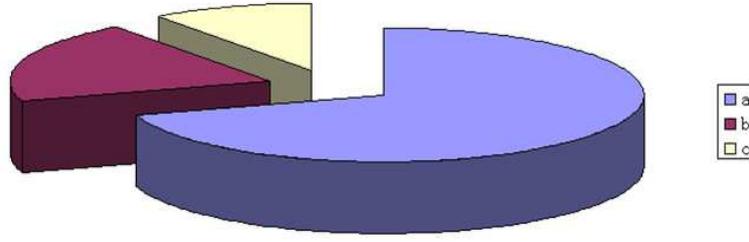
In matrices  $U'_{N \times k}$  and  $W'_{N \times k}$ , the parameter identity  $i_m$  means any item in the parameters' identity set  $\{i_1, i_2, \cdots, i_s\}$  corresponding to the module we focus on. So actually

for each module analysis, we used the same sampling and re-sampling matrix:  $U_{N \times k}$  and  $W_{N \times k}$ , in this case, the output variance would be the same for all the modules, and the sum of the variance from each module would be equal to the one obtained for the Sobol's analysis for the overall model based on the same  $U_{N \times k}$  and  $W_{N \times k}$ .

One side advantage is that every module analysis shares the same  $2N$  model evaluations of  $U_{N \times k}$  and  $W_{N \times k}$ , and we simply add the model evaluations for their respective  $U'_{N \times k}$  and  $W'_{N \times k}$ . For example, if we have  $r$  modules, and  $k_1, k_2, \dots, k_r$  are the numbers of parameters for each module,  $k_1 + k_2 + \dots + k_r = k$ ,  $k$  is the total number of parameters with uncertainty for the computed model. Using the improved Sobol's method in section.4.1.3, if we use the strategy in [Ruguet et al., 2002], the computing cost would be  $2N(k_1 + 1) + 2N(k_2 + 1) + \dots + 2N(k_r + 1) = 2N(k_1 + k_2 + \dots + k_r) + 2Nr = 2N(K + r)$ , instead in our strategy, the computing cost would be  $2N + 2Nk_1 + 2Nk_2 + \dots + 2Nk_r = 2N(1 + k_1 + k_2 + \dots + k_r) = 2N(K + 1)$ , it can save  $2N(r - 1)$  model evaluations, the most important point is that it keeps all the independent module analysis being comparable by keeping them in the same parameter space.

In fact the computing cost for module by module analysis in our strategy is the same if we make the analysis of the overall model. So why do not we just do the overall model analysis? Why bother to do this 'module by module analysis'? Firstly in [Cariboni et al., 2007], the author pointed out that the choice of most suitable technique for sensitivity analysis depends on the number of factors of the model and on the CPU time required to run it. As we cut the whole set of factors into several groups for the model, for each module when we do the SA, the number of factors is more appropriate for variance based technique which Sobol's method belongs to. Secondly it helps to get the analysis objectives more clearly, especially with the complex model with large number of parameters. And indeed for the computing cost, to compute separately module by module or all modules together have the same cost. But by combining the result of module by module analysis and group analysis, we can save a lot of higher order SA index computing cost.

- In section.3.5.2, we have mentioned that the total sensitivity index  $ST_i$  is used for screening parameters: when  $ST_i \approx 0$  parameter  $X_i$  can be screened from the parameters with uncertainty, which means that the variance of this *ith* parameter brings for the output can be ignored. So the module by module analysis in this step, we first check the index  $ST_i$  to screen the ignorable parameters in each module. Based on this strategy, the different indices obtained in the different modules can be compared together. Then  $ST_i$  is the quantitative standard to pick the most important parameters from each module.
- The comparable results of all the modules provide a quantitative standard to give the percentage of the variance explained by the selected parameters more precisely.
- Comprehensively consider the result in step 2 'group analysis': if we compute the sum of each  $S_i$  from step3 for the parameters in the module, and get the



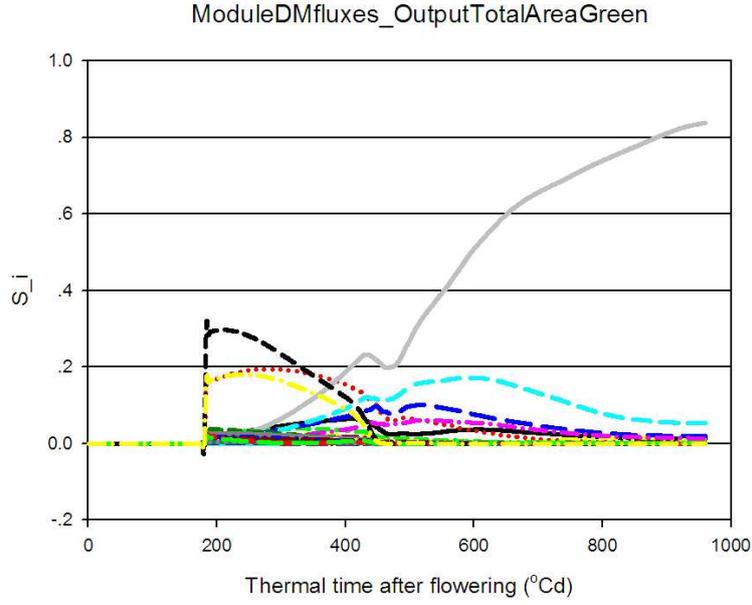
**Fig. 5.5 :**  $ST_{\Omega_i}^g$  for one module decomposition scheme. a. The sum of all the  $S_i$  for the parameters from one module in Step 3, b.  $S_{\Omega_i}^g$  for the module minus a, intra-module interactions from step 2, c.  $ST_{\Omega_i}^g$  minus  $S_{\Omega_i}^g$  for one module, inter-module interactions in step 2

difference of it with the  $S_{\Omega_i}^g$  (the first order sensitivity of each parameter plus the interactions between them within the module), for the corresponding module we will get the inner module interactions, which means the interactions between the parameters only in the module. Moreover as we mentioned in step 2, the difference between  $ST_{\Omega_i}^g$  and  $S_{\Omega_i}^g$  for one module stands for the interactions of the module to the others. These three parts: a. The sum of each  $S_i$  from step3 for the module, b.  $S_{\Omega_i}^g$  for the module minus a, c.  $ST_{\Omega_i}^g$  minus  $S_{\Omega_i}^g$  for one module, composite the  $ST_{\Omega_i}^g$  of one module as shown in fig.5.5.

**Step4: Overall model analysis** To make a complete sensitivity analysis with the selected parameters by the steps before. It is a ‘loop’ back to the general SA practice steps we presented in section.3.6, except that the decision for the input factors has been made.

## 5.4 Processing for functional output case

As we mentioned before, functional-structural plant models are dynamic models and the output of interest evolve with time. Screening methods in sensitivity analysis generally consider the case that the output is scalar. Although SA could be performed separately at each time step for FSPMs, in the case of dynamic crop models, simulations are usually computed at a daily time step and the sequential implementation of global sensitivity analysis at each simulation date can result in several hundreds of sensitivity indices, with one index per parameter per simulation date. It is not easy to identify the most important parameters based on such a large number of values. Moreover, this technique has the disadvantage of introducing a high level of redundancy because of the strong correlations between responses from one time step to the next one. Likewise, intuitively, to screen one parameter, the requirement should be that the sensitivity response curve is always near zero at all time steps, which is rarely the case in our analysis. We therefore devise a methodology to deal with the screening of the parameters in this dynamic context.



**Fig. 5.6 :** NEMA:  $S_i$  for 35 parameters in module dry mass fluxes (DMflux), with output of total green area (AreaGreenTotal).

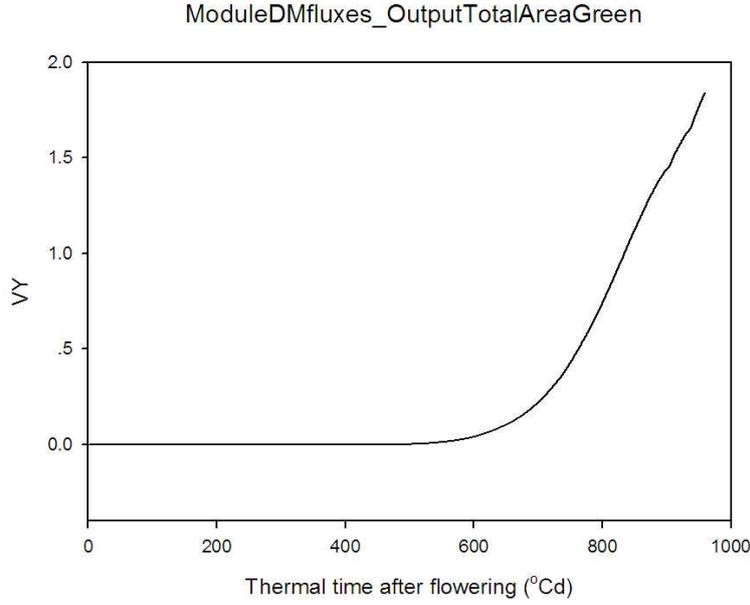
We use here a unitary concept based on the processing of averaging across time. Some parameters may have the biggest sensitivity indices at some stages for which the output variance is negligible. For this reason, with an averaging process, we weigh the sensitivity indices with the output variance at time  $t$ . In this way, we select the parameters that not only have the most important sensitivity at certain time steps, but also indicate the duration of this effect. We show a result for one module in NEMA model as example, see fig.5.6 and fig.5.7. There are 35 curves of parameter indices in fig.5.6, and few are close to zero through the whole time period, but if we check the output variance through time, we will see that until thermal time 450 °Cd, the total variance is very small, which means that the parameters dominating this period should be considered differently from the one dominating the period afterwards.

As such, we define a unitary index for each curve describing the first order and total order sensitivity averaging through time, and we call it ‘Time Generalized Index (TGI)’ for parameter with identity  $i$ :

$$TGIS_i = \frac{\sum_{t=1}^T V(Y(t))S_i(t)}{\sum_{t=1}^T V(Y(t))} \quad (5.1)$$

$$TGIST_i = \frac{\sum_{t=1}^T V(Y(t))ST_i(t)}{\sum_{t=1}^T V(Y(t))} \quad (5.2)$$

$T$  is the total number of time steps,  $V(Y(t))$  means the output variance at time step  $t$ ,  $S_i(t)$  and  $ST_i(t)$  means the sensitivity indices evolving with time  $t$ .



**Fig. 5.7 :** NEMA: Total variance of output, for 35 parameters in module dry mass fluxes (DMflux), with output of total green area (AreaGreenTotal).

$TGIS_i$  and  $TGIST_i$  are convex combinations of  $S_i(t)$  ( $1 \leq t \leq T$ ) so that the same properties as classical  $S_i$  remain the same. For example, if the model is linear, for all the parameters:

$$\sum_i TGIS_i = 1 \quad (5.3)$$

$$TGIST_i - TGIS_i = \frac{\sum_{t=1}^T V(Y(t))}{\sum_{t=1}^T V(Y(t))} (ST_i(t) - S_i(t)) \quad (5.4)$$

so it is characteristic of the total level of interaction for parameters.

The TGI helps us to simplify the screening processing, and at the same time, consider the evolution of the index with time, in order to guarantee that the generalized index we get can reflect the overall effect of the index on the output.

There has been other works that dealt with this issue. In [Ruget et al., 2002], the author calculated the average slope of the curve representing the output variable in relation to the modified input parameters. In [Campbell et al., 2006], the author used a principal component analysis of output temporal curves, then compute sensitivity indices of each input on each principal component coefficient; in [Lamboni et al., 2011] they developed the multivariate global sensitivity analysis method. It allows to aggregate the different sensitivity indices of the principal component coefficients in a unique index, called the generalized sensitivity index. Each generalized sensitivity index explains the influence of the corresponding input on the overall output curve variability. A comparison with these methods is currently under study.

## 5.5 Concluding remarks and discussions

The objective of this study is to explore how global sensitivity analysis (SA) can help the design of complex models in two aspects:

- To choose proper sensitivity analysis method to fulfill different aims.
- To combine more than one SA method in order to make best use of each method's advantages and to make them be complementary to each other.

For the first aspect, we specially presented the SRC method and non-linearity assessment to investigate the intrinsic non-linearity of the model and of its dynamic evolution throughout plant growth, thus to study model behavior and properties, to underline the occurrence of particular biological phenomena and hopefully, to improve the statistical analysis when confronting models to experimental data.

For the second aspect, we developed a methodology for complex biophysical models: module by module analysis inspired by [Ruget et al., 2002]. In contrast with previous types of module by module analysis, we proposed a more reliable and effective strategy of use:

- First, linearity analysis gives us the non-linearity stage information of the overall model. This preliminary information is also useful to adapt the following strategy accordingly.
- Second, group analysis provides the evolution of module importance  $S_{\Omega_i}^g$ , and help us check inter-module interactions by  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  and  $S_{\Omega_{ij}}^g$ .
- Thirdly, based on the same sampling points, we run the SA module by module, in order to provide index  $ST_i$  of each parameter for screening. It is the basic index for screening procedure but not the only standard. The question of how many parameters should be selected is decided by comprehensively considering the results of the first and second steps. Moreover, by analyzing the composition of  $ST_{\Omega_i}^g$ , we get to know the intra-module and inter-module interaction quantitatively.
- Finally, we run the SA for the overall model with the parameters selected module by module and check the parameter sensitivity indices for the overall model.

Conducted simulations using Sobol's method and an efficient computation technique derived from [Wu et al., 2011] will be presented in Chapter.7, and several outputs of interest are considered specially for NEMA to check how parameter effects change with the the outputs of interest.

Moreover, since we consider a dynamic system, the evolution of the sensitivity indices is computed. And when it is related to parameter screening, we proposed

a time averaging index called TGI to reduce the time dimension of the sensitivity indices. Though it has not been strictly proven to be valid, it worked very well for our application. Still it is interesting to try the method in [Campbell et al., 2006] and [Lamboni et al., 2011] based on principal component analysis to see which method is more effective and reliable.

Other methods of SA for screening for complex models can also be used, like Morris. There are references using the combination of Morris method and Sobol's method for the analysis of complex models. Though improved by [Campolongo and Braddock, 1999] and [Cropp and Braddock, 2002], Morris method can only compute indices up to second order and there are two indices for one parameter:  $\mu$  which assesses the overall influence of the factor on the output and  $\sigma$  which estimate the ensemble of the factor's higher order effect [Campolongo et al., 2007]. Though it has the advantage that it can get the coarse screening conclusion with a lower cost of model evaluations, the two indices for one parameter make it not so convenient to draw overall conclusions. Since the computing efficiency problem of Sobol's method got improved the estimator and hopefully by parallel computing, the Sobol's method remains the most interesting for us. That is the reason why we tried to devise a strategy to make full use of its advantage like  $ST_i$ , that can help to avoid higher order index computing and to get the interaction information for parameters. If we can get over the problem of time-consuming model evaluation, the real convenience of Sobol's method especially for quantitative analysis is obvious. In some situation, it might however be necessary to resort to Morris method.

What we have devised here is our practice for FSPM. It would be interesting to study how this method can be generalized to other types of complex systems, like industrial ones [Zio, 2009].

# 6. SOFTWARE PLATFORM: PYGMALION

## 6.1 Background and objectives

In the plant growth modelling community, many models have been developed following different formalisms to understand plant growth and physiology. Several plant modelling platforms have been proposed with the objective to ease the development of methods, with bricks (organogenesis, radiosity module to compute light interception, graphical tools, etc.) that can be combined.

- L-studio [Federl and Prusinkiewicz, 1999] and GroIMP [Hemmerling et al., 2008] offer integrated environment to develop new models: based on sophisticated formalisms and providing the 3D-geometric classes for modelling and visualization.
- OpenAlea [Pradal et al., 2008] is an open source project for plant research. It is developed with Python libraries and tools in plant architecture modelling. OpenAlea includes modules to analyse, visualize and model the functioning and growth of plant architecture.

These platforms allow the simulation of models but do not provide enough statistical tools for their analysis and evaluation, though GroIMP does provide a range of basic statistical tools applicable to simulated and imported measured structures still it does not have the tools for sensitivity analysis. On the other hand, there are also a lot of excellent platforms for sensitivity analysis of generic models like R package FME [Team, 2009], Open TURNS [Andrianov et al., 2007], SimLab [Saltelli et al., 2004], etc.

The R package ‘Sensitivity’ implements sensitivity analysis methods: linear and monotonic sensitivity analysis (SRC, PCC, SRRC, PRCC), the screening method of Morris, and non-linear global sensitivity analysis (the Sobol indices, the FAST method). The functions of this package generate the design of experiments (depending on the method of analysis) and compute the sensitivity indices based on the model inputs and outputs. All sensitivity indices can be estimated with the bootstrap technique which allows to estimate the bias, and basic bootstrap confidence intervals. Text and graphical outputs display the results of the analysis. <sup>1</sup>.

---

<sup>1</sup> <http://cran.r-project.org/web/packages/sensitivity/index.html>

Open TURNS <sup>2</sup> is a Unix/Linux software, for Treatment of Uncertainty, Risk'N Statistics in a structured industrial approach. It has three main components :a scientific C++ library including an internal data model and algorithms dedicated to the treatment of uncertainty; an independent application with a graphical user interface; a python module with high level operators in the probabilistic and statistical field. The targeted users are here research centers and the academic community.

SimLab <sup>3</sup> is a didactical software designed for Monte Carlo based uncertainty and sensitivity analysis. It is a professional tool for model developers to learn and use global uncertainty and sensitivity analysis techniques. Especially SimLab 3 supports a set of coding environments such as C, C++, Matlab and Fortran.

The softwares for plant growth mostly focus on the implementation and development of plant growth modelling, but rarely on tools for model evaluation, especially for sensitivity analysis. The mentioned platforms for sensitivity analysis are not specifically designed for plant or biological models.

Digiplant [Cournède et al., 2006; Cournède et al., 2011] is a first step towards integrating the plant growth models development and model evaluation with parametric identification on real experimental data. It is a quite flexible and efficient tool. It has been used for parameter estimation of maize, sugar beet, sunflower, tomato, grapevine, rice, etc. It provides powerful estimation tools, but has strong limitations since it is restricted to the GreenLab model.

An objective of the Digiplante team is to develop models of plant growth in interaction with the environment, to improve their predictive capacity, and to compare different models. Thus, it is of great interest to integrate the models within a general platform providing the evaluation and analysis tools. On the platform, the users develop models in the specific language grammar in order to plug them into the simulation kernel: all mathematical tools relating to model design like parameter estimation, sensitivity and uncertainty analysis can be shared for every model. PyGMAIion (A platform for "Plant Growth Models' Analysis and Identification") is developed based on this objective.

## 6.2 PyGMAIion

We recall here the steps of model design as detailed in section.2.2:

1. The conceptual model (mental work): identifying a system, state variables, inputs; writing the different state equations involving model parameters

---

<sup>2</sup> <http://www.openturns.org/>

<sup>3</sup> Simlab (2011) Software package for uncertainty and sensitivity analysis. Joint Research Centre of the European Commission. Downloadable for free at: <http://simlab.jrc.ec.europa.eu>

2. Numerical implementation: giving values to inputs, parameters and running model simulations
3. Mathematical analysis: behavior, limits, stability; parametric sensitivity analysis
4. Parameter identification and model evaluation: ‘fitting’ experimental data; estimate the model parameters from experimental data; estimate uncertainty of the estimated parameters; predictive capacity
5. Model comparison and model selection: computing some information criteria to compare models

PyGMAIion is a C++ template based framework which embeds the routines for:

- Numerical implementation of models
- Parameter estimation
- Sensitivity and uncertainty analysis
- Model comparison and selection

Thus PyGMAIion is a platform that concerns most steps of the modelling process. For modelling methodology study, the different mathematic tools can be reused under the same implementation scheme for different models. From a methodology point of view, it can be figured out whether the algorithm can be applied to all types of models or must be adapted for some models with special characteristics. From a modelling point of view, the integration of the same methodology with different models makes it possible to compare their performances in order to evaluate how well the models can achieve our modelling aims like predictive capacity, or the robustness.

### 6.3 Implementation in PyGMAIion

For plant growth models, we often describe the model as discrete dynamic systems or with discretized processes of a continuous dynamic system:

$$X_{n+1} = F_n(X_n, U_n, P) \quad (6.1)$$

In which:  $X_n$  is the vector of state variables (e.g. in GreenLab it stands for masses of plant components),  $F_n$  gives the biophysical laws (e.g. the biomass production and distribution of the organs, the photosynthesis formalism, etc.);  $P$  stands for the model parameters which are often of genetic origin;  $U_n$  is the control variables, which are often environmental conditions like soil water and nitrogen contents, temperature, etc.

In PyGMAIion, a model is implemented as a Class Template Reference:

```
PGMA::Model<StateClass, ParameterClass, EnvironmentClass>
```

which means that to define a model, you “simply” need:

- to define the 3 classes StateClass, ParameterClass, EnvironmentClass
- to implement a state function describing how to compute  $X(n+1)$  from  $X(n)$  as in eqn.6.1

We provide a simple example for an equation of the Lotka-Volterra family to demonstrate how to implement a model in PyGMAIion. The Lotka-Volterra equations, also known as the predator-prey equations, are a pair of first order, non-linear, differential equations frequently used to describe the dynamics of biological systems in which two species interact, a predator and its prey. They evolve in time according to the pair of equations:

$$X(n+1) = X(n) + (aX(n) - bX(n)Y(n))E \quad (6.2)$$

$$Y(n+1) = Y(n) - (cY(n) + dX(n)Y(n))E \quad (6.3)$$

where  $X$  is population of preys,  $Y$  is the population of predators,  $n$  is the time step,  $a, b, c, d$  are the parameters representing the interaction of the two species, and  $E$  is ecological environmental factor.

StateClass:

```
struct LVState
{
    double X;
    double Y;
    LVState()
    {
        X = 1.0;
        Y = 1.0;
    }
}
```

ParameterClass:

```
struct LVParameters
{
    double a;
    double b;
    double c;
    double d;

    LVParameters()
```

```

    {
        a = 0.1;
        b = 0.02;
        c = 0.15;
        d = 0.03;
    }
}

```

EnvironmentClass:

```

struct LVEnvironment
{
    double E;

    LVEnvironment()
    {
        E = 1.0;
    }
}

```

State function (the body of model):

```

class LVModel: public PGMA::Model<LVState, LVParameters,
    LVEnvironment>
{
    LVModel();
    LVState * nextState(PGMA::StateList<LVState> & list,
        const LVParameters & p, const LVEnvironment &env)
    {
        const LVState & Xn(list.last());
        LVState * Xnplus1 = new LVState;

        Xnplus1->X = (p.a+1)*Xn.X*env.E - p.b*Xn.X*Xn.Y*
            env.E;
        Xnplus1->Y = (-p.c+1)*Xn.Y*env.E + p.d*Xn.X*Xn.Y*
            env.E;
        return Xnplus1;
    };
};

```

Beside State, Parameters, Environment, and Model, there exist two other fundamental objects: Observer and ObservationList. Observer is a mechanism to save values or function of State variables during a Simulation. Observer will only be triggered thanks to a Timeline. ObservationList can be understood as a Database in which we will save these informations.

An Observer needs a State to Observe and an Observation to save the values through `observeFnc()`.

```
class MyObserver: public pgma::Observer<MyState,
    MyParameters>
{
    MyObserver(): pgma::Observer<MyState, MyParameters>("
        MyObserver") {}

    void observeFnc(const MyState & state, const
        MyParameters & parameters, Observation & obs)
    {
        obs.add(state.X, "X");
        obs.add(state.Y, "Y");
    }
};
```

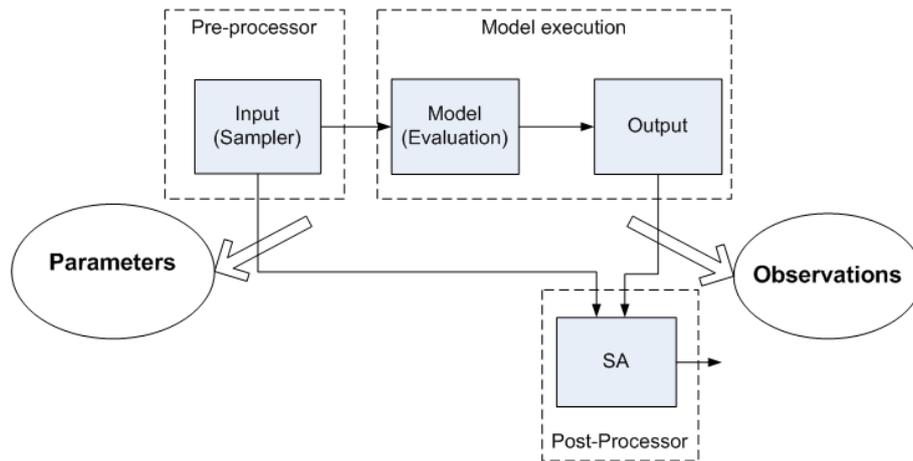
## 6.4 Sensitivity analysis in PyGMAIion

As we know, sensitivity analysis is processed in several steps:

- In the first step, we select an appropriate distribution for the input factors. These selected distributions will be used in the next step in samplers to generate random values for the input factors.
- In the second step, a set of points is generated in the parameter space obeying the distributions of the inputs specified in the first step with the samplers.
- In the third step, the model is evaluated at the sampled points and a set of model outputs are produced. In essence, these model evaluations create a mapping from the space of inputs to the space of outputs. This mapping is the basis for subsequent uncertainty and sensitivity analysis.
- In the fourth step, the results of model evaluations are used as the basis for uncertainty analysis. One way to characterize the uncertainty is with a mean value and a variance. Other model output statistics are provided.
- In the fifth step, the results of model evaluations are used as the basis for sensitivity analysis.

Following the logical configurations of the SA practice steps and the PyGMAIion ‘observation’ function for saving state variables, we developed a scheme of sensitivity analysis in PyGMAIion as shown in fig.6.1.

The SA processing in PyGMAIion is composed of three modules. These modules cover all the steps summarized above in the SA steps, and are inspired by the framework proposed in SimLab.



**Fig. 6.1** : PyGMAIion: Sensitivity analysis configuration

- The Pre-Processor module mainly works with samplers. It is in charge of the first and second steps. Most of the samplers come from Boost <sup>4</sup>, the standard C++ library. But when special samplers are needed like Morris sampling, we implement the samplers within PyGMAIion and add them to the function library.
- The Model Execution module accomplishes the third step with the simulation kernel of PyGMAIion.
- The Post-Processor module carries out the fourth and fifth steps. Different sensitivity index estimators are implemented in this module. Notice that both inputs and outputs are saved in a data file: the input sampling point are saved as parameter files, and the outputs are saved in observation list files for the chosen observation variables of the model as a function of time. So that for the users, the change of the analysis objective regarding the output of interest only leads to one command of observation function. It is not necessary to obtain the model evaluations again.

So far the SA module in PyGMAIion contains SRC, Sobol's method and Morris method to fulfill the SA requirements of models with different complexity. The SRC function generate the SRC index of each model factor and the non-linearity assessment of the model by giving the determination coefficient  $R^2$ . The Morris method is for the qualitative analysis to screen factors. As for the Sobol's method, we have implemented the classical Sobol's method [Sobol, 1993], the H-S Sobol's method [Homma and Saltelli, 1996] [Saltelli, 2002] and the new Sobol's proposed [Wu et al., 2011] in this thesis. Test cases of the methods on Ishigami function and Lotka-Volterra model are also given as examples for users. Graphical outputs are also produced to illustrate the results of the analysis.

<sup>4</sup> <http://www.boost.org/>



## **Part III**

# **APPLICATIONS AND SIMULATIONS**



# 7. METHODOLOGY PRACTICE: APPLICATION TO FSPMS

In the previous part, we discussed two important issues of SA for FSPMs:

- The computation issue for which we proposed an effective computational method to evaluate Sobol's indices and deduced the Monte Carlo simulation error estimation to a good control of the estimation accuracy.
- The strategy design for which we discussed the non-linearity assessment from SRC method and 'module by module analysis' that analyzes the model by functional modules and comprehensively use diverse sensitivity methods to combine different indices to avoid unnecessary high order index computing but still get the decomposition of the portions of higher order interaction.

In this chapter, we will present the result of the application of these methodologies to the three FSPMs with different levels of complexity introduced in section.2.4, and infer in each case what information can be drawn from this analysis.

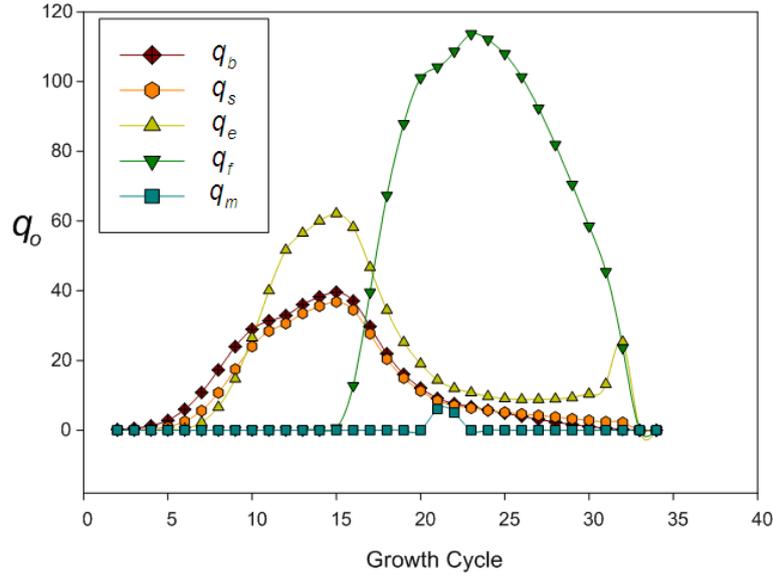
## 7.1 GreenLab maize

We first practice local SA method and its normalized version, then SRC, lastly Sobol's method. Our objective is to study the interest of global sensitivity analysis and its last developments for the GreenLab model, and more generally for a better understanding of source-sink dynamics and internal driving forces during plant growth.

### 7.1.1 Local sensitivity analysis and its normalized version

The time evolution of the biomass allocated to each type of organ  $q_o(i)$  (with  $o = b$ : leaf blade;  $o = s$ : sheath;  $o = e$ : internode;  $o = f$ : cob;  $o = m$ : tassel) for the mean values of uncertain input factors is given in fig.7.1.

We perform the local sensitivity analysis and its normalized version (presented in eqn.3.4 and eqn.3.5) respectively to the output variables  $q(i)$  (biomass production at growth cycle  $i$ ). The results are shown in fig.7.2.



**Fig. 7.1** : GreenLab Maize: biomass allocated to the different types of organs at each GC

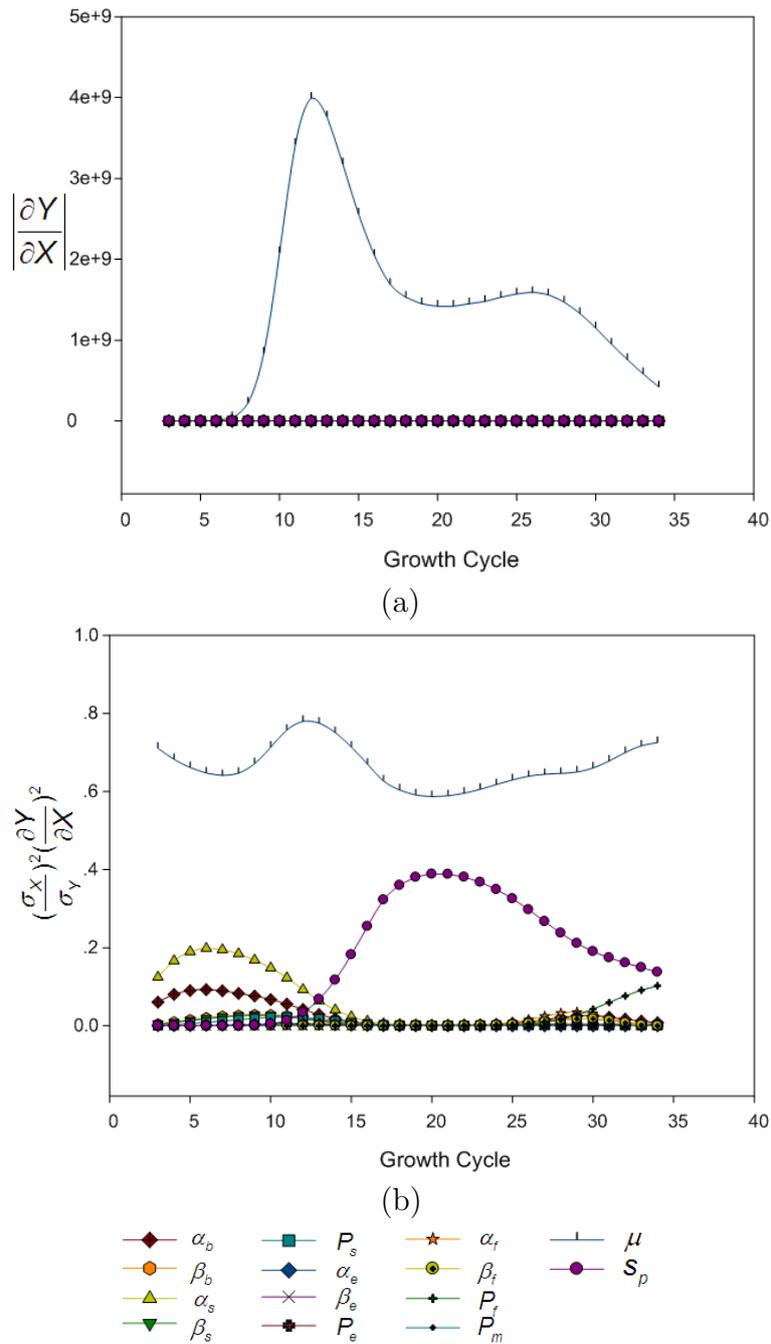
We used numerical simulation to get the derivatives of output  $q(i)$ . There are two ways to compute the standard deviations  $\sigma_y$ : the first one is through Monte Carlo simulation, which we used and is more reliable but with heavy computing cost. To avoid this shortcoming, another possibility is to use the following approximation:

$$(\sigma_y)^2 \cong \sum_i (\sigma_{x_i})^2 \left( \frac{\partial Y}{\partial X_i} \right)^2 \quad (7.1)$$

Such an approximation is justified when the system is highly linear, which we will evaluate in the following section.

In fig.7.2, we see the advantage of the normalized version. The pure local SA only reveals the importance of  $\mu$ . As for  $Sp$ , we can not even pick it out from other factors in the pure local analysis, but with the normalized one, the great contribution of  $Sp$  is more clearly demonstrated from about  $GC$  12 to the end of plant growth. The basic knowledge we get from this local analysis is that the factor  $\mu$  contributes the most to the variance of biomass generation from the beginning to the end.

Note that the point at which we calculate the derivatives is important, it could be not reliable for systems with parameter-to-output curve with many apices: the values at these special points are not representative of the whole information of sensitivity for this parameter. The results could be misleading, and we may miss the important details of the system. Therefore more reliable methods will be applied in the next section.



**Fig. 7.2 :** (a). Pure local sensitivity of  $q$  with respect to the factors listed in table.2.1 as functions of the growth cycle (absolute values). (b). Normalized local sensitivity of  $q^t$  with respect to the factors listed in table.2.1 (absolute values).  $Y$  stands for  $q(i)$  and  $X$  for the respective factors.

## 7.1.2 Standardized Regression coefficients

We perform a Standardized Regression method (see section.3.5.2) for the biomass production  $q$  across time with respect to all the parameters usually obtained by estimation

from experimental data in GreenLab. The input factors  $X_i$  being the parameters listed in table.2.1 and the output function  $Y$  being the biomass production at a given GC, we make a linear regression according to eqn.3.6 and eqn.3.7.

The SRC indices for each parameter evolving with time are shown in fig.7.3. The sum of SRC indices for all the parameters at each time step  $\sum(\gamma_{x_i})^2$  and model determination coefficient  $R^2$  are shown in fig.7.4. In Table.7.1, we present the SRC indices for all the parameters and  $R^2$  at the classical time steps which stand for different growth stage for maize.

Note that the SRCs results in fig.7.3 are really close to the squared normalized local measures, which is not a coincidence: for the linear system, the two should be equal. Fig.7.4 shows a very high  $R^2$ , (most of the time above 0.98) proving that the SRCs indices are reliable here. The regression analysis could be used as a preliminary step in SA to save computing time before we really do some more detailed analysis.

The crucial importance of the Radiation Use Efficiency  $\mu$  is demonstrated (SRC always above 0.6). It shows that a special care should be taken in its determination, including complementary experiments. The change of light interception can bring great variance for biomass production.  $Sp$  is the parameter that relates to competition between plants: at the beginning when blade and sheath are forming and the competition for the light has not start yet, the density does not bring so much variance. But at the 11<sup>th</sup> GC when the whole plant is grown and the fruit (cob for maize) appears, competition becomes dramatical and keeps increasing to the end. Allocation to blades (characterized by  $\alpha_b$  and  $\beta_b$ ) keeps a relative high importance, in the contrast  $P_m$  keeps low. Note that for  $P_f$  which relates to the cob sink strength, the ranking raises at the last stage of the whole plant growth.

**Tab. 7.1 :** GreenLab maize: SRC index of all the parameters and  $R^2$  at selected typical GC that can stand for growth stages

$S_i$ of $q_t$	1 <sup>th</sup>	5 <sup>th</sup>	11 <sup>th</sup>	19 <sup>th</sup>	30 <sup>th</sup>
$\alpha_b$	0.0797	0.0928	0.0335	0.0017	0.0205
$\beta_b$	0.0062	0.0251	0.0156	0.0015	0.0117
$\alpha_s$	0.1571	0.1997	0.0736	0.0013	0.0001
$\beta_s$	0.0040	0.0220	0.0135	0.0002	0.0000
$P_s$	0.0011	0.0118	0.0178	0.0010	0.0018
$\alpha_e$	0.0000	0.0015	0.0039	0.0000	0.0000
$\beta_e$	0.0000	0.0001	0.0009	0.0001	0.0000
$P_e$	0.0000	0.0001	0.0041	0.0017	0.0044
$\alpha_f$	0.0000	0.0000	0.0000	0.0004	0.0115
$\beta_f$	0.0000	0.0000	0.0000	0.0003	0.0095
$P_f$	0.0000	0.0000	0.0000	0.0003	0.0927
$P_m$	0.0000	0.0000	0.0000	0.0000	0.0000
$\mu$	0.6972	0.6315	0.8150	0.9778	0.8463
$R^2$	0.9453	0.9845	0.9778	0.9863	0.9983

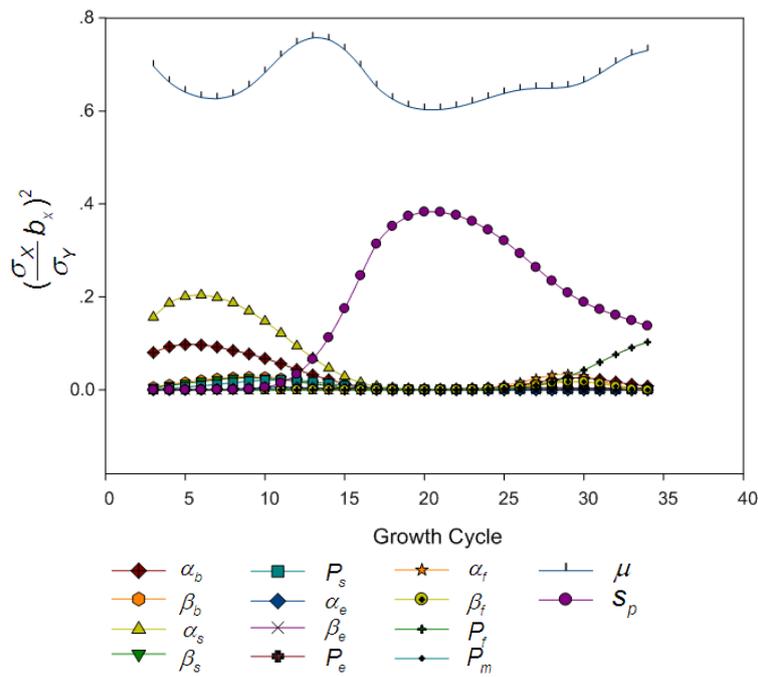


Fig. 7.3 : GreenLab maize:SRCs for the outputs  $q$  as functions of the growth cycle.

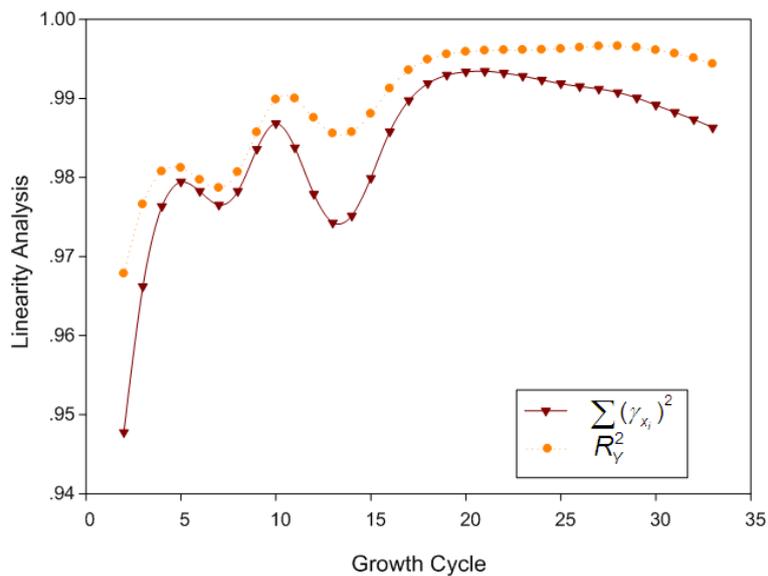
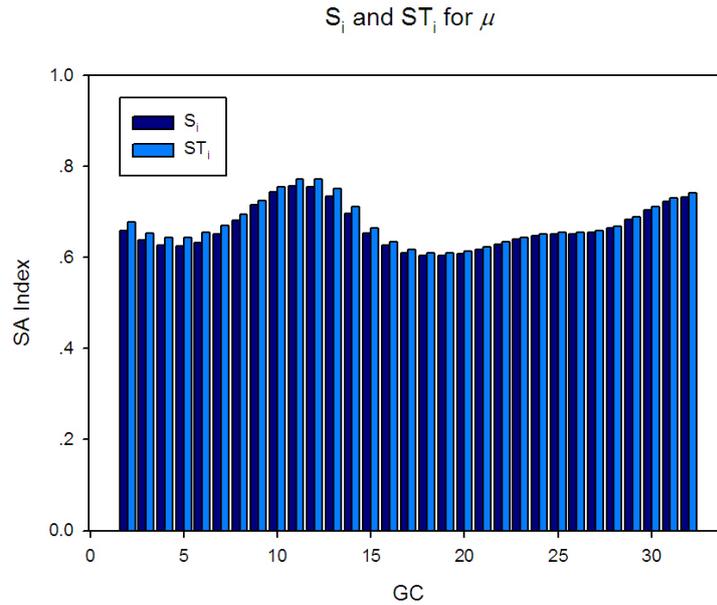


Fig. 7.4 : GreenLab maize:  $\sum(\gamma_{x_i})^2$  and  $R_Y^2$  for the output  $q$

### 7.1.3 Variance decomposition-based sensitivity measure: Sobol's method

The results of Sobol's method are similar to those of local and SRC methods due to the high linearity of the system: the parameters  $S_p$  and  $\mu$  take as much as 75%~98% of the sum for first-order index and the whole first-order sensitivity index contribute

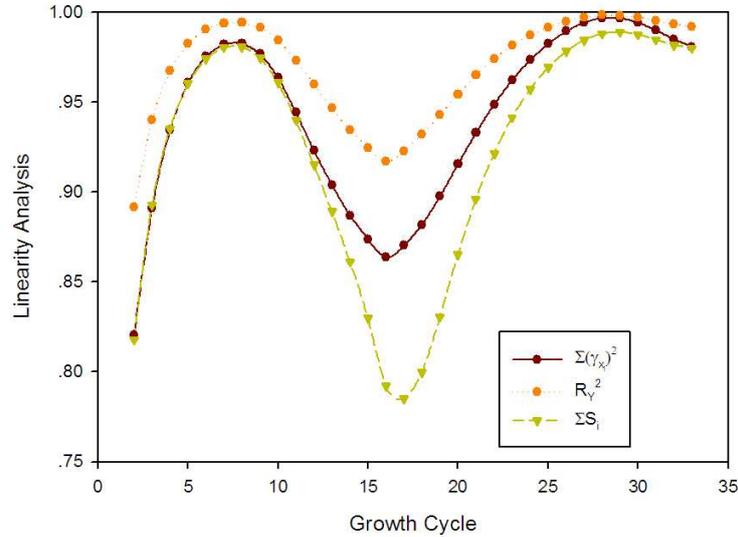


**Fig. 7.5 :** GreenLab model for maize:  $S_i$  and  $ST_i$  for parameter  $\mu$

to 93%~99% of the system variance. We computed Sobol's first order index  $S_i$  and total order index  $ST_i$ . The difference between the two indices can show whether the interaction of this parameter to others contributes significantly to the variance of the model output. From this point, we can still get that for the most important parameter  $\mu$ , the first order SA index can totally stand for its sensitivity to the model, which is shown in fig.7.5.

So as one alternative parameter space, we fixed  $Sp$  and  $\mu$  to their mean values. In fig.7.6, we show the two linearity indexes of the system, along with the summation of all  $S_i$ . We note that the trends of the three curves agree with each other. The summation of  $S_i$  is also characteristic of the system linearity. Actually it can be explained by the definition of  $S_i$ : the closer to 1 the sum of the first-order indexes is in eqn.3.16, the more mutually independent the factors are, and the more linear the system is. In the new parameter space without  $Sp$  and  $\mu$ , the system linearity is weaker, especially from 10th GC to 23th GC, which is the most non-linear stage of the growth, corresponding to abrupt changes in the allocation dynamics due to cob appearance. Since parametric estimation for GreenLab relies on multi-stage observations, it seems important not to miss this crucial period in the experimental data collection.

The first order Sobol's indices of the new set of parameters are illustrated in fig.7.7. Note that without the variance contributed by  $Sp$  and  $\mu$ ,  $\alpha_s$  and  $\alpha_b$  play the most important roles in the system during the first 20 GCs. Afterwards the sensitivity to  $\alpha_f$  becomes the biggest until the 30th GC, after which the sensitivity to  $P_f$  rises drastically while the sensitivities to all the other parameters drop. This trend can be explained by analyzing maize functioning. At the beginning of maize growth, before the fruit appears, the factors that control the competition for biomass acquisition must



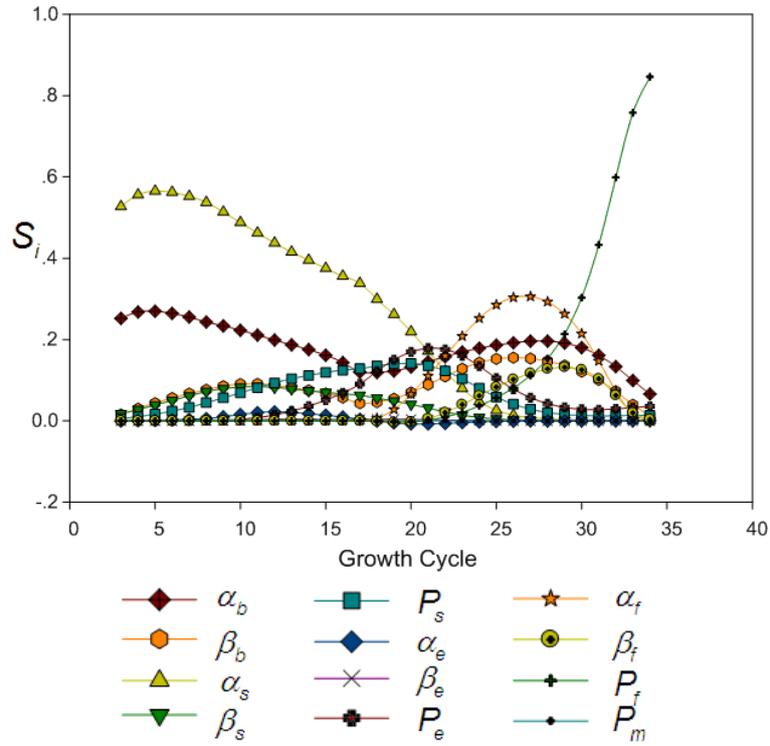
**Fig. 7.6 :** GreenLab maize: The three kinds of linearity indexes with parameter space of  $Sp$  and  $\mu$  fixed

play an important role, and the foliage is also privileged in the early stages of growth. It explains why the sensitivities to  $\alpha_s$  and  $\alpha_b$  are the most important ones. They correspond to the parameters driving the initial form of the sink variation functions of sheaths and blades. After the fruit (cob) appearance, it begins to attract biomass. The influence of  $\alpha_f$  (that is to say the parameter driving the initial form of the cob sink function) gets bigger. But its value remains relatively low because during this time period, the biomass demand of the other organs is still high. So the whole variance of the system output tends to be shared more uniformly during this stage between most parameters. Towards the end of the growth, the majority of biomass production is allocated to the cob and the sensitivity to  $P_f$  increases drastically.

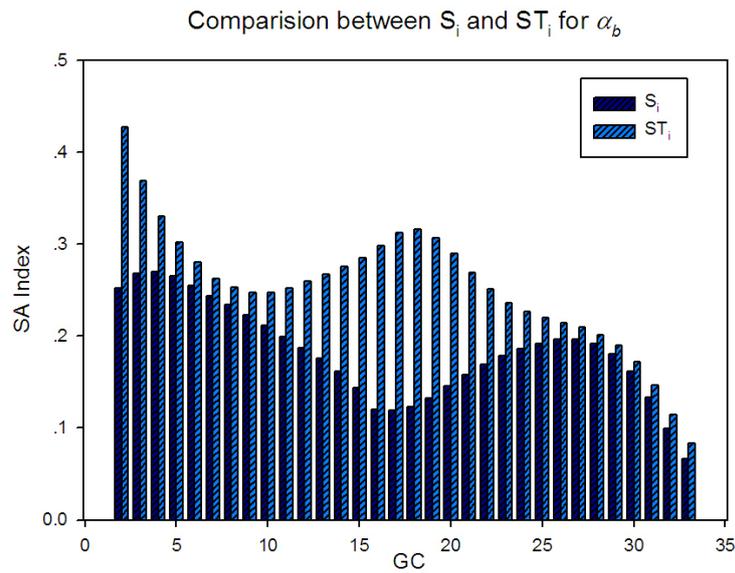
One thing we are more interested in is the interaction information that SA can bring us to know more about the source-sink dynamics and internal driving forces during plant growth. So  $S_i$  and  $ST_i$  are compared to see if interactions exist. Take  $\alpha_b$  for example in fig.7.8: we can see that the interaction for this parameter is active during the most dynamic growth duration: cob generation and internode expansion.

One more interesting point is that as we explore all the second-order Sobol's indices, we found that most of the interactions concentrate between  $\alpha_b$  and the others, in which the interaction between  $\alpha_b$  and  $\alpha_s$  outmatches by far the others, see fig.7.9.

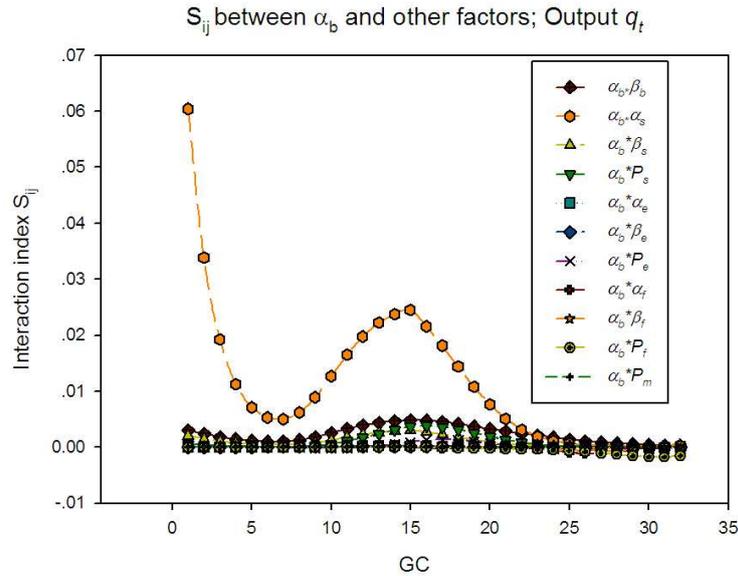
If we suppose that model parameterization can be linked to plant genes as illustrated by [Letort et al., 2008b], the understanding of parameter interaction may be interesting for genetic improvement. FSPM parameters may be linked to plant genes, and thus may help breeders to design ideotypes. If a parameter has little interaction with others (like  $P_f$ ), we can directly concentrate on this trait to improve plant per-



**Fig. 7.7 :** GreenLab maize: Sobol's method for SA with parameter space of  $Sp$  and  $\mu$  fixed



**Fig. 7.8 :** GreenLab maize: Difference between  $ST_i$  and  $S_i$  for  $\alpha_b$

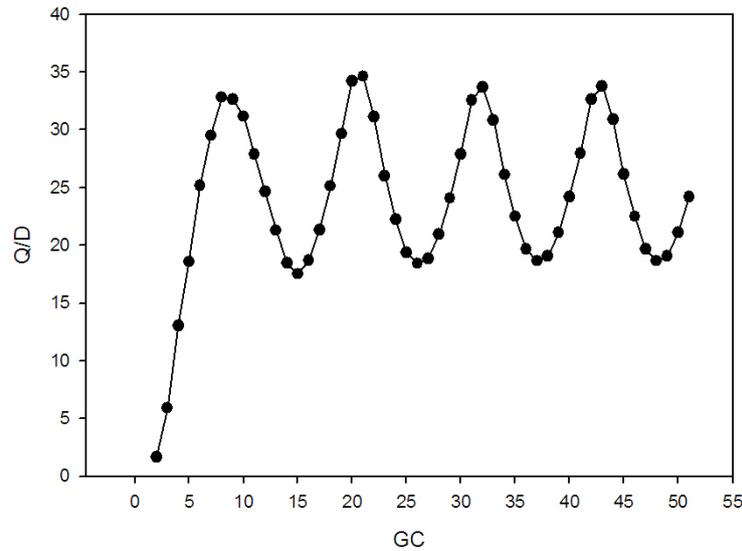


**Fig. 7.9** : GreenLab maize: Interactions between  $\alpha_b$  and the other parameters

formance. If the interaction is strong, like for  $\alpha_b$  (cf fig.7.9, the interaction between  $\alpha_b$  and  $\alpha_s$  being the most important), it is more complex. If the parameters are strongly genetically related (determination by the same genes, which for  $\alpha_b$ ,  $\alpha_s$  is probable since they are both characteristic of leaf geometry), the model parameterization should be improved to take into account this fundamental interaction. If they are not genetically related, breeding strategy should rely on multi-dimensional optimization to handle the interacting processes [Qi et al., 2010].

#### 7.1.4 Discussion

The most important parameters for plant growth model of maize are  $\mu$  (energetic efficiency) and  $Sp$  (characteristic surface related to competition between plants), which represent from 75% to 98% of the first-order sensitivity index. During the process of parameter estimation from experimental data, there is not usually a direct convergence to the proper set of parameters because of the non-convexity of the generalized least-square function used as fitting criteria. Moreover, it was shown that the confidence interval on the estimated parameters might be improved by fixing some parameters [Guo et al., 2006], [Lamboni et al., 2011]. The sensitivity analysis gives us hints on how to improve the calibration process for Maize: first fix all parameters to reasonable values from literature, then find estimates of  $\mu$  and  $Sp$ , then find simultaneously new estimates for the set of  $\alpha_o$  parameters together with  $\mu$  and  $Sp$ , and finally find simultaneously new estimates for the set of sink parameters  $P_o$ , together with the set of  $\alpha_o$  parameters and  $\mu$  and  $Sp$ . The sensitivity analysis indicates that fixing  $\beta_o$  parameters is reasonable since their influence is limited.

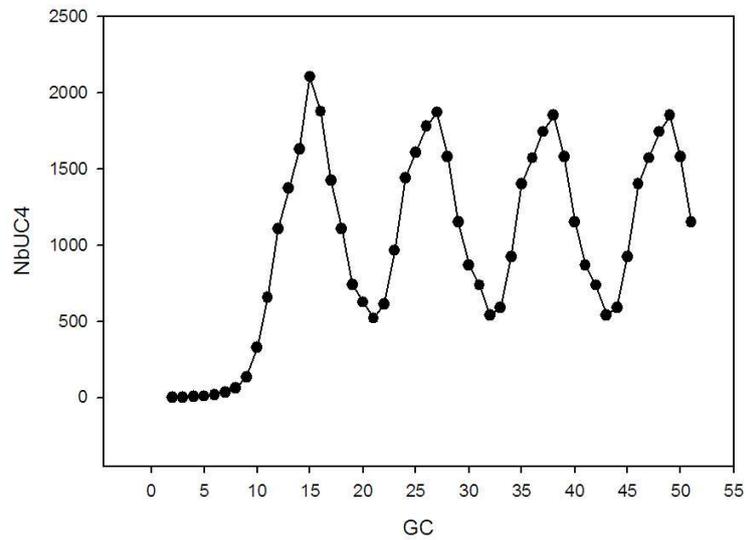


**Fig. 7.10 :** Poplar tree: Oscillation of the ratio of available biomass to organs' demand ( $Q/D$ ). First the ratio increases due to an increase in biomass production at a constant level of demand. When its value exceeds a threshold, new organs appear, which leads to an increase in plant demand and hence a decrease of the ratio. The ratio falls below the threshold, reducing the appearance of new organs, and a decrease in the demand and so on.

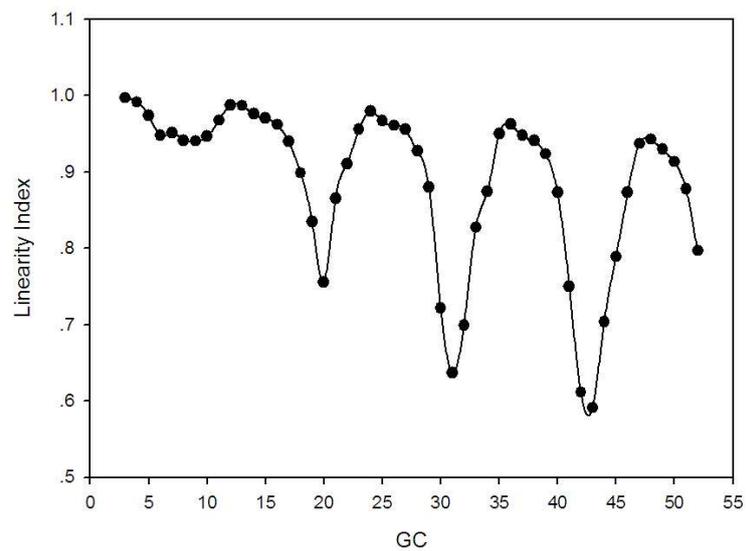
## 7.2 Poplar tree model with retroaction of functioning on organogenesis

In fig.7.10, is shown the evolution of the ratio of biomass production to organs' demand, which is the key variable controlling tree organogenesis in the GreenLab model of tree growth. The sensitivity analysis method was applied to the tree annual biomass production computed with eqn.2.7. As mentioned in [Wu and Cournède, 2009], linearity index is firstly checked before starting global SA. The linearity index is shown in fig.7.12. The initial states show the high linearity indices, but it decreases rapidly with the increasing influence of the trophic competition on organogenesis, particularly the appearance of the first branch at growth cycle 5. At this stage, the biomass production is low, because of the small number of leaves (fig. 7.11) induced by a lower ratio of biomass to demand(fig.7.10). The sensitivity index of the model for parameter  $\mu$  follows the pattern of biomass production (fig.7.13).

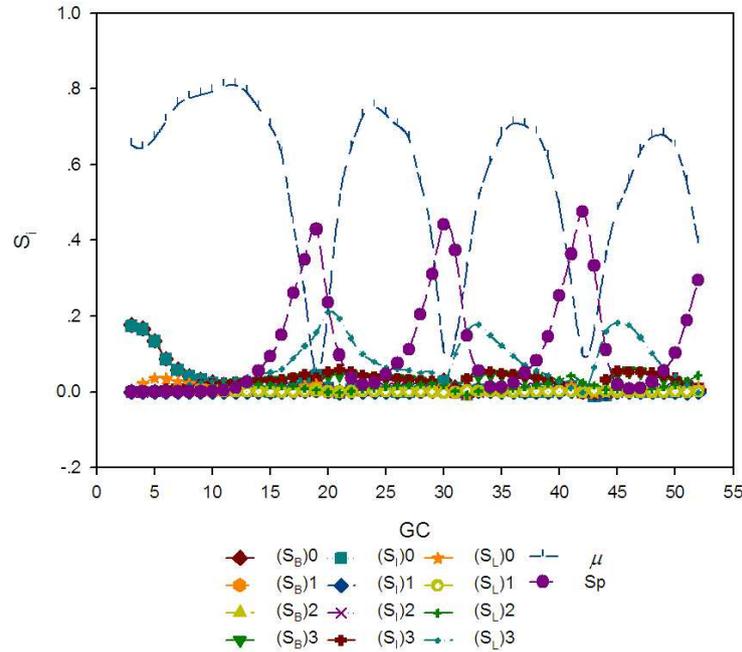
Due to the exponential negative function, the biomass production is very sensitive to changes in total leaf surface area when this one is small, but on the contrary, there is a value beyond which an increase in leaf surface area will induce a very little increase in biomass production. As the leaf surface area depends on the number of organs, the biomass production alternates between the phase close to saturation and the linear phase. In the linear phase, the model is more sensitive to the parameter  $S_p$ . On the



**Fig. 7.11** : Poplar tree: Number of phytomers of physiological age 4. Their number depends on the ratio of available biomass to demand.



**Fig. 7.12** : Poplar tree: Evolution of linearity index

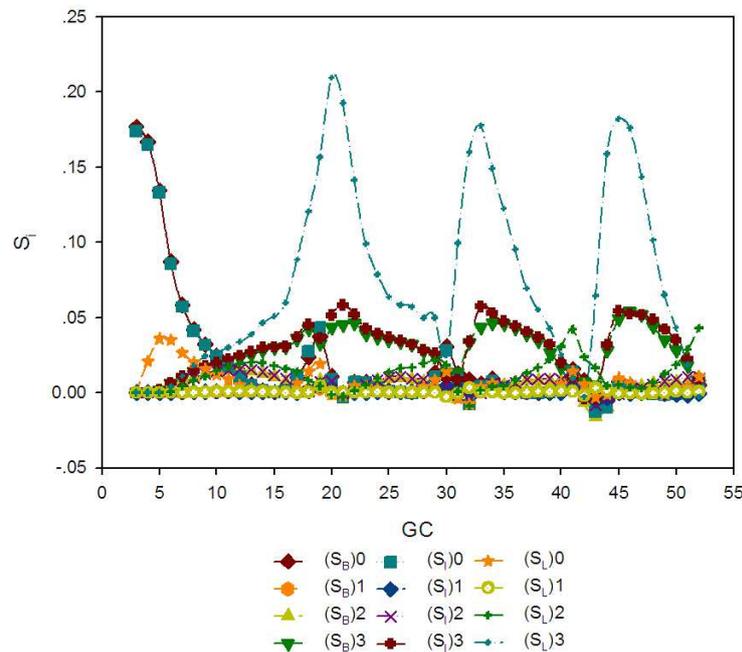


**Fig. 7.13 :** Poplar tree: Sensitivity indexes for all the model parameters

contrary, in the saturation phase, the model is more sensitive to the parameter  $\mu$  as illustrated by fig.7.13.

These two parameters are the most important ones regarding biomass production. As in the study of GreenLab maize, we exclude these two parameters to get closer insight in the other parameters driving biomass allocation - sink strengths - that impact plant production through the computation of leaf surface area (eqn.2.8), as shown in fig.7.14. During the youth of the tree, the trunk starts growing and no branch appears due to the low value of the ratio of biomass to demand. At this point, the model is mainly sensitive to the parameters of phytomers of physiological age 1 ( $(S_B)0$  and  $(S_I)0$  in fig.7.14). Then, the ratio of available biomass to demand increases fast, and several branches appear together. The phytomers of physiological age 4 are the most numerous in the tree as they correspond to the twigs. Their number increases till time step 15 and then oscillates (fig.7.11), with a period corresponding to that of the ratio of biomass to demand (fig.7.10). Hence, the model output is sensitive to their sink strengths ( $(S_B)3$ ,  $(S_L)3$  and  $(S_I)3$  on fig.7.14) with the same period, in the phases corresponding to high levels of  $Q/D$ , when a large number of twigs will appear. In the phase of low levels of  $Q/D$ , the most important parameter is the layer sink for secondary growth since it corresponds to the largest part of plant demand when primary growth is restricted.

We inspected 14 parameters related to tree biomass production and allocation, and follow dynamically the sensitivity indices for 50 years. Note that the computation of



**Fig. 7.14 :** Poplar tree: Sensitivity indices when fixing  $\mu$  and  $S_p$

such a long period of growth remains quite reasonable for the GreenLab model thanks to the structural factorization described in [Cournède et al., 2006]. The test case chosen corresponds to a specific model showing alternating patterns in growth phases resulting from the complex interactions between functioning and organogenesis. The sensitivity analysis offered interesting insight in the understanding of this interaction. Moreover, by using variance-based techniques, an analyst is capable not only of obtaining the parameter contribution to the output variance but also of gaining insights on the model structure by using moment-independent indicators [Borgonovo, 2006]. Such method provides insights on the influence of the input uncertainty on the output distribution [Borgonovo, 2007]. So it would be interesting for our future work to gain insight about the output distribution (in our tree modelling case, for the variables describing tree height and diameter for example). It should open a new way to deeper studies on more complex functional-structural plant models.

### 7.3 Model of C-N dynamics (NEMA)

Basic biological modules are identified: namely in our test case Carbon distribution (DMflux), Nitrogen distribution (Nflux), Carbon acquisition via photosynthesis (Photosynthesis), Nitrogen acquisition by roots (RootNuptake), Senescence (TissueDeath).

Several outputs of interest are considered for both intra-module and inter-module analysis: a) total green area of the plant (AreaGreenTotal), b) total dry mass production of the plant (Production), c) dry mass of the grains (DMgrains), d) nitrogen

mass of the grains (Ngrains) and e) root nitrogen uptake (RootNuptake). Moreover, the evolution of the sensitivity indices are computed for a better investigation of the dynamics of plant growth. When we aim at screening parameters, we consider the TGI mentioned in section.5.4. The full model involves around 80 parameters. Full model description can be found in [Bertheloot et al., 2011a]. We used the parameterization of [Bertheloot et al., 2011b].

We use subscriptions to identify the factors for different plant organs as follows: g for Grain, r for Root, La for Lamina, Sh for Sheath, In for Internode, Pe for Peduncle, Ch for Chaff.

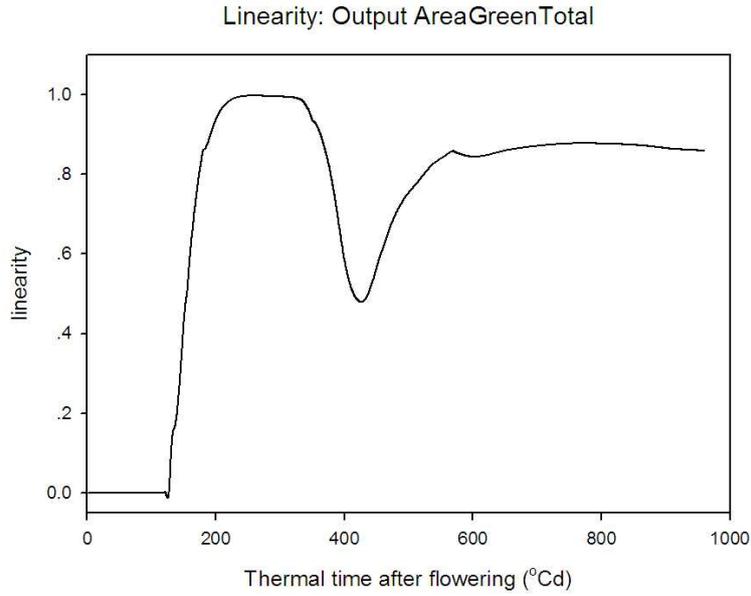
As shown in tab.7.2, we present the result summary of our analysis following the strategy design mentioned in section.5.3. We will present the result according to the different outputs first, then we will make a comparison to see the common points and differences. Some steps are skipped due to the intermediate conclusions reached, for example, by the linearity analysis at the first step, we may be able to skip the interaction analysis, either for the group analysis or for the overall model analysis based on selected parameters.

**Tab. 7.2 :** NEMA: List of tables and figure describing the results of SA. Five main outputs considered, a) AreaGreenTotal, b) Production, c) DMgrains, d) Ngrains and e) RootNuptake, so each output was checked separately. The mark ‘-’ means the corresponding step was skipped for the output.

		a	b	c	d	e
Linearity		fig.7.15	fig.7.21	fig.7.27	fig.7.30	fig.7.36
Group	$S_{\Omega_i}^g$	fig.7.16	fig.7.22	fig.7.28	fig.7.31	fig.7.37
	$ST_{\Omega_i}^g - S_{\Omega_i}^g$	fig.7.17	fig.7.23	-	fig.7.32	fig.7.38
	$S_{\Omega_{ij}}^g$	fig.7.18	fig.7.24	-	fig.7.33	fig.7.39
Intra module	$TGIS_i, TGIST_i$	tab.7.3,7.4,7.5, 7.6,7.7,7.8	tab.7.12,7.13	tab.7.14,7.15	tab.7.16,7.17	tab.7.21,7.22
	$TGIS_{\Omega_i}^g - \sum TGIS_i$	tab.7.9	-	-	tab.7.18	tab.7.23
Inter module	$TGIST_{\Omega_i}^g - TGIS_{\Omega_i}^g$	tab.7.10	-	-	tab.7.19	tab.7.24
Overall model	$S_i$	fig.7.19	fig.7.25	fig.7.29	fig.7.34	fig.7.40
	$ST_i - S_i$	fig.7.20	fig.7.26	-	fig.7.35	fig.7.41
	$S_{ij}$	tab.7.11	-	-	tab.7.20	-

### 7.3.1 Output AreaGreenTotal

As we mentioned in section.5.3, we first applied SRC to check the linearity information evolving with time as in fig.7.15. Between 374 °Cd and 531 °Cd, there is one valley for the linearity, during which the lowest linearity appears at 417 °Cd to the linearity of 0.48. Such a period in which a strong non-linearity occurs may be characteristic of very specific biological phenomena during plant growth and of high level of interactions



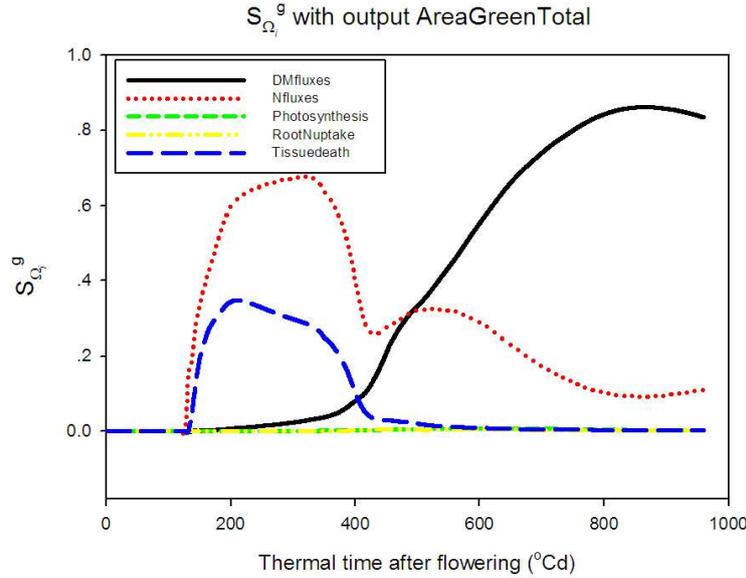
**Fig. 7.15** : NEMA: Linearity: Output AreaGreenTotal, the turn point of linearity starts at 374 °Cd and ends at 531 °Cd, the lowest linearity appears at 417 °Cd.

between parameters, either known by the modeler or unknown, in which case they should probably be investigated more.

We then check the result of group analysis for each module, as shown in fig.7.16, fig.7.17 and fig.7.18. Fig.7.16 shows the evolution of the first order group index  $S_{\Omega_i}^g$  and we observe a change in importance ranks of modules around 411 °Cd. For  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  in fig.7.17 and  $S_{\Omega_{ij}}^g$  in fig.7.18, which stand for the inter-module interactions, the peaks of most of the curves overlap also around 411 °Cd, which coincides with the lowest linearity of the model.

Our analysis points out this important stage and guides us for the parameter selection procedure in the module by module analysis part. After the parameter selection we can see that the lowest sum of first order  $S_i$  in fig.7.19 and highest value of  $ST_i - S_i$  in fig.7.20 still appear at the same stage. After the parameter selection step, it is important to check if the simplified analysis keeps the same attribute such as non-linearity period. Moreover, it also helps us focus on the  $S_{ij}$  matrix at this typical time point to know the parameter interaction attribute as shown in tab.7.11.

From fig.7.16, we can see that the first order sensitivity mainly goes to module DMfluxes and Nfluxes, and for short period of the module Tissuedeath. Fig.7.17 and fig.7.18 show that the interactions between modules mainly exist for DMfluxes and Nfluxes. We also noticed that for the output AreaGreenTotal, the module Photosynthesis and RootNuptake contribute little to the variance of the output from the beginning to the end. Moreover, the inter-module interactions  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  for the two modules are also very low at all stages. This information leads us to screen the parameters for output AreaGreenTotal. Thus we do not need to consider the uncertainty

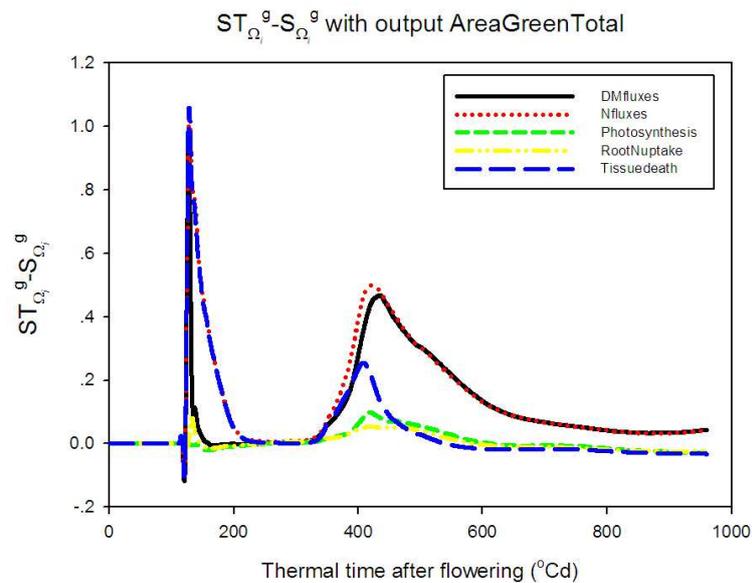


**Fig. 7.16 :** NEMA:  $S_{\Omega_i}^g$ : Output AreaGreenTotal, module Nfluxes and Tissuedeath dominate the first period between 131 °Cd and 411.15 °Cd, the turning point of the  $S_{\Omega_i}^g$  for the modules' ranking change is at 411.15 °Cd, afterwards, the importance of module DMfluxes increases dramatically, and on the contrary the indices of Nfluxes and Tissuedeath start to decrease. So module DMfluxes is the most important module after 411.15 °Cd. Module Photosynthesis and RootNuptake keep low sensitivity to system output AreaGreenTotal from the beginning to the end.

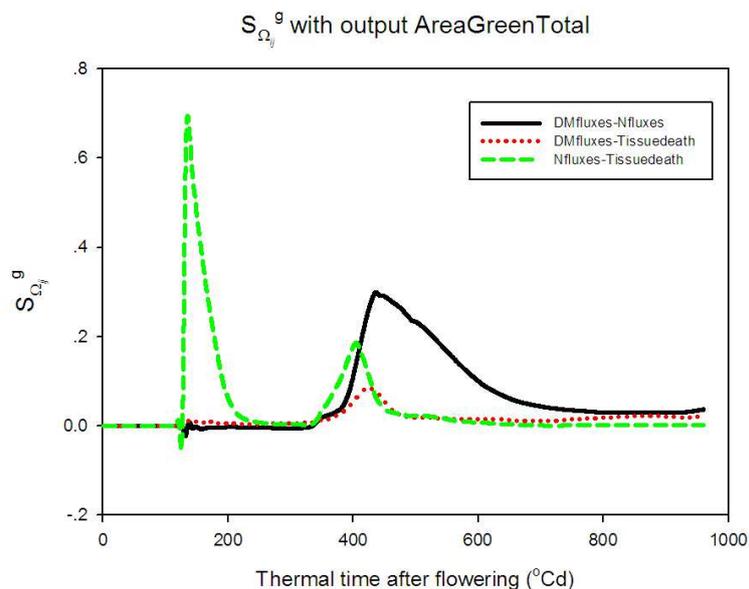
from these parameters while performing parameter estimation and the other model design steps that the number of parameters could be a bottleneck for processing.

As for the third step: internal module analysis, we list the TGI  $S_i$  and TGI  $ST_i$  of all the parameters module by module here to present the whole parameter selection procedure. The selected parameters from the 5 modules are marked with grey colour in the tables. From tab.7.3 to tab.7.7, all the parameters are listed by modules, and besides the TGI  $S_i$  and TGI  $ST_i$ , we give out the ranking of each index intra-module and the ranking of TGI  $ST_i$  in the overall model. The two types of intra-module ranking let us see the different positions of each parameter for different indices. And based on the same sampling points in the parameter space, the indices can be compared inter-module, so that the overall model ranking can be used here for the main standard of selection.

We have mentioned before in the group analysis for output AreaGreenTotal that the parameters belonging to modules Photosynthesis and RootNuptake may be screened because the  $S_{\Omega_i}^g$  and  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  are nearly nil all through the whole time period. It gets validated when we check the ranking of TGI  $ST_i$  for the overall model analysis. The best ranking in these two modules is for parameter  $\omega_{La,2}$ , whose index is about 0.6%. Moreover, the difference between TGI  $ST_i$  and TGI  $S_i$  is less than 0.2%, which can be neglected in terms of interactions. As such, we can confirm that we can screen



**Fig. 7.17** : NEMA:  $ST_{\Omega_i}^g - S_{\Omega_i}^g$ : Output AreaGreenTotal. The first peak of the curves appears at  $127^\circ\text{Cd}$ , but it only lasts for several discrete time points, and then disappears. We can consider that the appearance of the first peak comes from the adaptation stage from the zero output period to the beginning of response, so that the transition period is characterized by a lot of interactions but they don't last long. The second peak of the curves appears at  $411.15^\circ\text{Cd}$ , while module DMfluxes and Nfluxes share the same trend as the modules with the largest inter-module interactions.



**Fig. 7.18** : NEMA:  $S_{\Omega_{ij}}^g$ : Output AreaGreenTotal, the peaks of the three curves appear between  $400^\circ\text{Cd}$  and  $420^\circ\text{Cd}$ . The interaction between module DMfluxes and Nfluxes is the most important one and lasts for the longest time.

the parameters of these two modules without any risk.

We stopped the selection procedure at the 12th parameter  $d_{La}$ . It is related to the computation accuracy issue. If we check the indices after the TGI  $ST_i$  ranking 12th, the TGI  $ST_i < 0.65\%$ , TGI  $S_i < 0.48\%$  and TGI  $ST_i - S_i < 0.37\%$ . The indices on one hand are too little for us to consider, on the other hand, their values are close to the error of computing accuracy, therefore we only selected the first 12 parameter for the overall model analysis. In tab.7.8, we compare the effect of the analysis before and after the selection. We can see that after the selection, the abandoned main effect of all the parameters to the model is 1.9% which only takes about 2% of the overall main effect 87.3%. So we can be sure that the selected parameters can represent the whole model by most.

**Tab. 7.3 :** NEMA: Module:DMfluxes; Output:AreaGreenTotal

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
$\sigma_g^M$	0.012301	4	0.040917	5	7
$\alpha_g$	0.003365	7	0.006452	9	13
$\beta_g$	0.000367	11	0.011586	7	9
$tt_g^{Macc}$	0.008182	6	0.057255	3	4
$\sigma_r^M$	0.018347	3	0.043396	4	5
$\alpha_r$	0.012154	5	0.030498	6	8
$\beta_r$	0.061913	2	0.136818	2	2
$tt_r^{Macc}$	0.621337	1	0.746507	1	1
$\delta_r^M$	1.61E-05	32	0.000332	26	46
$\sigma_{La}^M$	4.99E-05	24	0.000615	22	34
$\alpha_{La}$	1.01E-05	34	8.09E-05	31	74
$\beta_{La}$	1.8E-05	31	0.002305	13	21
$tt_{La}^{Macc}$	9.32E-05	20	0.005109	11	16
$\delta_{La}^M$	4.28E-05	27	0.000899	19	30
$\sigma_{Sh}^M$	9.96E-05	19	0.000634	21	33
$\alpha_{Sh}$	3.86E-05	29	0.000151	30	68
$\beta_{Sh}$	0.000176	14	0.002018	15	25
$tt_{Sh}^{Macc}$	0.001505	8	0.003112	12	20
$\delta_{Sh}^M$	0.000159	16	0.000343	25	44
$\sigma_{In}^M$	0.000169	15	0.000784	20	31
$\alpha_{In}$	8.35E-05	22	0.000324	27	47
$\beta_{In}$	0.000152	17	0.001302	17	27
$tt_{In}^{Macc}$	0.000889	9	0.002198	14	22
$\delta_{In}^M$	0.000222	13	0.001416	16	26
$\sigma_{Pe}^M$	3.94E-05	28	3.97E-05	34	81
$\alpha_{Pe}$	2.21E-05	30	5.3E-05	33	77
$\beta_{Pe}$	4.84E-05	25	0.000316	28	50
$tt_{Pe}^{Macc}$	0.000352	12	0.005502	10	15
$\delta_{Pe}^M$	9.16E-05	21	0.001274	18	28
$\sigma_{Ch}^M$	4.31E-05	26	7.17E-05	32	75
$\alpha_{Ch}$	1.44E-05	33	0.000223	29	60
$\beta_{Ch}$	0.000109	18	0.000403	24	41
$tt_{Ch}^{Macc}$	0.000833	10	0.008561	8	11
$\delta_{Ch}^M$	5.75E-05	23	0.000511	23	37

As we have concluded in section.5.3 step 3 that the difference between  $TGIS_{\Omega_i}^g$  and  $\sum TGIS_i$  for each module stands for the intra-module interactions. And obviously, the difference between  $TGIST_{\Omega_i}^g$  and  $TGIS_{\Omega_i}^g$  means the inter-module interactions.

**Tab. 7.4** : NEMA: Module:Nfluxes;Output:AreaGreenTotal

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
$P_g$	0	28	0.000216	16	61
$\gamma$	0.08179	1	0.081467	1	3
$P_r$	-5.7E-10	29	0.000216	17	62
$\delta_r^N$	1.02E-07	24	0.00017	19	65
$P_{La}$	1.56E-08	27	0.000236	15	59
$\sigma_{La}^{Nph}$	0.009076	3	0.010294	3	10
$\delta_{La}^N$	0.023419	2	0.041367	2	6
$k_{La,1}$	0.000732	5	0.000516	5	36
$k_{La,2}$	0.000743	4	0.003654	4	17
$P_{Sh}$	1E-06	20	0.000157	20	66
$\sigma_{Sh}^{Nph}$	2.72E-05	11	3.61E-05	29	82
$\delta_{Sh}^N$	0.000106	8	0.000127	23	69
$k_{Sh,1}$	4.16E-06	14	0.000487	8	39
$k_{Sh,2}$	6.47E-06	13	6.3E-05	27	76
$P_{In}$	2.88E-06	16	0.000121	24	70
$\sigma_{In}^{Nph}$	3.47E-06	15	4.13E-05	28	80
$\delta_{In}^N$	2.2E-05	12	0.00034	11	45
$k_{In,1}$	6.26E-08	25	0.000195	18	63
$k_{In,2}$	6.1E-07	21	0.000245	14	58
$P_{Pe}$	5.63E-07	22	0.000358	9	43
$\sigma_{Pe}^{Nph}$	2.82E-05	10	0.0001	25	72
$\delta_{Pe}^N$	0.000182	7	0.00049	6	38
$k_{Pe,1}$	1.92E-06	18	0.000271	12	56
$k_{Pe,2}$	1.35E-06	19	0.000151	21	67
$P_{Ch}$	3.77E-08	26	0.000358	9	43
$\sigma_{Ch}^{Nph}$	8.89E-05	9	0.0001	25	72
$\delta_{Ch}^N$	0.000493	6	0.00049	6	38
$k_{Ch,1}$	1.99E-06	17	0.000271	12	56
$k_{Ch,2}$	2.26E-07	23	0.000151	21	67

**Tab. 7.5** : NEMA: Module:Photosynthesis;Output:AreaGreenTotal

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
$\epsilon_{La}$	0.000158	3	0.00027	8	57
$\omega_{La,2}$	0.004761	1	0.00623	1	14
$\epsilon_{Sh}$	5.73E-05	4	0.000322	5	49
$\omega_{Sh,2}$	3.21E-05	6	0.000378	3	42
$\epsilon_{In}$	8.14E-07	9	0.000294	6	51
$\omega_{In,2}$	3.86E-07	10	0.000178	9	64
$\epsilon_{Pe}$	3.25E-06	8	4.41E-05	10	79
$\omega_{Pe,2}$	3.83E-05	5	0.000324	4	48
$\epsilon_{Ch}$	2.72E-05	7	0.000436	2	40
$\omega_{Ch,2}$	0.000169	2	0.00028	7	52

**Tab. 7.6** : NEMA: Module:RootNuptake;Output:AreaGreenTotal

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
$U_{r,max}$	0.00134	2	0.002048	3	24
$k_{r,1}$	0.0003	3	0.002174	2	23
$k_{r,2}$	4.41E-05	4	0.000521	4	35
$\beta_N$	0.00169	1	0.003523	1	18
$\beta_C$	1.28E-05	5	0.000273	5	55

**Tab. 7.7** : Module:Tissueddeath;Output:AreaGreenTotal

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
$d_{La}$	0.004237	1	0.008199	1	12
$d_{Sh}$	2.36E-05	3	4.96E-05	4	78
$d_{In}$	5.01E-06	5	2.12E-05	5	83
$d_{Pe}$	2.31E-05	4	0.000276	3	54
$d_{Ch}$	4.89E-05	2	0.000648	2	32

**Tab. 7.8** : NEMA: Overall parameter selection analysis;Output:AreaGreenTotal

	$\sum TGIS_i$	$\sum SelectedTGIS_i$	Difference
DMfluxes	0.7433	0.735434	0.007866
Nfluxes	0.116733	0.114285	0.002448
Photosynthesis	0.005248	0	0.005248
RootNuptake	0.003386	0	0.003386
Tissuedeath	0.004338	0.004237	0.000101
Overall model	0.873005	0.853955	0.019049

Table.7.9 shows the decomposition for the intra-module part. Generally, most intra-module interactions exist in module DMfluxes, which is as high as 8.75% in TGI. And the total intra-module interactions for all the five modules is 8.79%. Table.7.10 shows the decomposition for the inter-module part. Corresponding to the conclusion we drew from fig.7.17, most of the inter-module interaction comes between module DMfluxes and Nfluxes.

**Tab. 7.9** : NEMA: Intra-Module Interaction analysis;Output:AreaGreenTotal

	$TGIS_{\Omega_i}^g$	$\sum TGIS_i$	Intra-Module Interaction
DMfluxes	0.830802	0.7433	0.087502
Nfluxes	0.117132	0.116733	0.000399
Photosynthesis	0.00531	0.005248	6.18E-05
RootNuptake	0.003411	0.003386	2.47E-05
Tissuedeath	0.004348	0.004338	1.04E-05
Overall model	0.961002	0.873005	0.087997

**Tab. 7.10** : NEMA: Inter-Module Interaction; Output:AreaGreenTotal

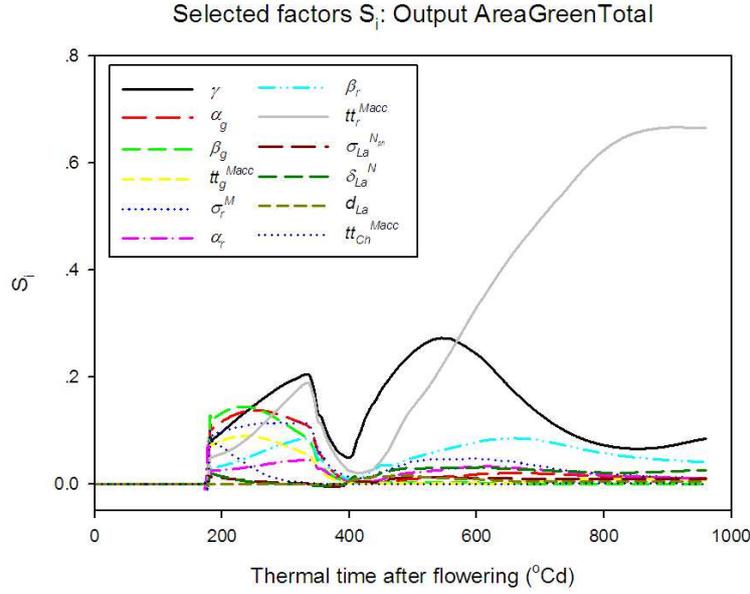
	$TGIS_{\Omega_i}^g$	$TGIST_{\Omega_i}^g$	Inter-Module Interaction
DMfluxes	0.830802	0.874785	0.043983
Nfluxes	0.117132	0.15515	0.038019
Photosynthesis	0.00531	0.005408	9.8E-05
RootNuptake	0.003411	0.003518	0.000108
Tissuedeath	0.004348	0.004679	0.000331

After we decided the set of selected parameters, we run the Sobol's method with the relatively small number of parameters to evaluate the overall parameter sensitivity. Fig.7.19 shows the  $S_i$  and fig.7.20 shows the  $ST_i - S_i$  for the selected parameters. Factor  $\gamma$  from module Nfluxes which stands for the relative rate of potential grain N filling during cell division and factor  $tt_r^{Macc}$  from module DMfluxes which stands for the duration during which roots can accumulate dry mass are the two most active factors for the model while output AreaGreenTotal is considered.

For the main effect  $S_i$ , these two factors share the same shape at the first stage before the time point 411.15 °Cd with the lowest sum of  $S_i$  or linearity. Afterwards  $\gamma$  reaches a peak at around 543 °Cd and then decrease little by little while  $tt_r^{Macc}$  keeps increasing quasi linearly and surpasses  $\gamma$  at around the peak point of  $\gamma$ .

For the interaction distribution presented by fig.7.20, most of the interaction peaks appear in the period between 374 °Cd and 531 °Cd, which corresponds to the non-linearity stage we found in the step 1 linearity analysis. As such, we pick the time point 411.15 °Cd that proves to be the most important time point (for all the important transit and interactions, etc). To locate the interactions between parameters at this most non-linear and interaction-active stage, we present the  $S_{ij}$  matrix as in tab.7.11.

As shown in the table, factor  $\gamma$  is the only one that has relative high interactions with all the other factors, like the  $ST_i - S_i$  curve for  $\gamma$  (black line) in fig.7.20 reaches



**Fig. 7.19** : NEMA: Selected factors:  $S_i$ ; Output AreaGreenTotal; the lowest sum of  $S_i$  of all the selected factors appears at 411.15 °Cd, which means around this point, the higher order sensitivity should take more part of the system variance.

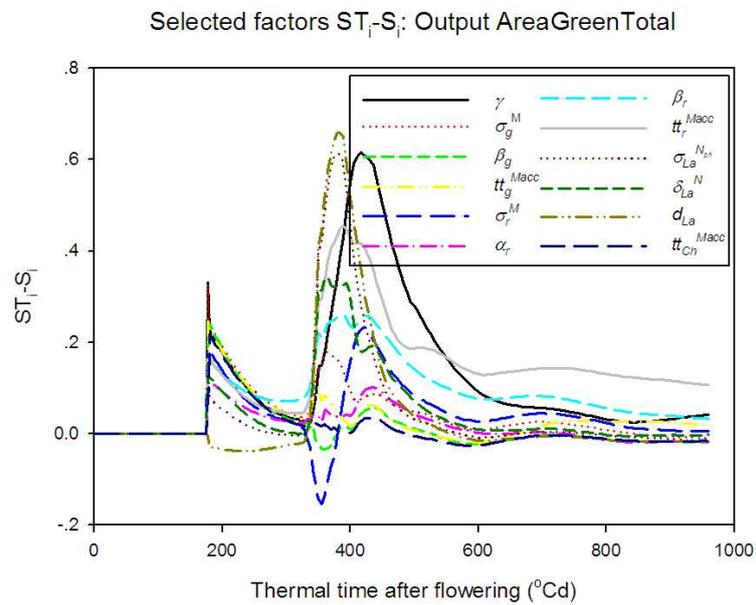
its maximum at this point. Besides, factor  $tt_r^{Macc}$  also has interactions with  $\sigma_{La}^{Nph}$ ,  $\delta_{La}^N$ ,  $d_{La}$  which we can not ignore. The interactions stand for the main part of inter-module interactions between module DMfluxes ( $tt_r^{Macc}$ ) and module Nfluxes ( $\sigma_{La}^{Nph}$ ,  $\delta_{La}^N$ ), between module DMfluxes ( $tt_r^{Macc}$ ) and module Tissuedeath ( $d_{La}$ ).

Among all these factors that have interactions with  $\gamma$ ,  $\sigma_r^M$  has the strongest interaction 10.97% with  $\gamma$ . Another typical example is  $d_{La}$ , though the total ranking of  $d_{La}$  is the last among the selected factors and the absolute value of its TGI  $S_i$  is only 0.004, but it still has 12.31% totally of interactions at 411.15 °Cd when we do the overall model analysis with the selected factors. We can see that it has even higher interactions at 374 °Cd in fig.7.20. Thanks to the comprehensive consideration of group analysis and module by module analysis before, we can have an appropriate standard to decide how many factors should be selected for the overall model analysis in order not to miss such important interactions when the main effect of certain factor is low but the interaction is strong.

To demonstrate the strategy for the module by module analysis specifically, as an example, we have presented the related indices of all the parameters with uncertainty for the NEMA model in this section. In the next sections, for the other model output of interest, we use the same procedure of parameters selection that we have illustrated in detail here, so we only list the indices for the finally selected factors, and focus on the main features for the specified output.

**Tab. 7.11** : NEMA:  $S_{ij}$  for selected factors at thermal time after flowering 411.15 °Cd, output AreaGreenTotal

	$\gamma$	$\sigma_g^M$	$\beta_g$	$tt_g^{Macc}$	$\sigma_r^M$	$\alpha_r$	$\beta_r$	$tt_r^{Macc}$	$\sigma_{La}^{Nph}$	$\delta_{La}^N$	$d_{La}$	$tt_{Ch}^{Macc}$
$\gamma$	-	0.027509	0.047476	0.038851	0.10968	0.046097	0.056311	0.047124	0.042893	0.01245	0.055861	0.032027
$\sigma_g^M$	0.027509	-	0.000967	0.0008	0.003352	7.95E-05	0.003306	-0.00043	0.002496	0.003398	0.000969	0.000381
$\beta_g$	0.047476	0.000967	-	-0.00086	-0.00011	-0.00169	-0.00492	-0.00062	0.004646	0.009388	0.003912	3.77E-05
$tt_g^{Macc}$	0.038851	0.0008	-0.00086	-	-0.00044	-0.00185	-0.00333	0.001642	0.003269	0.003101	0.001038	-0.00013
$\sigma_r^M$	0.10968	0.003352	-0.00011	-0.00044	-	0.001966	5.91E-05	0.001801	0.000651	0.002221	0.005045	0.000459
$\alpha_r$	0.046097	7.95E-05	-0.00169	-0.00185	0.001966	-	-0.00212	9.12E-05	0.010585	0.017425	0.002878	-0.00038
$\beta_r$	0.056311	0.003306	-0.00492	-0.00333	5.91E-05	-0.00212	-	0.004816	0.007596	0.008951	0.005149	0.003311
$tt_r^{Macc}$	0.047124	-0.00043	-0.00062	0.001642	0.001801	9.12E-05	0.004816	-	0.031849	0.054932	0.024656	0.012335
$\sigma_{La}^{Nph}$	0.042893	0.002496	0.004646	0.003269	0.000651	0.010585	0.007596	0.031849	-	0.000103	0.011705	0.001646
$\delta_{La}^N$	0.01245	0.003398	0.009388	0.003101	0.002221	0.017425	0.008951	0.054932	0.000103	-	0.01029	0.000257
$d_{La}$	0.055861	0.000969	0.003912	0.001038	0.005045	0.002878	0.005149	0.024656	0.011705	0.01029	-	0.001625
$tt_{Ch}^{Macc}$	0.032027	0.000381	3.77E-05	-0.00013	0.000459	-0.00038	0.003311	0.012335	0.001646	0.000257	0.001625	-
$\sum^j S_{ij}$	0.516279	0.042822	0.058221	0.042092	0.124683	0.073079	0.079125	0.178195	0.117438	0.122515	0.123128	0.051574



**Fig. 7.20** : NEMA: Selected factors:  $ST_i - S_i$ ; Output AreaGreenTotal; the peaks of most of the curves appear at the period between  $375^{\circ}Cd$  and  $411.15^{\circ}Cd$ .

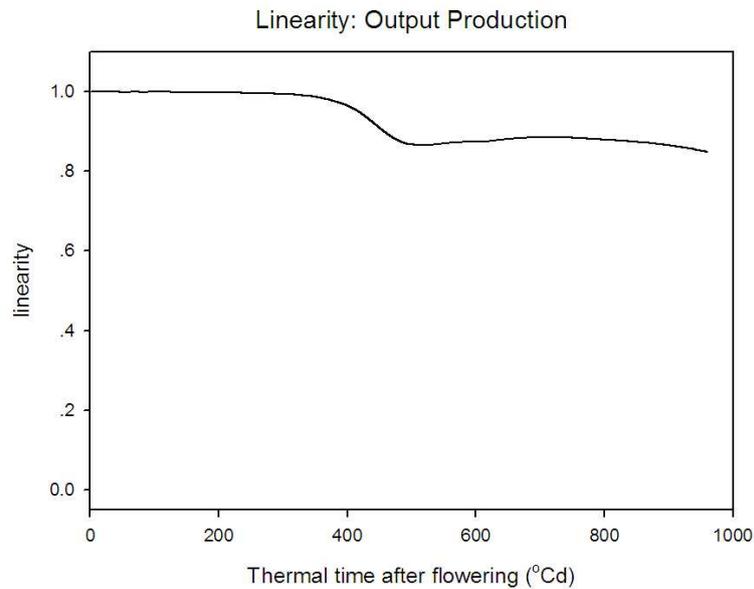
### 7.3.2 Output Production

As for the output Production, the linearity shown in fig.7.21 keeps a relatively high value above 0.86. So things are more or less simplified here, but since it is not typically quasi-linear ( $> 0.95$ ) for nearly half of the whole period, to be cautious we still keep the interaction checking between modules in the group analysis as in fig.7.23 and fig.7.24, and fig.7.26.

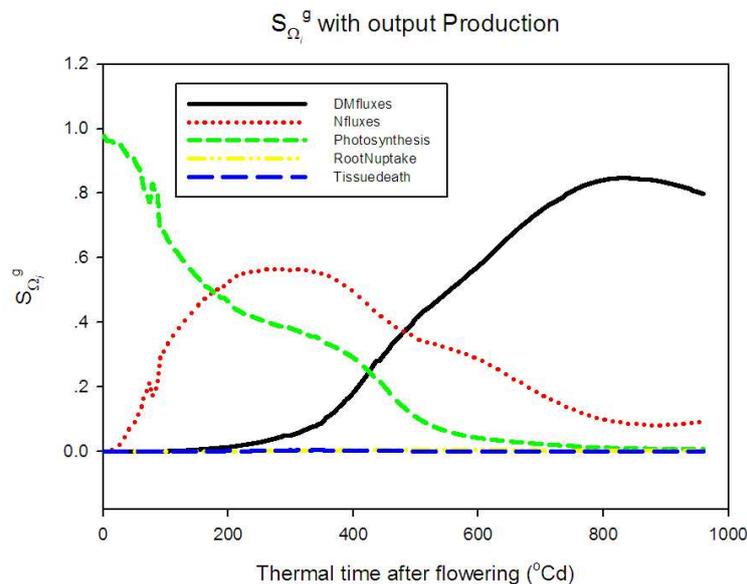
The main information given by fig.7.22 is that the first order sensitivity is mainly shared by modules Photosynthesis, Nfluxes and DMfluxes in order. Furthermore, there is little effect from module RootNuptake and Tissuedeath, likewise for the inter-module interactions. The inter-module interactions only concentrate between modules DMfluxes and Nfluxes as shown in fig.7.23 and fig.7.24. Combining the information we got from the group analysis with the  $ST_i$  we get from the intra-module analysis in our strategy, 12 factors among 83 total for the model were selected as shown in tab.7.12. The distribution is 8 factors from DMfluxes, 3 from Nfluxes and 1 from Photosynthesis. In tab.7.13, we can see that the screening processing for the model has ignored 1.29% of the main effect from 84.8%, which is a very safe amount to guarantee the reliability of the results afterwards.

As we can see in fig.7.25, in the first period before 465 °Cd,  $\omega_{La,2}$  (the only factor selected from module Photosynthesis),  $\delta_{La}^N$  and  $\sigma_{La}^{Nph}$  (the 2 out of 3 factors from module Nfluxes) take most portion of the uncertainty. For the last period,  $tt_r^{Macc}$  from module DMfluxes and  $\gamma$  from module Nfluxes take the role in change. It corresponds to what was observed regarding module indices in the group analysis, as shown in fig.7.22. Again, basic features are kept after the parameter selection.

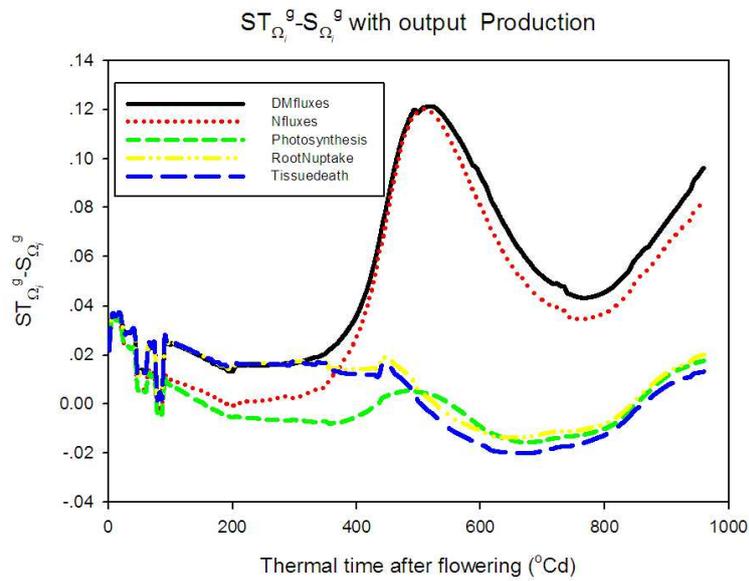
For interactions between selected factors, the 3 ones  $\gamma$ ,  $\beta_r$  and  $tt_r^{Macc}$  were identified for this issue as shown in fig.7.26, in which  $tt_r^{Macc}$  is responsible for most part of the interaction.



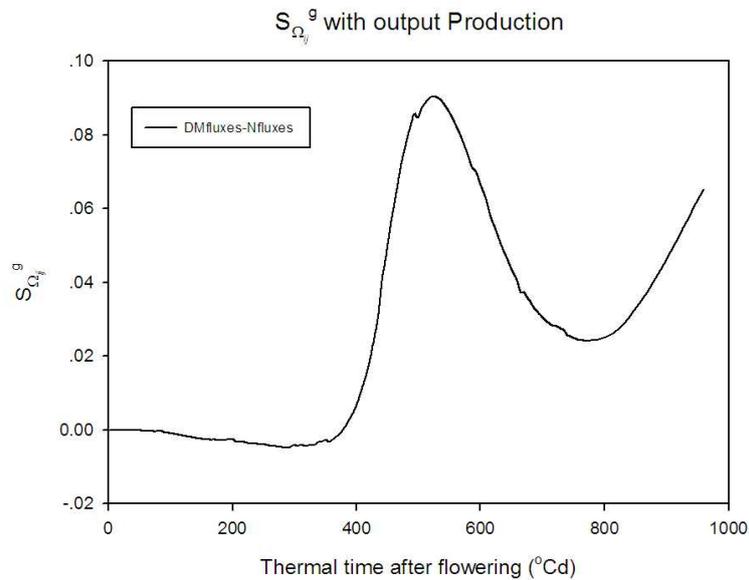
**Fig. 7.21 :** NEMA: Linearity: Output Production. There is a small slope between 360 °Cd and 474 °Cd, the drop is from 0.99 to 0.86. Before and after this slope, there are two flat levels with the linearity around 0.99 and 0.86 respectively.



**Fig. 7.22 :** NEMA:  $S_{\Omega_i}^g$  Output Production. Module Photosynthesis ranks first from the beginning to 165 °Cd, at which point module Nfluxes surpasses it. Module DMfluxes starts to increase steadily at point 341 °Cd with the index 0.08, after which it crosses with module Photosynthesis at point 423 °Cd with value 0.243 and crosses with module Nfluxes at point 474 °Cd with value 0.36. After these two crossed points, module DMfluxes gets to dominate the model uncertainty.



**Fig. 7.23** : NEMA:  $ST_{\Omega_i}^g - S_{\Omega_i}^g$ : Output Production. The negative values are caused by the computing accuracy limitation. The only interaction is between modules DMfluxes and Nfluxes. They share the same evolution. At 520 °Cd, they reach the highest value as 0.12.



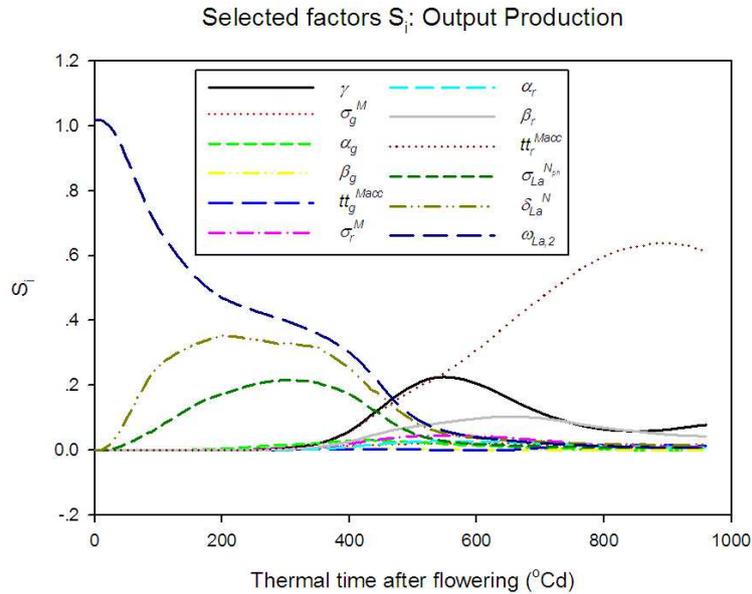
**Fig. 7.24** : NEMA:  $S_{\Omega_{ij}}^g$ : Output Production. The peak value is 0.09 at 520 °Cd.

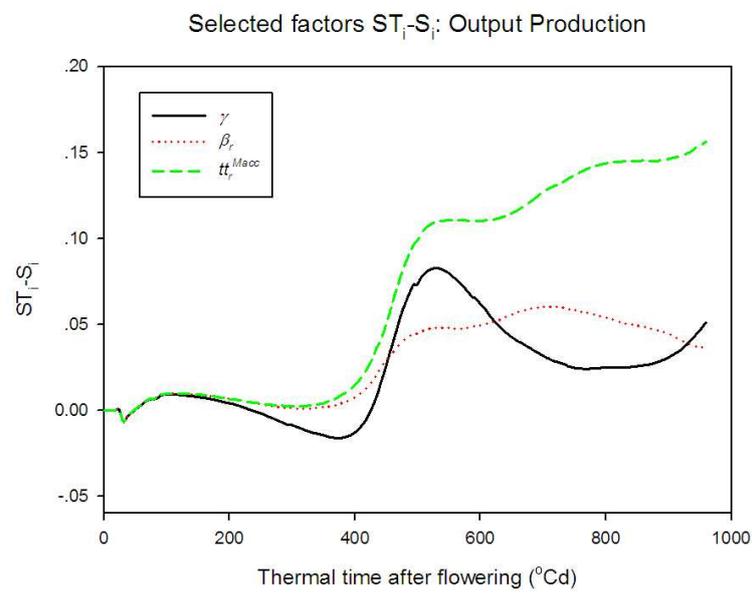
**Tab. 7.12** : NEMA: Selected factors; Output:Production

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
Module:	DMfluxes				
$\sigma_g^M$	0.010966	6	0.026162	6	9
$\alpha_g$	0.001963	8	0.012057	8	12
$\beta_g$	0.001687	9	0.013025	7	11
$tt_g^{Macc}$	0.013809	5	0.06488	3	4
$\sigma_r^M$	0.016135	3	0.046696	5	6
$\alpha_r$	0.013812	4	0.051463	4	5
$\beta_r$	0.04353	2	0.105708	2	2
$tt_r^{Macc}$	0.629453	1	0.768214	1	1
Module:	Nfluxes				
$\gamma$	0.067811	1	0.098038	1	3
$\sigma_{La}^{Nph}$	0.007753	3	0.024685	3	10
$\delta_{La}^N$	0.017416	2	0.034581	2	8
Module:	Photosynthesis				
$\omega_{La,2}$	0.010495	1	0.039516	1	7

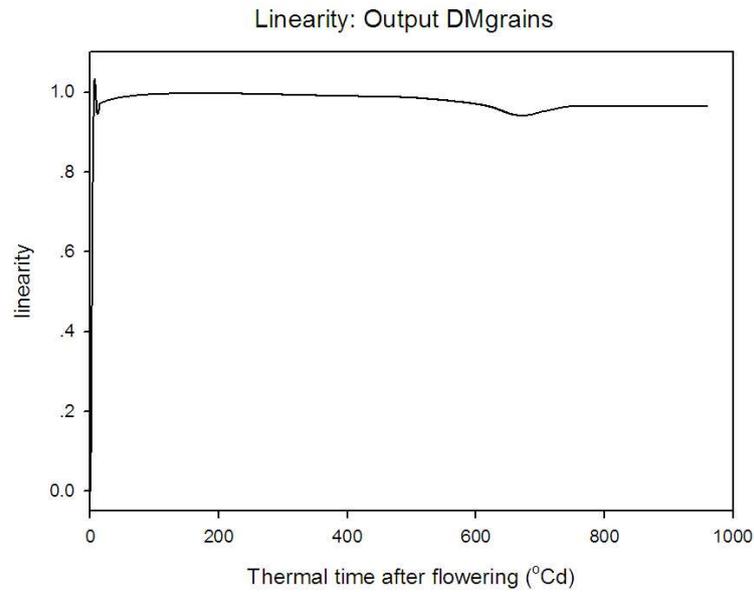
**Tab. 7.13** : NEMA: Overall parameter selection analysis;Output:Production

	$\sum TGIS_i$	$\sum SelectedTGIS_i$	Difference
DMfluxes	0.737226	0.731355	0.005871
Nfluxes	0.095379	0.092979	0.002399
Photosynthesis	0.011457	0.010495	0.000962
RootNuptake	0.00332	0	0.00332
Tissuedeath	0.000443	0	0.000443
Overall model	0.847825	0.834829	0.012996

**Fig. 7.25** : NEMA: Selected factors  $S_i$ ; Output Production.  $\omega_{La,2}$ ,  $\delta_{La}^N$  and  $\sigma_{La}^{Nph}$  dominate the model from beginning to 465 °Cd. Afterwards,  $\gamma$  and  $tt_r^{Macc}$  rank first, the difference is that  $\gamma$  starts decreasing after a short period of increasing and  $tt_r^{Macc}$  keeps a steady increase.



**Fig. 7.26** : NEMA: Selected factors  $ST_i - S_i$ ; Output Production. The difference between  $ST_i$  and  $S_i$  concentrate only for  $\gamma$ ,  $\beta_r$  and  $tt_r^{Macc}$ , which are the first 3 factors in the overall model ranking (see tab.7.12).



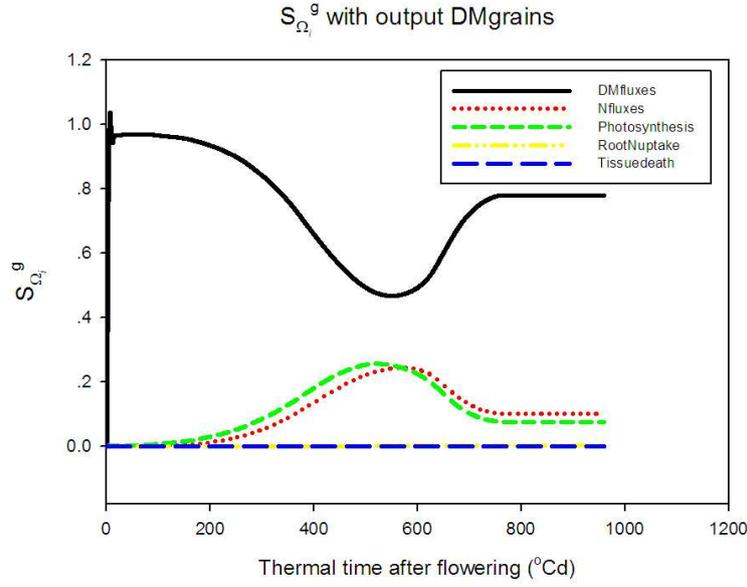
**Fig. 7.27** : NEMA: Linearity: Output DMgrains

### 7.3.3 Output DMgrains

As in fig.7.27, the linearity of the model with DMgrains as the output is always above 0.95, which is high enough for us to take the model as quasi-linear. So we can ignore the interactions either between the modules or between the parameters in the analysis. So all the steps related to the interaction analysis are skipped here.

As for the group analysis in fig.7.28, we can see the module DMfluxes has the obvious priority over all the other modules all along time. Module Nfluxes and Photosynthesis have the same trend over time. These two sets of curves go through a reverse change, from 446 °Cd to 629 °Cd, which is explained by fig.7.29: the transit is caused by the decrease in sensitivity of  $\sigma_g^M$ ,  $\alpha_g$  from module DMfluxes and a small increase of  $\omega_{La,2}$  from module Photosynthesis,  $\delta_{La}^N$  from module Nfluxes. Module RootNuptake and Tissuedeath have little effect upon the output variance.

Considering the group analysis result and the intra-module analysis, we selected 12 factors out of 83 as shown in tab.7.15. Due to the high linearity of the system with output DMgrains, the selected factors shown in tab.7.15 have the same ranking for  $S_i$  and  $ST_i$ . The difference of the sensitivity indices ignored by the selection is 0.089 as shown in tab.7.15.



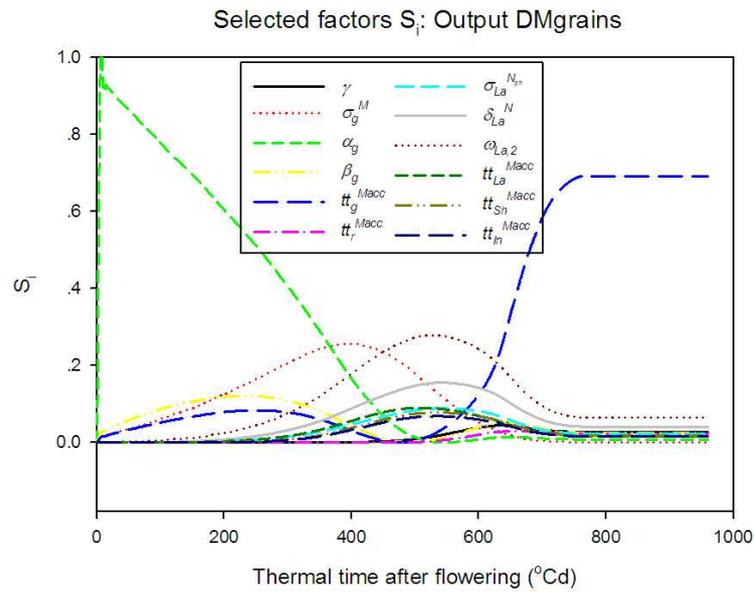
**Fig. 7.28** : NEMA:  $S_{\Omega_i}^g$  Output DMgrains. Module DMfluxes has the highest first-order index all the way. Module Nfluxes and Photosynthesis have the same trend. Module RootNuptake and Tissuedeath nearly have nil sensitivity along the time period. From 446 °Cd to 629 °Cd, the trend of DMfluxes and Nfluxes, Photosynthesis goes through a reverse change, which indicates a transition period.

**Tab. 7.14** : NEMA: Selected factors;Output:DMgrains

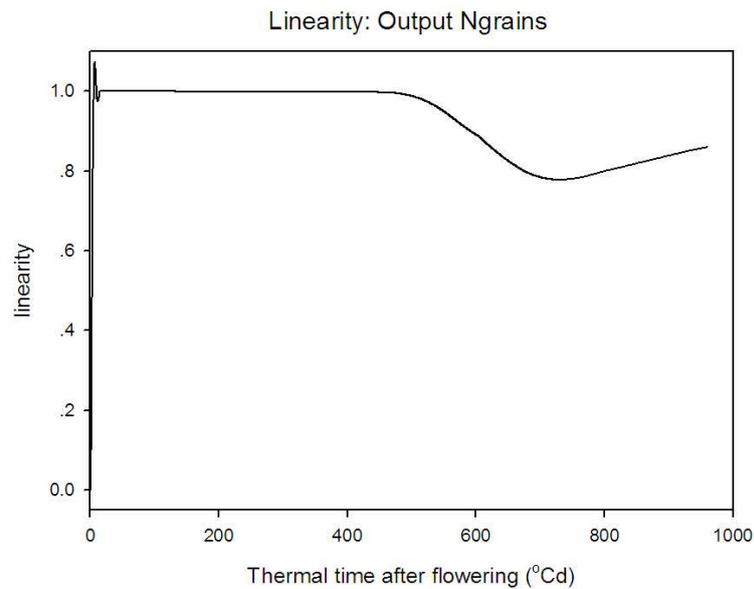
Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
Module:	DMfluxes				
$\sigma_q^M$	0.062628	3	0.048309	4	5
$\alpha_g$	0.262091	2	0.272972	2	2
$\beta_g$	0.05951	4	0.05488	3	4
$tt_q^{Macc}$	0.298627	1	0.298451	1	1
$tt_r^{Macc}$	0.007711	9	0.000449	8	12
$tt_l^{Macc}$	0.017336	5	0.005114	6	9
$tt_s^{Macc}$	0.016069	6	0.00701	5	7
$tt_n^{Macc}$	0.015578	7	0.001234	7	11
Module:	Nfluxes				
$\gamma$	0.015428	3	0.006748	2	8
$\sigma_{Nph}^{Lq}$	0.020627	2	0.005104	3	10
$\delta_{La}^{Lq}$	0.040441	1	0.024408	1	6
Module:	Photosynthesis				
$\omega_{La,2}$	0.074522094	1	0.079008	1	3

**Tab. 7.15** : NEMA: Overall parameter selection analysis;Output:DMgrains

	$\sum TGIS_i$	$\sum SelectedTGIS_i$	Difference
DMfluxes	0.809829	0.739551	0.070278
Nfluxes	0.084324	0.076495	0.007829
Photosynthesis	0.084416	0.074522	0.009894
RootNuptake	0.000765	0	0.000765
Tissuedeath	0.000193	0	0.000193
Overall model	0.979528	0.890568	0.08896



**Fig. 7.29** : NEMA: Selected factors  $S_i$ : Output DMgrains.  $\alpha_g$  starts with the first ranking with high portion of sensitivity but decrease to nil until 549 °Cd, after which  $tt_g^{Macc}$  increases until around 0.7 at the end period. The indices of  $S_i$  for  $\sigma_g^M$  and  $\omega_{La,2}$  meet at 446 °Cd, and those of  $S_i$  for  $\omega_{La,2}$  and  $tt_g^{Macc}$  meet at 629.125 °Cd, between the two points,  $\omega_{La,2}$  from module Photosynthesis and  $\delta_{La}^N$  from module Nfluxes rank the first two factors. It is interpreted by the transition period in module analysis in fig.7.28.



**Fig. 7.30** : NEMA: Linearity: Output Ngrains. The system is linear from the beginning to 494 °Cd. After a steady decrease until 748 °Cd with linearity 0.78, the system basically maintains the same linearity level though with a slight increase.

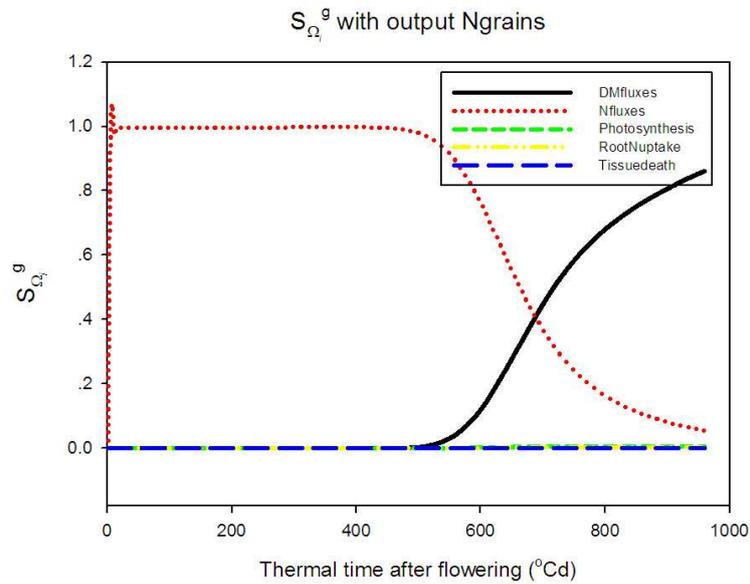
### 7.3.4 Output Ngrains

Shown in fig.7.30, as for the linearity when the system output is Ngrains, the model is slightly non-linear after 600 °Cd. We followed the same analysis steps as when system output is AreaGreenTotal.

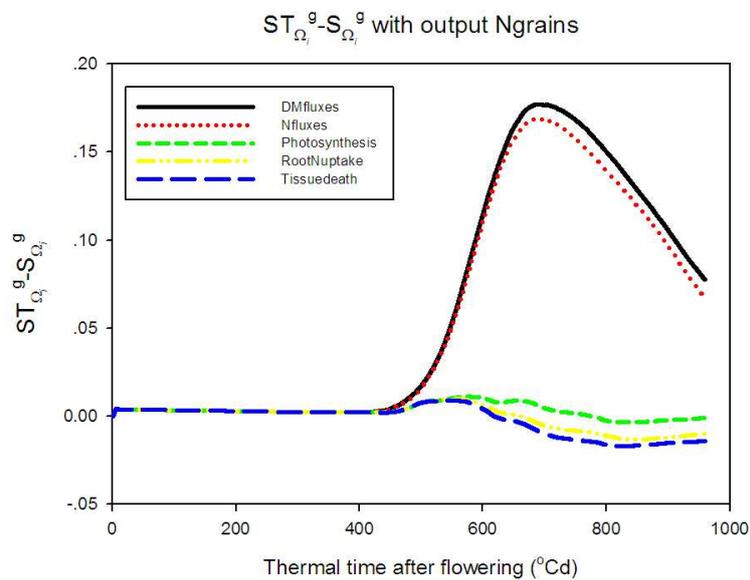
In fig.7.31, modules DMfluxes and Nfluxes have the complementary trends for  $S_{\Omega_i}^g$  while the other three modules have nearly no effect. Then it is not a surprise that the main inter-module interaction goes directly between module DMfluxes and Nfluxes as shown in fig.7.32 and fig.7.33.

We follow the standard procedure for the parameter selection detailed in section.7.3.1. For the output Ngrains, the sensitivity main effect is very concentrated on 6 factors from two modules: DMfluxes and Nfluxes, as shown in tab.7.16. From tab.7.17, we can see that even only with 6 factors among 83 for the overall model, they have already explained the main effect 84.34% of the variance out of 86.37% for all the 83 factors, with only a loss of 2%.

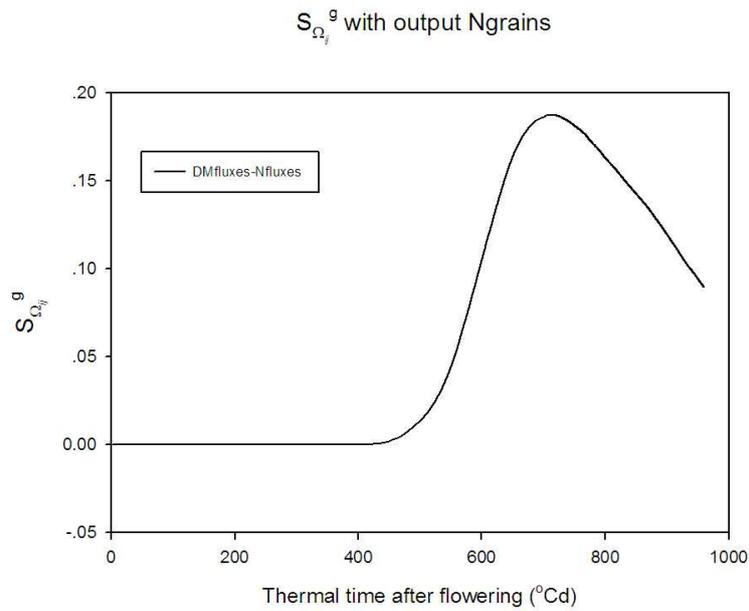
A similar concentration also appears for the intra-module interaction shown in tab.7.18 and the inter-module interaction shown in tab.7.19: most intra-module interaction only exists in module DMfluxes and the inter-module interaction only happens between DMfluxes and Nfluxes.



**Fig. 7.31 :** NEMA:  $S_{\Omega_i}^g$  Output Ngrains. Module Nfluxes has the most important effect to the system until 685 °Cd, after which DMfluxes gets more and more importance and Nfluxes effect decreases steadily.



**Fig. 7.32 :** NEMA:  $ST_{\Omega_i}^g - S_{\Omega_i}^g$ , Output Ngrains. It is obvious that only modules DMfluxes and Nfluxes have interaction regarding the output Ngrains. Note that the peak also appears at the transit point 685 °Cd between Nfluxes and DMfluxes.



**Fig. 7.33 :** NEMA:  $S_{\Omega_{ij}}^g$  Output Ngrains

**Tab. 7.16 :** NEMA: Selected factors;Output:Ngrains

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
Module: DMfluxes					
$tt_g^{Macc}$	0.004262	6	0.006578	5	6
$\sigma_r^M$	0.014819	3	0.02549	3	4
$\alpha_r$	0.011334	4	0.018858	4	5
$\beta_r$	0.035907	2	0.053437	2	3
$tt_r^{Macc}$	0.31217	1	0.438664	1	2
Module: Nfluxes					
$\gamma$	0.464929	1	0.553625	1	1

**Tab. 7.17 :** NEMA: Overall parameter selection analysis; Output:Ngrains

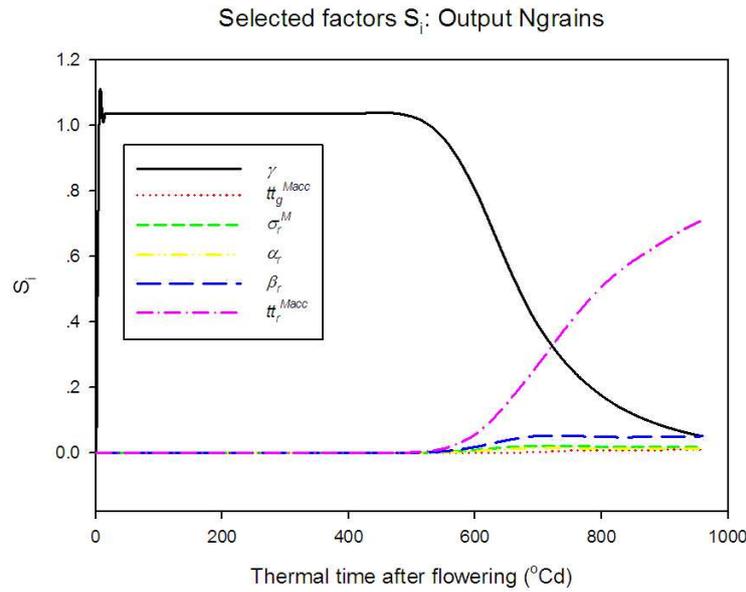
	$\sum TGIS_i$	$\sum SelectedTGIS_i$	Difference
DMfluxes	0.392619	0.378493	0.014126
Nfluxes	0.467102	0.464929	0.002173
Photosynthesis	0.002535	0	0.002535
RootNuptake	0.00142	0	0.00142
Tissuedeath	1.18E-05	0	1.18E-05
Overall model	0.863687	0.843421	0.020266

**Tab. 7.18 :** NEMA: Intra-Module Interaction analysis; Output:Ngrains

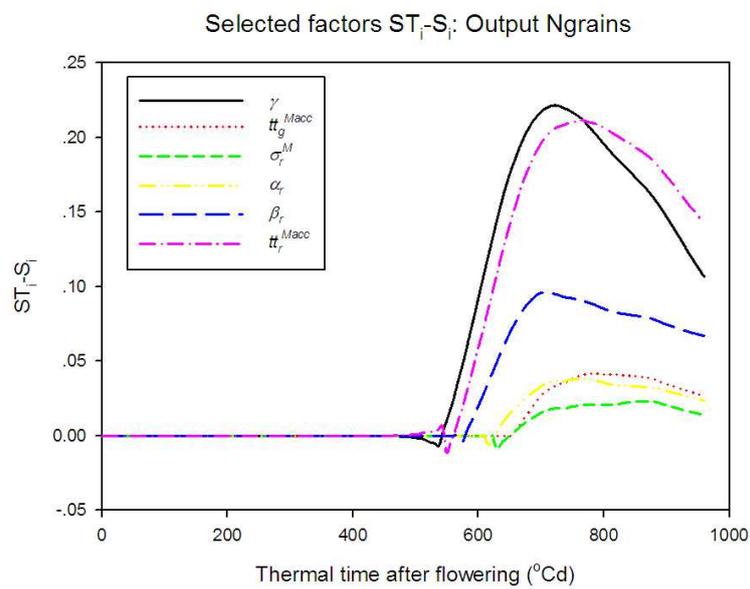
	$TGIS_{\Omega_i}^g$	$\sum TGIS_i$	Intra-Module Interaction
DMfluxes	0.446579	0.392619	0.053961
Nfluxes	0.467633	0.467102	0.000531
Photosynthesis	0.002995	0.002535	0.00046
RootNuptake	0.001532	0.00142	0.000112
Tissuedeath	2.02E-05	1.18E-05	8.37E-06
Overall model	0.918759	0.863687	0.055072

**Tab. 7.19** : NEMA: Inter-Module Interaction; Output:Ngrains

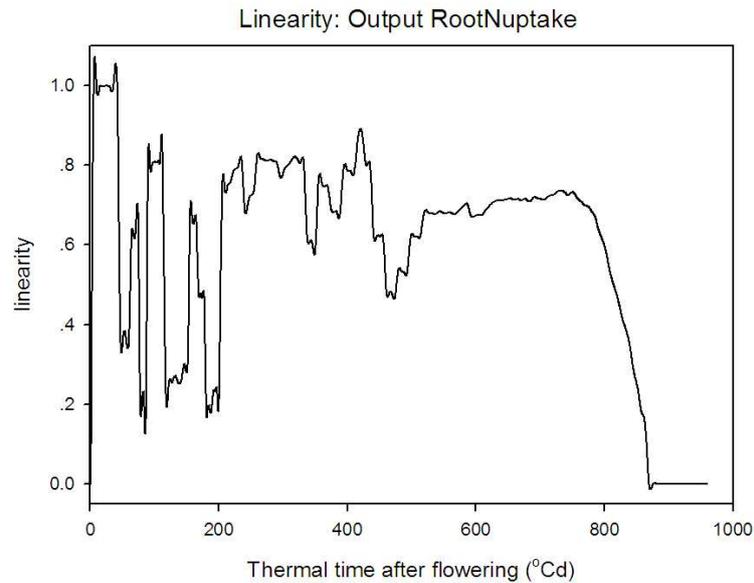
	$TGIS_{\Omega_i}^g$	$TGIST_{\Omega_i}^g$	Inter-Module Interaction
DMfluxes	0.446579	0.535207	0.088628
Nfluxes	0.467633	0.542965	0.075332
Photosynthesis	0.002995	0.00439	0.001396
RootNuptake	0.001532	-0.00316	-0.0047
Tissuedeath	2.02E-05	-0.00736	-0.00738

**Fig. 7.34** : NEMA: Selected factors  $S_i$  Output Ngrains. Similar transition exists between factor  $\gamma$  from module Nfluxes and factor  $tt_r^{Macc}$  from module DMfluxes as in fig.7.31, which means that the two factors can represent most part of the variance from these two modules respectively.**Tab. 7.20** : NEMA:  $S_{ij}$  for selected factors at thermal time after flowering  $685^{\circ}\text{Cd}$ , output Ngrains. The interaction between  $\gamma$  and  $tt_r^{Macc}$  takes most part of the interaction between parameters at this time point.

	$\gamma$	$tt_g^{Macc}$	$\sigma_r^M$	$\alpha_r$	$\beta_r$	$tt_r^{Macc}$
$\gamma$	-	0.005294	0.031051	0.003267	0.021093	0.129051
$tt_g^{Macc}$	0.005294	-	0.002653	-0.00193	0.001179	0.006328
$\sigma_r^M$	0.031051	0.002653	-	-0.00089	0.001278	0.009428
$\alpha_r$	0.003267	-0.00193	-0.00089	-	0.000261	0.00506
$\beta_r$	0.021093	0.001179	0.001278	0.000261	-	0.011148
$tt_r^{Macc}$	0.129051	0.006328	0.009428	0.00506	0.011148	-
$\sum^j S_{ij}$	0.189756	0.01352	0.043517	0.00576	0.034959	0.161015



**Fig. 7.35** : NEMA: Selected factors  $ST_i - S_i$  Output Ngrains.  $\gamma$  and  $tt_r^{Macc}$  have the same curve shape of the higher order interactions. At 685 °Cd, besides the strong interaction for  $\gamma$  and  $tt_r^{Macc}$ ,  $\beta_r$  also contribute around 10% to the higher order sensitivity.



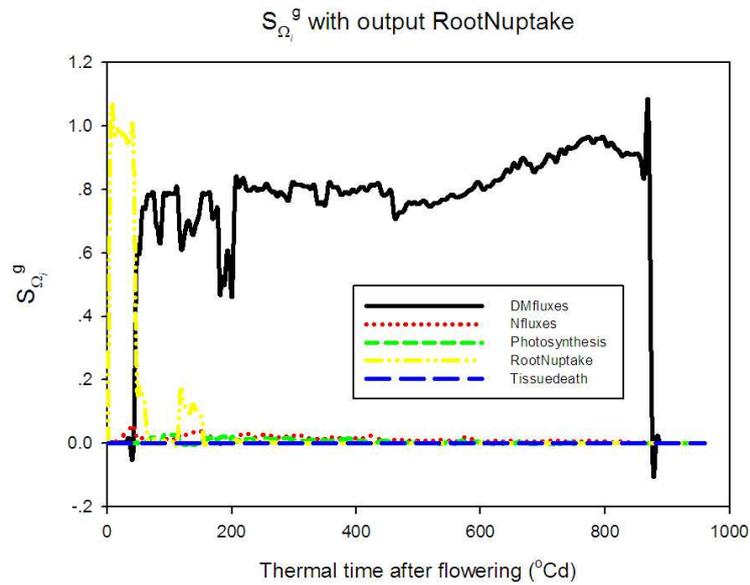
**Fig. 7.36** : NEMA: Linearity: Output RootNuptake. The non-linearity is very strong and presents strong variability.

### 7.3.5 Output RootNuptake

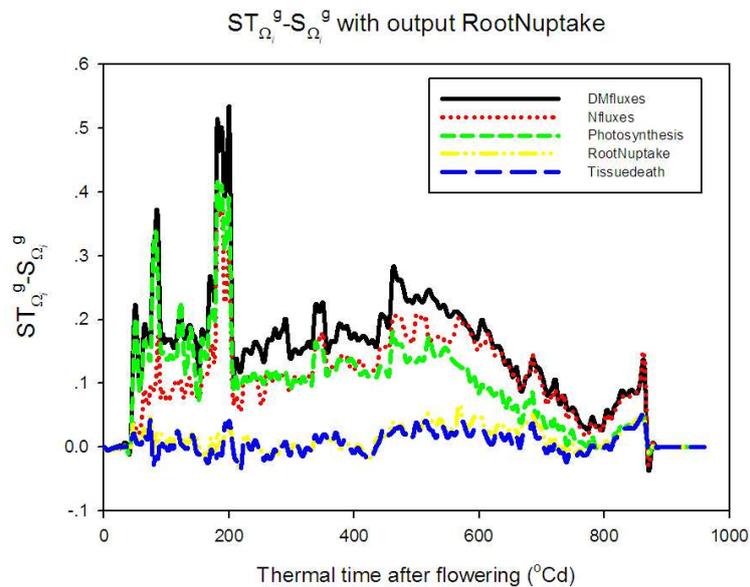
Generally, the system for output RootNuptake is very non-linear and with strong variability shown in fig.7.36. The first order sensitivity concentrates on module RootNuptake as for a short period at the beginning and then module DMfluxes as in fig.7.37. However the interactions between modules are more complicated. By checking  $S_{\Omega_i}^g$  in fig.7.38 and  $S_{\Omega_{ij}}^g$  in fig.7.39 together, two types of inter-module interactions were identified: DMfluxes and Nfluxes, DMfluxes and Photosynthesis.

The factor selection procedure only cut 0.6% of the first order sensitivity as shown in tab.7.22, which means 16 factors out of 83 in total can reflect the main effect of the system when the output is RootNuptake. By analyzing the selected factors in tab.7.21, intra-module interactions (mainly in DMfluxes as high as 24% when averaging with time) are stronger than inter-module interactions (between 0.11 and 0.14 appropriately) as in tab.7.23 and tab.7.24.

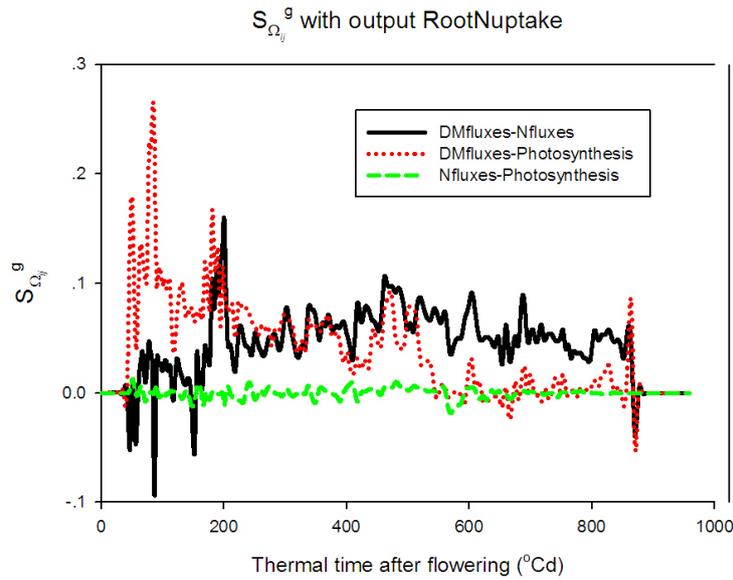
In fig.7.40, we can see that parameter  $tt_r^{Macc}$  stands for most part of the main effect of the system, but the interactions between parameters are distributed evenly for the selected factors as shown in fig.7.41. So it is necessary to check by modules for the interactions to get more oriented precise insights.



**Fig. 7.37 :** NEMA:  $S_{\Omega_i}^g$  Output RootNuptake. Module RootNuptake only dominates one short period at the beginning and module DMfluxes takes the most important role until the end. As for the main effect, module Nfluxes, Photosynthesis and Tissuedeath keeps low values all the time.



**Fig. 7.38 :** NEMA:  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  Output RootNuptake. Interactions exist among module DMfluxes, Nfluxes and Photosynthesis.



**Fig. 7.39 :** NEMA:  $S_{\Omega_{ij}}^g$  Output RootNuptake

**Tab. 7.21 :** NEMA: Selected factors;Output:RootNuptake

Factor	TGI $S_i$		TGI $ST_i$		
	Index	Intra-Module Ranking	Index	Intra-Module Ranking	Model Ranking
Module: DMfluxes					
$\sigma_g^M$	0.0082	5	0.169337	6	6
$\alpha_g$	0.006524	7	0.089024	8	11
$\beta_g$	0.003358	8	0.123071	7	7
$tt_g^{Macc}$	0.024149	3	0.232268	3	3
$\sigma_r^M$	0.012747	4	0.1813	5	5
$\alpha_r$	0.007595	6	0.18315	4	4
$\beta_r$	0.042309	2	0.331893	2	2
$tt_r^{Macc}$	0.491072	1	0.857027	1	1
$tt_{La}^{Macc}$	0.000751	10	0.070719	12	16
$tt_{Sp}^{Macc}$	8.18E-05	22	0.0784	11	14
$tt_{Tg}^{Macc}$	0.000482	11	0.07992	10	13
$tt_{Ch}^{Macc}$	0.001178	9	0.084181	9	12
Module: Nfluxes					
$\gamma$	0.007203	1	0.113062	1	8
$\sigma_{Nph}^{La}$	0.00126	3	0.075261	3	15
$\delta_{La}^N$	0.002005	2	0.09054	2	10
Module: Photosynthesis					
$\omega_{La,2}$	0.001996381	1	0.101541	1	9

**Tab. 7.22 :** NEMA: Overall parameter selection analysis; Output:RootNuptake

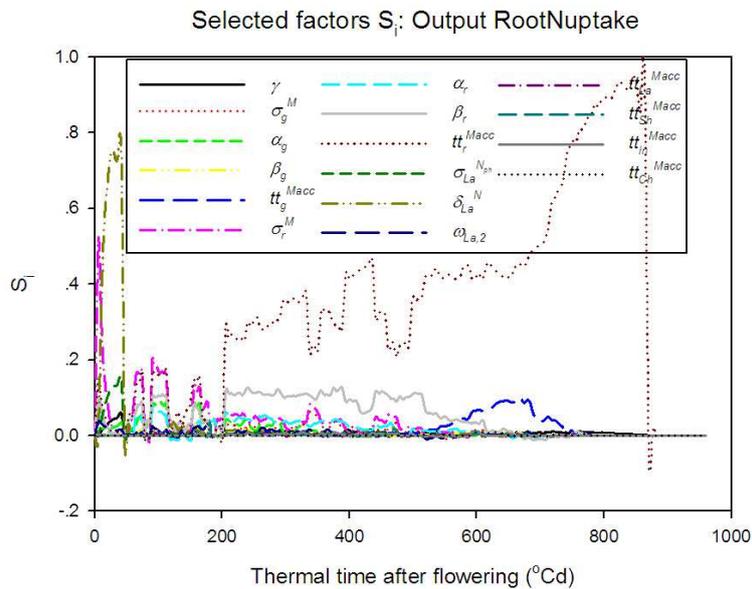
	$\sum TGIS_i$	$\sum SelectedTGIS_i$	Difference
DMfluxes	0.600248	0.598448	0.0018
Nfluxes	0.011304	0.010468	0.000837
Photosynthesis	0.002105	0.001996	0.000108
RootNuptake	0.003351	0	0.003351
Tissuedeath	7.08E-05	0	7.08E-05
Overall model	0.617079	0.610912	0.006167

**Tab. 7.23** : NEMA: Intra-Module Interaction analysis; Output:RootNuptake

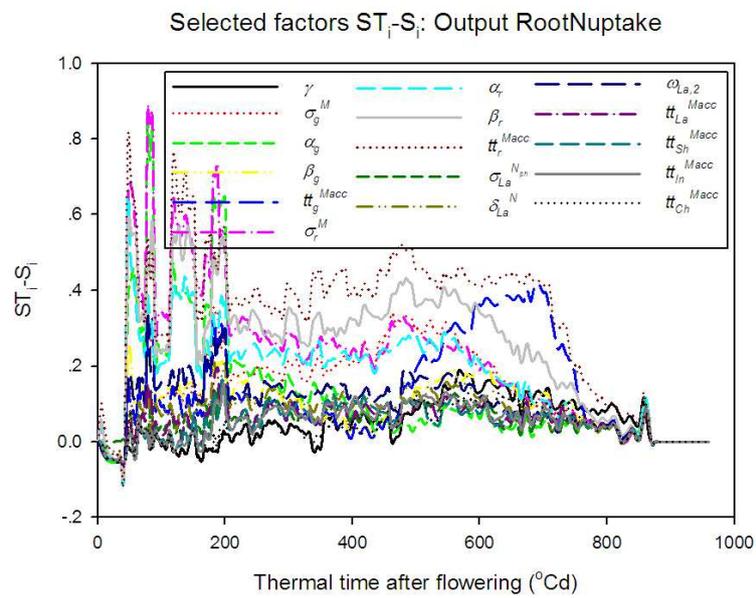
	$TGIS_{\Omega_i}^g$	$\sum TGIS_i$	Intra-Module Interaction
DMfluxes	0.840487	0.600248	0.240238
Nfluxes	0.01109	0.011304	-0.00021
Photosynthesis	0.003427	0.002105	0.001323
RootNuptake	0.003808	0.003351	0.000457
Tissuedeath	0.00013	7.08E-05	5.93E-05
Overall model	0.858941	0.617079	0.241862

**Tab. 7.24** : NEMA: Inter-Module Interaction; Output:RootNuptake

	$TGIS_{\Omega_i}^g$	$TGIST_{\Omega_i}^g$	Inter-Module Interaction
DMfluxes	0.840487	0.981854	0.141367
Nfluxes	0.01109	0.127559	0.116469
Photosynthesis	0.003427	0.087084	0.083657
RootNuptake	0.003808	0.021292	0.017484
Tissuedeath	0.00013	0.009546	0.009416



**Fig. 7.40** : NEMA: Selected factors  $S_i$ : Output RootNuptake



**Fig. 7.41** : NEMA: Selected factors  $ST_i - S_i$ : Output RootNuptake

### 7.3.6 Comprehensive comparison

Mathematically for sensitivity analysis, the definition of one model for the analysis includes the equations describing the phenomenon and one scalar output variable. So when the outputs are different though the equations are the same for our analysis, we take each case as different mathematical model. That's the reason why we presented the results for each output in the previous sections.

Even though from a mathematical point of view the models are different, as for the modelers, we still hope that from the result of several outputs we can get a common view for the model. So we listed the selected parameters for all the interesting outputs we have analyzed in tab.7.25 along together with their rankings corresponding to the first order index  $S_i$  in the overall model.

From tab.7.25 we can see that though with differences, the selected factors that can take most portion for the variance of the system concentrate on 17 factors out of 83 in total for the model. This conclusion has already helped us a lot to reduce the number of parameters with uncertainties when doing the parameter estimation.

What's more, if we check the diversity in details, we can find that the distribution of the selected factors is 12 out of 34 factors for module DMfluxes, 3 out of 28 factors for module Nfluxes, 1 out of 5 factors for module Tissuedeath, 1 out of 10 factors for module Photosynthesis and 0 out of 5 for module RootNuptake. And as we used  $ST_i$  for the screening, when we check the  $S_i$  ranking here, we can see that not all the parameters selected are the ones with the biggest  $S_i$ . For example, within the 12 selected factors,  $\beta_g$ , output AreaGreenTotal ranks 21st regarding  $S_i$ . So as we mentioned in section.5.3, we can not rely on selecting the factors according to  $S_i$  ranking in the intra-module analysis. The right way is to check the  $ST_i$  ranking on the basis of the same sampling points for each module analysis, plus comprehensively considering the group analysis for each module.

For NEMA model itself, the most important factors set for module Nfluxes are pretty steady for all the outputs: 1)  $\gamma$  standing for the relative rate of potential grain N filling during cell division, 2)  $\sigma_{La}^{Nph}$  standing for relative rate of photosynthetic N synthesis associated to xylem influx for entity Lamina, 3)  $\delta_{La}^N$  standing for relative degradation rates of remobilizable N for entity Lamina, in which  $\gamma$  rules all the outputs and most of the time has the very high ranking.  $\gamma$  appears to be a crucial factor for the general NEMA model. For the 28 factors in this module, the sensitivity is very concentrated on these 3 ones. However, for module DMfluxes, it is not the case. The 12 factors for DMfluxes have different rankings for different outputs. Generally, 1)  $tt_r^{Macc}$  standing for the duration during which roots can accumulate dry mass 2)  $tt_g^{Macc}$  standing for stands for the duration during which grains can accumulate dry mass, those two rank among the most important for all the 5 outputs. Secondly, 1)  $\sigma_g^M$  standing for relative sink strength of grains, 2)  $\sigma_r^M$  standing for relative sink strength of roots, 3)  $\alpha_r, \beta_r$  standing for Beta function indices for roots, have also

noticeable effect for 4 of the outputs. Compared to modules DMfluxes and Nfluxes, modules Tissuedeath and Photosynthesis tend to have specific important parameters for different outputs, like  $d_{La}$  (standing for proportion of maximum specific N mass at which tissues die for Lamina) for output of AreaGreenTotal and  $\omega_{La,2}$  (standing for proportion coefficient linking photosynthesis at saturating PAR and N mass per unit) specially for output DMgrains.

If we categorize the factors in tab.7.25 by organs, we can see that in the 4 types of parameters presented in tab.2.3 for module DMfluxes, the ones for organ grain and root tend to have significant effect for all outputs. The parameter characteristic of lamina tend to have the control for Nfluxes, Tissuedeath and Photosynthesis.

**Tab. 7.25 :** NEMA: Summary of selected factors'  $S_i$  ranking in overall model for different outputs. There are 5 main outputs we considered in the analysis, a) AreaGreenTotal, b) Production, c) DMgrains, d) Ngrains and e) RootNuptake. Factors that are selected for all the 5 outputs are marked in dark grey and for 4 outputs are marked in light grey. '-' means the factor is not selected for output.

$S_i$ ranking	a	b	c	d	e
Module:DMfluxes					
$\sigma_g^M$	6	8	4	-	5
$\alpha_g$	-	12	2	-	8
$\beta_g$	21	13	5	-	9
$tt_g^{Macc}$	9	7	1	7	3
$\sigma_r^M$	5	5	-	4	4
$\alpha_r$	7	6	-	5	6
$\beta_r$	3	3	-	3	2
$tt_r^{Macc}$	1	1	13	2	1
$tt_{Lg}^{Macc}$	-	-	8	-	16
$tt_{Sp}^{Macc}$	-	-	9	-	32
$tt_{Ip}^{Macc}$	-	-	10	-	17
$tt_{Ch}^{Macc}$	17	-	-	-	15
Module:Nfluxes					
$\gamma$	2	2	11	1	7
$\sigma_{La}^{Nph}$	8	10	7	-	14
$\delta_{La}^N$	4	4	6	-	10
Module:Tissuedeath					
$d_{La}$	11	-	-	-	-
Module:Photosynthesis					
$\omega_{La,2}$	-	9	3	-	11

## 7.4 Conclusions and discussion

In this chapter, we implemented the strategy developed in this thesis to 3 FSPMs with different levels of complexity, and inferred in each case what information can be drawn from this analysis.

Firstly we practiced on a simple source-sink model of maize growth, which is used to specifically study the process of carbon (C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations. We first practiced local SA method and its normalized version, then SRC, lastly Sobol's method. We studied the interest of global sensitivity analysis and its latest developments for the GreenLab model. SA helped us to get better understanding of source-sink dynamics and internal driving forces during plant growth.

Secondly we practiced on the GreenLab model of tree growth (applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis, which describes tree structural plasticity in response to trophic competition. We inspected 14 parameters related to tree biomass production and allocation, and follow dynamically the sensitivity indices for 50 years. The test case chosen corresponds to a specific model showing alternating patterns in growth phases resulting from the complex interactions between functioning and organogenesis.

Lastly we practiced on a functional-structural model, NEMA, describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering. This model is more mechanistic and more complex than the two previous ones. Basic biological modules are identified: namely in our test case Carbon distribution, Nitrogen distribution, Carbon acquisition via photosynthesis, Nitrogen acquisition by roots, Senescence. Several outputs of interest are considered for both intra-module and inter-module analysis: a) total green area of the plant, b) total dry mass production of the plant, c) dry mass of the grains, d) Nitrogen mass of the grains and e) root Nitrogen uptake. Basically, for each output of interest, we did 1) the non-linearity assessment, 2) the analysis of function module rankings, 3) the intra module analysis, 4) the inter module analysis, and finally 5) the overall model analysis based on the result of the 4 analyse before.

The real advantage of Sobol's method was particularly illustrated for the different steps of our method for module by module analysis, and its application to NEMA. For the specific analysis of NEMA, an important step has not been fully completed. It corresponds to the biological diagnosis of the model. Once sensitivity analysis revealed the importance of some very specific stage of growth, it would be interesting to really analyze from a biological point of view the processes that make these stages of interest. From a modelling point to view, being able to show that 17 parameters out of 83 can explain most of the variance is an important step forward, particularly for parameterization. However, we should also question why no parameter from the module of RootNuptake was selected: either the uncertainty of the parameters in

this module was underestimated, or level of description of the model structure is not appropriate.

## **Part IV**

# **CONCLUSIONS AND PERSPECTIVES**

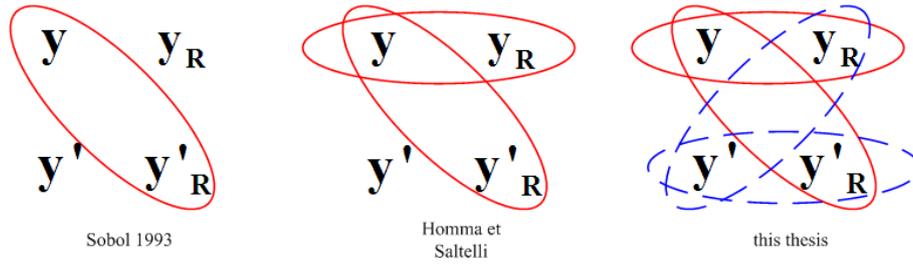


# 8. CONCLUSIONS AND PERSPECTIVES

## 8.1 Contributions

Global sensitivity analysis has a key role to play in the design and parameterization of functional-structural plant growth models (FSPM) which combine the description of plant structural development (organogenesis and geometry) and functional growth (biomass accumulation and allocation). Models of this type generally describe many interacting processes, count a large number of parameters, and their computational cost can be important. The general objective of this thesis is to develop a proper methodology for the sensitivity analysis of functional structural plant models and to investigate how sensitivity analysis can help for the design and parameterization of such models as well as providing insights for the understanding of underlying biological processes. Our contribution can be summarized as follows:

- In order to face the challenge of the computing cost and to meet the necessity of using Sobol's indices for the quantitative information about sensitivity of models, especially the interaction information, we improved a computing method inspired by [Homma and Saltelli, 1996] so that the model evaluations can be made best use of. We derived an estimator of the error of sensitivity indices evaluation with respect to the sampling size for this generic type of computational methods so that better control of the convergence of the estimations of Sobol's indices can be achieved.
- We designed a methodology adapted to FSPMs. We first discussed the use of non-linearity assessment to identify the occurrence of particular biological phenomena, and then processed a strategy to conduct 'module by module analysis' in order to comprehensively integrate different SA methods and indices when a complex biophysical system characterized by the interaction of several processes described by sub-models/modules is analyzed.
- We applied the developed methodology of sensitivity analysis to 3 FSPMs with different levels of complexity, and inferred in each case what information can be drawn from this analysis. Better understanding of source-sink dynamics and internal driving forces during plant growth are achieved. Especially for NEMA model, 'module by module analysis' helped to understanding the model behavior from the classical simulation approach.



**Fig. 8.1 :** Brief illustration of improved Sobol's estimator. 'Sobol 1993' see [Sobol, 1993], 'Homma et Saltelli' see [Homma and Saltelli, 1996]. Matrices in the frame of the same line type are involved in one factor index estimation.  $\mathbf{y}$ : model output of sampling matrix,  $\mathbf{y}_R$ : model output of re-sampling matrix,  $\mathbf{y}'$ : model output of sampling matrix with  $i$ th column re-sampled,  $\mathbf{y}'_R$ : model output of re-sampling matrix with  $i$ th column re-sampled.

## Computational issue

It is necessary for us to use a precise global sensitivity analysis method like Sobol's to locate the quantitative interaction information between parameters. However FSPMs usually have a large number of parameters and the model evaluation is computationally heavy, so the implementation of the SA strategy faces a great challenge regarding the computing cost issue. It is crucial to not only devise efficient computing techniques, in order to make best use of model evaluations, but also to have a good error estimation to check whether the SA computing has properly converged.

Based on the idea of 'making best use of model evaluations' and making the sampling-resampling processing 'smoother' by 'averaging', we proposed a new Sobol's estimator to improve the computing efficiency and convergence of the method. For the Homma-Saltelli method, to estimate each of the  $k$  first order indices, the computing cost is  $N(k + 2)$  model runs, with  $N$  model runs for the initial sampling matrix output evaluation  $\mathbf{y}$ ,  $N$  model runs for the initial re-sampling matrix output evaluation  $\mathbf{y}_R$ ,  $Nk$  model runs for the re-sampling matrix generated for conditional variance of output  $\mathbf{y}'_R$ . What we proposed is to add the complementary 'sampling' matrix  $\mathbf{y}'$  into the evaluation which increases the number of sampling matrices involved for the SA index for one factor from 3 to 4. With the 4 matrices, when the role of 'sampling' and 're-sampling' matrices change, we get two sets of SA index evaluations. As in fig.8.1, the set of matrices within real line frame and dashed line frame are two sets of SA index evaluations. When we average the two sets of indices, since the sampling gets more 'balanced', a better convergence has been achieved by the improved estimator. Plus, by adding with  $Nk$  more model runs, that is to say with the computing cost of  $N(2k + 2)$  model runs while we make full use of the  $2N$  samplings, two sets of SA indices estimation can be done, so that some computing cost has been saved. The computing test on Ishigami function has proved to yield the better performance of our new Sobol's estimator.

Previous work as in [Homma and Saltelli, 1996] gave interesting result about error

estimation, but the conclusions are based on some restrictive assumptions. We also derived for the generic type of computational methods as illustrated in fig.8.1 an error estimation of sensitivity indices with respect to the sampling size. It is based on some rules to compute ‘the moments of functions of random variables’. It allows a detailed control of the balance between accuracy and computing time and can also be used to give confidence bounds of the result.

## Strategy design

A good sensitivity analysis practice does not only need well designed estimators (which is one of the thesis objectives as recalled in the previous paragraph), but it also requires a good understanding of the issues that can be tackled by different methods, and how these methods can be combined to benefit from their respective advantages. These considerations underlies the strategy design for SA.

Typically to show how to make best use of SRC for FSPMs, we discussed the non-linearity assessment based on a side result of SRC: the determination coefficient  $R^2$ . The knowledge of the intrinsic non-linearity of the model and of its dynamic evolution throughout plant growth identified by  $R^2$  proved to be useful to study model behavior and properties, to underline the occurrence of particular biological phenomena or to improve the statistical analysis when confronting models to experimental data.

For complex biological systems which are characterized by several interacting processes with sub-modules describing each of them, especially for some FSPMs, the strategy design should be divided into several steps for which we choose different SA methods to fulfill different requirements. An interesting point is the evaluation of the importance of the ‘function’ modules of FSPMs, which means checking the sensitivity effects of the groups of parameters. This is how ‘module by module’ analysis for complex biophysical systems is proposed.

In our ‘module by module’ analysis, the following sensitivity indices and some of their combinations are used. We recall the type of information the different indices and methods provide:

- Non-linearity index from SRC: The level of non-linearity evolution of the model and the general view of the model dynamic property.
- Sobol’s first order index  $S_i$ : The main effect of factor  $X_i$ , mainly used for ‘Factor Priorization’.
- Sobol’s total order index  $ST_i$ : The total effect of factor  $X_i$ , mainly used for ‘Factor fixing’ or screening.
- The difference between  $ST_i$  and  $S_i$ : The sum of interactions of all orders between  $X_i$  and the other factors.

- Sobol's second order index  $S_{ij}$ : The interaction between factors  $X_i$  and  $X_j$ .
- First order group indices  $S_{\Omega_i}^g$ : The main effect of the set of factors  $\Omega_i$  which usually gathers the inputs of one module.
- Total order group index  $ST_{\Omega_i}^g$ : The total effect of factor set  $\Omega_i$ .
- The difference between  $ST_{\Omega_i}^g$  and  $S_{\Omega_i}^g$ : The sum of interactions of all orders between factor set  $\Omega_i$  and the other sets.
- Second order group indices  $S_{\Omega_{ij}}^g$ : The interaction between factor sets  $\Omega_i$  and  $\Omega_j$ .

The proposed strategy follows the steps:

- First, linearity analysis gives us the non-linearity stage information of the overall model. This preliminary information is also useful to adapt the following strategy accordingly.
- Second, group analysis provides the evolution of module importance  $S_{\Omega_i}^g$ , and helps us to check inter-module interactions by  $ST_{\Omega_i}^g - S_{\Omega_i}^g$  and  $S_{\Omega_{ij}}^g$ .
- Thirdly, based on the same sampling points, we run the SA module by module, in order to provide index  $ST_i$  of each parameter for screening. It is the basic index for screening procedure but not the only standard. The question of how many parameters should be selected is decided by comprehensively considering the results of the first and second steps. Moreover, by analyzing the composition of  $ST_{\Omega_i}^g$ , we get to know the intra-module and inter-module interaction quantitatively.
- Finally, we run the SA for the overall model with the parameters selected module by module and check the parameter sensitivity indices for the overall model.

Analyses are conducted using Sobol's method and the efficient computation technique derived from [Saltelli, 2002], and several outputs of interest are considered specially for NEMA to check how parameter effects change with the outputs of interest.

Moreover, since we consider a dynamic system, the evolution of the sensitivity indices is computed. When it is related to parameter screening, we use a time averaging index called TGI to reduce the time dimension of the sensitivity indices.

## SA practice for FSPMs

We implemented the strategy developed in this thesis to 3 FSPMs with different levels of complexity, and inferred in each case what information can be drawn from this analysis.

1. Firstly we practiced on a simple source-sink model of maize growth, which is used to specifically study the process of carbon (C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations. We first practiced local SA method and its normalized version, then SRC, lastly Sobol's method. We studied the interest of global sensitivity analysis and its latest developments for the GreenLab model. SA helped us to get better understanding of source-sink dynamics and internal driving forces during plant growth. The following conclusions can be noted:
  - The most important parameters for GreenLab maize are  $\mu$  (energetic efficiency) and  $Sp$  (characteristic surface related to competition between plants). A non-linear period was identified thanks to  $R^2$  index. It indicates a key stage in terms of biophysical processes corresponding to the transition between two allocation phases: the first one corresponding mostly to leaf area increase and the second one to grain filling.
  - The parametric estimation of GreenLab by model inversion relies on multi-stage observations, so the information given by the linearity analysis can be taken advantage of. For maize, it is important to have detailed and frequent observations between cycles 14 and 20, while the measurements can be less frequent after cycle 20.
  - The sensitivity analysis gives us hints on how to improve the calibration process for maize model: first fix all parameters to reasonable values from literature, then find estimates of  $\mu$  and  $Sp$ , then find simultaneously new estimates for the set of  $\alpha_o$  parameters together with  $\mu$  and  $Sp$ , and finally find simultaneously new estimates for the set of sink parameters  $P_o$ , together with the set of  $\alpha_o$  parameters and  $\mu$  and  $Sp$ . The sensitivity analysis indicates that fixing  $\beta_o$  parameters is reasonable since their influence is limited.
  - The interaction information given by SA enlightens the source-sink dynamics and internal driving forces during plant growth. Most of the interactions concentrate between  $\alpha_b$  and the others, in which the interaction between  $\alpha_b$  and  $\alpha_s$  outmatch by large.
  - The understanding of parameter interaction is crucial for genetic improvement. FSPM parameters may be linked to plant genes, and thus may help breeders to design ideotypes. If a parameter has little interaction with others, we can directly concentrate on this trait for the design of ideotypes. If the interaction is strong, it is more complex. If the parameters are strongly genetically related (determination by the same genes), the model parameterization should be improved to take into account this fundamental interaction. If they are not genetically related, breeding strategy should rely on multi-dimensional optimization to handle the interacting processes.
2. Secondly we practiced on the GreenLab model of tree growth (applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis,

which describes tree structural plasticity in response to trophic competition. We inspected 14 parameters related to tree biomass production and allocation, and follow dynamically the sensitivity indices for 50 years. The test case chosen corresponds to a specific model showing alternating patterns in growth phases resulting from the complex interactions between functioning and organogenesis. The sensitivity analysis offered interesting insight in the understanding of this interaction:

- Due to the exponential negative function in the production function in GreenLab, the biomass production is very sensitive to changes in total leaf surface area when it is small, but on the contrary, there is a value beyond which an increase in leaf surface area will induce very little increase in biomass production. In the linear phase, the model is more sensitive to the parameter  $S_p$ . On the contrary, in the saturation phase, the model is more sensitive to the parameter  $\mu$ .
  - In the youth stage of the tree, the trunk starts growing and no branch appear due to the low value of the ratio of biomass to demand. At this point, the model is mainly sensitive to the parameters of phytomers of physiological age 1 ( $(S_B)0$ : blade sink and  $(S_I)0$ :internode sink). The ratio of available biomass to demand increases fast, and several branches appear together. The phytomers of physiological age 4 are the more numerous in the tree as they correspond to the twigs. Their number increases till time step 15 and then oscillates. The model output is sensitive to their sink strengths ( $(S_B)3$ ,  $(S_L)3$  and  $(S_I)3$ ) with the same period.
  - In the phases corresponding to high levels of Q/D, a large number of branches will appear. In the phase of low levels Q/D, the most important parameter is the layer sink for secondary growth since it corresponds to the largest part of plant demand when primary growth is restricted.
3. Lastly we practiced on a functional-structural model, NEMA, describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering. This model is more mechanistic and more complex than the two previous ones. Basic biological modules are identified: namely in our test case Carbon distribution, Nitrogen distribution, Carbon acquisition via photosynthesis, Nitrogen acquisition by roots, Senescence. Several outputs of interest are considered for both intra-module and inter-module analysis: a) total green area of the plant, b) total dry mass production of the plant, c) dry mass of the grains, d) Nitrogen mass of the grains and e) root Nitrogen uptake. Basically, for each output of interest, we did 1) the non-linearity assessment, 2) the analysis of function module rankings, 3) the intra module analysis, 4) the inter module analysis, and finally 5) the overall model analysis based on the result of the 4 analyse before. Some detailed conclusions can be drawn:
- The most important achievement of this practice work for NEMA is that 17 factors out of 83 total can be selected as priority factors for all the 5 outputs

of interest. It is a big step towards the model simplification of NEMA and can potentially lead to more optimized work on parameter estimation in the next modelling step.

- Relying on selecting the factors according to  $S_i$  ranking in the intra-module analysis is not appropriate. The right way is to check the  $ST_i$  ranking on the basis of the same sampling points for each module analysis, plus comprehensively considering the group analysis for each module.
- The most important factors for module Nfluxes are fairly steady for all the outputs: 1)  $\gamma$  (relative rate of potential grain N filling during cell division), 2)  $\sigma_{La}^{N_{ph}}$  (relative rate of photosynthetic N synthesis associated to xylem influx for entity Lamina), 3)  $\delta_{La}^N$  (relative degradation rates of remobilizable N for entity Lamina), in which  $\gamma$  rules all the outputs and most of the time has the very high ranking.  $\gamma$  appears to be a crucial factor for the general NEMA model. For the 28 factors in this module, the sensitivity is very concentrated on these 3 ones.
- For module DMfluxes, the later statement is not appreciate. The 12 factors for DMfluxes have different rankings for different outputs. Generally, 1)  $tt_r^{Macc}$  (the duration during which roots can accumulate dry mass) 2)  $tt_g^{Macc}$  (the duration during which grains can accumulate dry mass), those two rank among the most important for all the 5 outputs. Secondly, 1)  $\sigma_g^M$  (relative sink strength of grains), 2)  $\sigma_r^M$  (relative sink strength of roots), 3)  $\alpha_r, \beta_r$  (Beta function indices for roots), have also a noticeable effect on 4 of the outputs.
- Compared to modules DMfluxes and Nfluxes, modules Tissuedeath and Photosynthesis tend to have specific important parameters for different outputs, like  $d_{La}$  (proportion of maximum specific N mass at which tissues die for Lamina) for output of AreaGreenTotal and  $\omega_{La,2}$  (proportion coefficient linking photosynthesis at saturating PAR and N mass per unit) specially for output DMgrains.
- If we categorize the factors by organs, for module DMfluxes, the ones for organ grain and root tend to have a significant effect on all outputs. The parameter characteristic of lamina tend to have the control over Nfluxes, Tissuedeath and Photosynthesis.
- The module RootNuptake has very little influence on the 5 outputs of interest. It questions the modelling part for this module and invites to work again on the structure and parameterization of this module.

## 8.2 Discussion and perspectives

Sensitivity analysis (SA) is a fundamental tool in the building, use and understanding of mathematical models [Saltelli et al., 2008]. Though we have done some work on this topic in this thesis, a lot more is expected for us to keep improving SA's role in FSPMs community.

## Evaluation of the distribution of the inputs

Most of the time, the distributions of input factors are not easy to obtain. A possible strategy is to perform an initial exploratory analysis with rather crude definition for the distribution of the inputs and use sensitivity analysis to identify the most important inputs; then, resources can be concentrated on characterizing the uncertainty of these inputs and a second sensitivity analysis can be carried out with these improved uncertainty characterizations [Helton et al., 2006b]. In this thesis, we have implemented the first step. A study to explore the variability of parameters in a family of genotypes for the SUNFLO model [Lecoeur et al., 2011] will be implemented in our future work.

## Computational issue

Fighting with computation cost is a key issue in the case of a model with a huge number of parameters as potentially for FSPMs. Though we have proposed an improved Sobol's estimator to achieve better computing efficiency and convergence characteristics compared to its previous version, there is still room for improvement. In a recent paper [Saltelli et al., 2010], the authors compared different types of estimators for  $ST_i$ , and Jansen's estimator [Jansen, 1999] is shown to be the most efficient. It would thus be interesting to adapt our strategy to make best use of model evaluations to other estimators to check whether the same convergence characteristics can be obtained. A proper sampling design is also an important aspect of computational issue. It relates to the convergence properties of the Monte Carlo simulation [Tarantola et al., 2012]. Thus it is also interesting for us to investigate which type of sampling design is the most appropriate for our SA practice in FSPM.

## Parallel computing

As the main bottleneck of the computing cost is model evaluation for the output matrices, and model evaluations are actually independent from each other, it would be interesting to use parallel computing to reduce the computing time for sensitivity analysis. Efficient exchange of data would still be a challenge. A study on this issue is in process in the Digiplante team.

## Sensitivity analysis and model selection

The estimation of the uncertain parameters from experimental data is an important step and model performances depend a lot on the accuracy of the parameter estimation ([Butterbach-Bahl et al., 2004]; [Gabrielle et al., 2006]; [Lehuger et al., 2009]; [Makowski et al., 2006]). In general, it is impossible to estimate all parameters of complex models simultaneously [Bechini et al., 2006]. During the process of parameter estimation from experimental data, there is not usually direct convergence to the proper set of parameters because of the non-convexity of the generalized least-square

function used as fitting criterion. Moreover, it was shown that the confidence interval on the estimated parameters might be improved by fixing some parameters [Guo et al., 2006], [Lamboni et al., 2011]. A common strategy consists in selecting a subset of parameters to be calibrated using sensitivity analysis, and fixing the others to some nominal values ([Monod et al., 2006]; [Wallach et al., 2002]).

Parameters fixed as recommended values are treated as constants in models. Therefore the problem to choose parameters needed to be estimated turns out to be a statistical model selection problem in which candidate models may have different numbers of parameters. Thus, model selection procedure need to be done after processing parameter estimation by the help of sensitivity analysis for ‘Factor prioritization’ so that the model can be validated.

## **Sensitivity analysis with dynamic output and multi-variate outputs**

Dynamic functional structural plant models frequently simulate state variables across discrete time step. The dynamic structure of these models introduces a strong temporal correlation between the different model outputs. For a discrete-time model, global sensitivity analysis methods can be applied sequentially at each simulation time step. The sequential implementation of global sensitivity analysis at each time step can result in several hundreds of sensitivity indices, with one index per time step. It is not easy to identify the most important parameters based on such a large number of values [Campolongo et al., 2007]. In addition, there is a high level of redundancy from the discrete-time outputs.

Another problem is that when we analyze one model, usually, there is not only one model output of interest. Generally there are several variables of interest, as usually in more general complex systems [Zio, 2003]. Therefore, one factor has several different sensitivity indices for the different outputs. It is difficult to consider all the indices for all outputs for a given factor to evaluate its effect on the whole model.

We used in this thesis a unitary concept based on the averaging across time called ‘Time Generalized Index (TGI)’ to resolve the problem of time dimension redundancy of the SA indices. Since the final result we obtained in NEMA is that all the analysis for different outputs of interest underlines the important effect to a relatively similar set of factors, we did not consider a synthetic index to summarize the indices for all types of outputs.

In [Campbell et al., 2006], the author used a principal component analysis of output temporal curves, then compute sensitivity indices of each input on each principal component coefficient; in [Lamboni et al., 2011] they developed the multivariate global sensitivity analysis method. It allows to aggregate the different sensitivity indices of the principal component coefficients in a unique index, called the generalized sensitivity index. Each generalized sensitivity index synthesizes the effects of uncertain

factors on all the dynamic outputs obtained from dynamic models, and thus explains the influence of the corresponding factor on all outputs at all time steps. It will be interesting to use this kind of index for the analysis of a broader range of models.

## Deterministic and stochastic models

All the models we considered in this thesis are deterministic models and the methods we used are mostly appropriate for this type of model. However, in the FSPM modelling community, stochastic models have been widely developed in the past decades [Kang et al., 2008] [Loi and Cournède, 2008] [Pallas et al., 2011]. They also need sensitivity analysis to help for the parameter prioritization or for model simplification. There are several references about this issue [Ginot et al., 2006], [Lurette et al., 2009].

## Correlated input factors

Though it is not easy to know the correlation between input parameters, it may be of interest and was not considered in this thesis. In [Saltelli and Tarantola, 2002], the authors devised a strategy for sensitivity analysis that could work for correlated input factors, based on the first-order and total-order indices from variance decomposition. Specially, we will use this strategy to investigate SA of plant models with uncertain inputs representing genotype parameters [Buck-Sorlin et al., 2005] in a family linked through correlation matrix [Letort et al., 2008b].

## Influence of the input uncertainty on the output distribution

Sensitivity analysis offers interesting insight for the understanding of the importance ranking and interactions of factors. Moreover, by using variance-based techniques, an analyst is capable not only of obtaining the parameter contribution to the output variance but also of gaining insights on the model structure by using moment-independent indicators [Borgonovo, 2006]. Such methods provide indications of the influence of the input uncertainty on the output distribution [Borgonovo, 2007]. It would be interesting for our future work to gain insight about the output distribution (in our tree modelling case, for the variables describing tree height and diameter for example).

## 8.3 To conclude

In this thesis, we attempt to explore and demonstrate the benefit of SA to the FSPM community in which we still find papers in the recent years processing SA by only considering local methods [Evers et al., 2007]. The potential use of SA in the plant modelling process is very interesting, both in terms of model parameterization, but also as a tool for model diagnosis. An effort to promote the recent methods and make them easily available in the FSPM community is made through the platform

---

PyGMAIion for plant modelling, which is now used in some other plant modelling groups specifically for SA.



# APPENDIX



# A. APPROXIMATION THE MOMENTS OF FUNCTIONS OF RANDOM VARIABLES

It is possible to approximate the moments of a function  $f$  of random variables  $X$ ,  $Y$  using Taylor expansions, provided that  $f$  is sufficiently differentiable and that the moments of  $X$  and  $Y$  are finite [Bevington and Robinson, 2003]. Let us denote the expectations  $E(X) = \mu_X$ ,  $E(Y) = \mu_Y$ , and variances  $V(X) = \sigma_X^2$ ,  $V(Y) = \sigma_Y^2$ . We thus get the following results:

**Theorem A.0.1:** Suppose  $Z = f(X)$ , then

$$E(Z) \approx f(\mu_X) + \frac{f''(\mu_X)}{2}\sigma_X^2 \quad (\text{A.1})$$

$$V(Z) \approx [f'(\mu_X)\sigma_X]^2 \quad (\text{A.2})$$

**Proof** Make the second order Taylor expansion for  $Z = f(X)$  at the point of  $X = \mu$ , get:

$$Z = f(X) = f(\mu) + f'(\mu)(X - \mu) + \frac{f''(\mu)}{2}(X - \mu)^2 + R_2(X)$$

where  $R_2(X)$  is the remainder for the second order Taylor expansion, ignore this remainder, then:

$$Z \approx f(\mu) + f'(\mu)(X - \mu) + \frac{f''(\mu)}{2}(X - \mu)^2$$

Add the expectation calculator to it:

$$\begin{aligned} E(Z) &\approx E\left[f(\mu) + f'(\mu)(X - \mu) + \frac{f''(\mu)}{2}(X - \mu)^2\right] \\ &= E[f(\mu)] + E[f'(\mu)(X - \mu)] + E\left[\frac{f''(\mu)}{2}(X - \mu)^2\right] \\ &= f(\mu) + f'(\mu)E(X - \mu) + \frac{f''(\mu)}{2}E(X - \mu)^2 \\ &= f(\mu) + \frac{f''(\mu)}{2}\sigma^2 \end{aligned}$$

Make the first order Taylor expansion for  $Z = f(X)$  at the point of  $X = \mu$ , get:

$$Z = f(X) = f(\mu) + f'(\mu)(X - \mu) + R_1(X)$$

where  $R_1(X)$  is the remainder for the first order Taylor expansion, ignore this remainder, then:

$$Z \approx f(\mu) + f'(\mu)(X - \mu)$$

Add the variance calculator to it:

$$\begin{aligned} V(Z) &\approx V[f(\mu) + f'(\mu)(X - \mu)] \\ &= V[f(\mu)] + V[f'(\mu)(X - \mu)] \\ &= 0 + f'^2(\mu)V(X - \mu) \\ &= f'^2(\mu)V(X) \\ &= f'^2(\mu)\sigma^2 \end{aligned}$$

**Theorem A.0.2:** suppose  $Z = f(X, Y)$ , then

$$\begin{aligned} E(Z) &\approx f(\mu_X, \mu_Y) \\ &+ \frac{1}{2}f_{xx}(\mu_X, \mu_Y)\sigma_X^2 \\ &+ f_{xy}(\mu_X, \mu_Y)cov(X, Y) \\ &+ \frac{1}{2}f_{yy}(\mu_X, \mu_Y)\sigma_Y^2 \end{aligned} \tag{A.3}$$

$$\begin{aligned} V(Z) &\approx [f_x(\mu_X, \mu_Y)\sigma_X]^2 \\ &+ [f_y(\mu_X, \mu_Y)\sigma_Y]^2 \\ &+ 2f_x(\mu_X, \mu_Y)f_y(\mu_X, \mu_Y)cov(X, Y) \end{aligned} \tag{A.4}$$

**Proof** Make the second order Taylor expansion for  $Z = f(X, Y)$  at the point of  $(\mu_X, \mu_Y)$ , get:

$$\begin{aligned} Z &= f(X, Y) \\ &= f(\mu_X, \mu_Y) + [f_x(\mu_X, \mu_Y)(X - \mu_X) + f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &+ \frac{1}{2}[f_{xx}(\mu_X, \mu_Y)(X - \mu_X)^2 + 2f_{xy}(\mu_X, \mu_Y)(X - \mu_X)(Y - \mu_Y) + f_{yy}(\mu_X, \mu_Y)(Y - \mu_Y)^2] \\ &+ R_2(X, Y) \end{aligned}$$

where  $R_2(X, Y)$  is the remainder for the second order Taylor expansion, ignore this remainder, add the expectation calculator to it, then:

$$\begin{aligned} E(Z) &\approx E[f(\mu_X, \mu_Y)] + E[f_x(\mu_X, \mu_Y)(X - \mu_X) + f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &+ \frac{1}{2}E[f_{xx}(\mu_X, \mu_Y)(X - \mu_X)^2 + 2f_{xy}(\mu_X, \mu_Y)(X - \mu_X)(Y - \mu_Y) + f_{yy}(\mu_X, \mu_Y)(Y - \mu_Y)^2] \\ &= f(\mu_X, \mu_Y) + 0 + \frac{1}{2}f_{xx}(\mu_X, \mu_Y)E[(X - \mu_X)^2] + f_{xy}(\mu_X, \mu_Y)E[(X - \mu_X)(Y - \mu_Y)] \\ &+ \frac{1}{2}f_{yy}(\mu_X, \mu_Y)E[(Y - \mu_Y)^2] \\ &= f(\mu_X, \mu_Y) + \frac{1}{2}f_{xx}(\mu_X, \mu_Y)\sigma_X^2 + f_{xy}(\mu_X, \mu_Y)cov(X, Y) + \frac{1}{2}f_{yy}(\mu_X, \mu_Y)\sigma_Y^2 \end{aligned}$$

Make the first order Taylor expansion for  $Z = f(X, Y)$  at the point of  $(\mu_X, \mu_Y)$ , get:

$$Z = f(X, Y) = f(\mu_X, \mu_Y) + f_x(\mu_X, \mu_Y)(X - \mu_X) + f_y(\mu_X, \mu_Y)(Y - \mu_Y) + R_1(X, Y)$$

where  $R_1(X, Y)$  is the remainder for the first order Taylor expansion, ignore this remainder, add the expectation calculator to it, then:

$$\begin{aligned} V(Z) &\approx V[f(\mu_X, \mu_Y) + f_x(\mu_X, \mu_Y)(X - \mu_X) + f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &= V[f_x(\mu_X, \mu_Y)(X - \mu_X) + f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &= V[f_x(\mu_X, \mu_Y)(X - \mu_X)] + 2E[f_x(\mu_X, \mu_Y)(X - \mu_X)f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &\quad + V[f_y(\mu_X, \mu_Y)(Y - \mu_Y)] \\ &= f_x^2(\mu_X, \mu_Y)V(X - \mu_X) + 2f_x(\mu_X, \mu_Y)f_y(\mu_X, \mu_Y)E[(X - \mu_X)(Y - \mu_Y)] \\ &\quad + f_y^2(\mu_X, \mu_Y)V(Y - \mu_Y) \\ &= f_x^2(\mu_X, \mu_Y)\sigma_X^2 + 2f_x(\mu_X, \mu_Y)f_y(\mu_X, \mu_Y)cov(X, Y) + f_y^2(\mu_X, \mu_Y)\sigma_Y^2 \end{aligned}$$

In particular,

- if  $Z = X^2$ , then

$$E[Z] = \mu_X^2 + \sigma_X^2 \quad (\text{A.5})$$

$$V[Z] \approx 4\mu_X^2\sigma_X^2 \quad (\text{A.6})$$

- and if  $Z = X \pm Y$ , then:

$$E[Z] = \mu_X \pm \mu_Y \quad (\text{A.7})$$

$$V[Z] = \sigma_X^2 + \sigma_Y^2 \pm 2cov(X, Y) \quad (\text{A.8})$$

- and if  $Z = X/Y$ , then:

$$E[Z] \approx \frac{\mu_X}{\mu_Y} - \frac{1}{\mu_Y^2}cov(X, Y) + \frac{\mu_X \cdot \mu_Y}{\mu_Y^4} \cdot \sigma_Y^2 \quad (\text{A.9})$$

$$V[Z] \approx \frac{1}{\mu_Y^2} \cdot \sigma_X^2 - 2 \cdot \frac{\mu_X}{\mu_Y^3} \cdot cov(X, Y) + \frac{\mu_X^2}{\mu_Y^4} \cdot \sigma_Y^2 \quad (\text{A.10})$$



# BIBLIOGRAPHY

- M. Allen, P. Prusinkiewicz, and T. Dejong. Using L-systems for modeling source-sink interactions, architecture and physiology of growing trees, the L-peach model. *New Phytologist*, 166:869–880, 2005.
- G. Andrianov, S. Burriel, S. Cambier, A. Dutfoy, I. Dutka-Malen, E. de Rocquigny, and B. Sudret. Open turns, an open source initiative to treat uncertainties, risks'n statistics in a structured industrial approach. In *Proceedings of ESREL 2007: The 18th European Safety and Reliability Conference*, 2007. ESREL2007, Stavenger, Norway.
- D. Barthélémy and Y. Caraglio. Plant architecture: a dynamic, multilevel and comprehensive approach of plant form, structure and ontogeny. *Annals of Botany*, 99:375–407, 2007.
- L. Bechini, S. Bocchi, T. Maggiore, and R. Confalonieri. Parameterization of a crop growth and development simulation model at sub-model components level. an example for winter wheat (*triticum aestivum* l.). *Environmental Modelling & Software*, 21(7):1042 – 1054, 2006.
- J. Bertheloot, P. Martre, and B. Andrieu. Dynamics of light and nitrogen distribution during grain filling within wheat canopy. *Plant Physiology*, 148(3):1707–1720, 2008.
- J. Bertheloot, P.-H. Cournède, and B. Andrieu. Nema, a functional-structural model of nitrogen economy within wheat culms after flowering. i. model description. *Annals of Botany*, 108(6):1085–1096, 2011a.
- J. Bertheloot, Q.-L. Wu, P.-H. Cournède, and B. Andrieu. Nema, a functional-structural model of nitrogen economy within wheat culms after flowering. ii. evaluation and sensitivity analysis. *Annals of Botany*, 108(6):1097–1109, 2011b.
- P.-R. Bevington and D.-K. Robinson. *Data reduction and error analysis for the physical sciences*. Boston, MA: McGraw-Hill, 2003.
- B. Blocken and C. Gualtieri. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21(5):602 – 614, 2006.
- E. Borgonovo. Measuring uncertainty importance: Investigation and comparison of alternative approaches. *Risk Analysis*, 26:1349–1361, 2006.

- E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771 – 784, 2007.
- G. Buck-Sorlin and K. Bachmann. Simulating the morphology of barley spike phenotypes using genotype information. *Agronomie*, 20:691–702, 2000.
- G. Buck-Sorlin, O. Kniemeyer, and W. Kurth. Barley morphology, genetics and hormonal regulation of internode elongation modelled by a relational growth grammar. *New Phytologist*, 166(3):859–867, 2005.
- K. Butterbach-Bahl, M. Kesik, P. Miehele, H. Papen, and C. Li. Quantifying the regional source strength of n-trace gases across agricultural and forest ecosystems with process based models. *Plant and Soil*, 260:311–329, 2004.
- D. Cacuci. *Sensitivity and Uncertainty Analysis: Theory*, volume I. Chapman & Hall, 2003.
- D. Cacuci, I.-B. Mihaela, and N. Michael. *Sensitivity And Uncertainty Analysis: Applications to Large-Scale Systems*, volume II. Chapman & Hall, 2005.
- K. Campbell, M. D. McKay, and B. J. Williams. Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, 91(10-11):1468–1472, 2006.
- F. Campolongo and R. Braddock. The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering & System Safety*, 64(1):1 – 12, 1999.
- F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22:1509–1518, 2007.
- J. Cariboni, D. Gatelli, R. Liska, and A. Saltelli. The role of sensitivity analysis in ecological modelling. *Ecological modelling*, 203:167–182, 2007.
- M. Chelle. Phylloclimate or the climate perceived by individual plant organs: What is it? how to model it? what for? *New Phytologist*, 166(3), 2005.
- P.-H. Cournède, V. Letort, A. Mathieu, M.-Z. Kang, S. Lemaire, S. Trevezas, F. Houllier, and P. de Reffye. Some parameter estimation issues in functional-structural plant modelling. *Mathematical Modelling of Natural Phenomena*, 6(2):133–159, 2011.
- P.-H. Cournède, M.-Z. Kang, A. Mathieu, J.-F. Barczi, H.-P. Yan, B.-G. Hu, and P. de Reffye. Structural Factorization of Plants to Compute their Functional and Architectural Growth. *Simulation*, 82(7):427–438, 2006.
- R. A. Cropp and R. D. Braddock. The new morris method: an efficient second-order screening method. *Reliability Engineering & System Safety*, 78(1):77 – 83, 2002.

- R. Cukier, C. Fortuin, K. Shuler, A. Petschek, and J. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. *J. Chem. Phys.*, 59 (i):3873–3878, 1973.
- R. Cukier, H. Levine, and K. Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. *J. Chem. Phys.*, 26(1), 1978.
- P. de Reffye and B.-G. Hu. Relevant choices in botany and mathematics for building efficient dynamic plant growth models: the greenlab case. In B. Hu and M. Jaeger, editors, *Plant Growth Models and Applications*, pages 87–107. Tsinghua University Press and Springer, 2003.
- P. De Reffye, F. Fourcaud, T. and Blaise, D. Barthélémy, and F. Houllier. A functional model of tree growth and tree architecture. *Silva Fennica*, 31(3):297–311, 1997.
- P. De Reffye, E. Heuvelink, D. Barthélémy, and P.-H. Cournède. Plant growth models. In S. Jorgensen and B. Fath, editors, *Ecological Models. Vol. 4 of Encyclopedia of Ecology (5 volumes)*, pages 2824–2837. Elsevier (Oxford), 2008.
- P. de Reffye, M.-Z. Kang, J. Hua, and D. Auclair. Stochastic modelling of tree annual shoot dynamics. *Annals of Forest Science*, 69(2), 2012.
- E. De Rocquigny, N. Devictor, and S. Tarantola. *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley&Sons, the primer edition, 2008.
- N. Draper and H. Smith. *Applied Regression Analysis*. John Wiley & Sons, New York, 1981.
- J.-L. Drouet and L. Pagès. Graal-cn: A model of growth, architecture and allocation for carbon and nitrogen dynamics within whole plants formalized at the organ level. *Ecological Modelling*, 206(3-4):231 – 249, 2007.
- J. Evers, J. Vos, M. Chelle, B. Andrieu, C. Fournier, and P. Struik. Simulating the effects of localized red: far-red ratio on tillering in spring wheat (*Triticum aestivum*) using a three-dimensional virtual plant model. *New Phytologist*, 176(2): 325–336, 2007.
- J. Evers, J. Vos, X. Yin, P. Romero, P. Van Der Putten, and P. Struik. Simulation of wheat growth and development based on organ-level photosynthesis and assimilate allocation. *Journal of Experimental Botany*, 61(8):2203–2216, 2010.
- P. Federl and P. Prusinkiewicz. Virtual laboratory: An interactive software environment for computer graphics. *Computer Graphics International Conference*, 0:93, 1999.
- T. Fourcaud, X. Zhang, A. Stokes, H. Lambers, and C. Körner. Plant growth modelling and applications: The increasing importance of plant architecture in growth models. *Annals of Botany*, 101(8):1053–1063, 2008.

- C. Fournier and B. Andrieu. A 3d architectural and process-based model of maize development. *Annals of Botany*, 81(2):233–250, 1998.
- C. Fournier, B. Andrieu, S. Ljutovac, and S. Saint-Jean. Adel-wheat: a 3d architectural model of wheat development. In B. Hu and M. Jaeger, editors, *Plant Growth Models and Applications*, pages 54–66. Tsinghua University Press and Springer, 2003.
- B. Gabrielle, P. Laville, O. Duval, B. Nicoullaud, J. C. Germon, and C. Hénault. Process-based modeling of nitrous oxide emissions from wheat-cropped soils at the subregional scale. *Global Biogeochemical Cycles*, 20(GB4018), 2006.
- V. Ginot, S. Gaba, R. Beaudouin, F. Aries, and H. Monod. Combined use of local and anova-based global sensitivity analyses for the investigation of a stochastic dynamic model: Application to the case study of an individual-based model of a fish population. *Ecological Modelling*, 193(3-4):479 – 491, 2006.
- C. Godin, J. Hanan, W. Kurth, A. Lacoïnte, A. Takenaka, P. Prusinkiewicz, T. De-Jong, C. Beveridge, and B. Andrieu. Proceedings of the fourth international workshop on functional-structural plant models (fspm04). 2004.
- A. Grievank. *Evaluating derivatives, Principles and techniques of algorithmic differentiation*. SIAM publisher, 2000.
- Y. Guo, Y. Ma, Z. Zhan, B. Li, M. Dinkuhn, D. Luquet, and P. de Reffye. Parameter optimization and field validation of the functional-structural model greenlab for maize. *Annals of Botany*, 97:217–230, 2006.
- M. Hafsi, W. Mechmeche, L. Bouamama, A. Djekoune, M. Zaharieva, and P. Monneveux. Flag leaf senescence, as evaluated by numerical image analysis, and its relationship with yield under drought in durum wheat. *Journal of Agronomy and Crop Science*, 185(4):275–280, 2000.
- G. Hammer, M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, and D. Podlich. Models for navigating biological complexity in breeding improved crop plants. *Trends in plant science*, 11(12):587–593, 10 2006.
- J. Helton, J. Johnson, C. Salaberry, and C. Storlie. Survey of sampling based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety*, 91:1175–1209, 2006a.
- J. Helton, J. Johnson, C. Sallaberry, and C. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10-11):1175 – 1209, 2006b. The Fourth International Conference on Sensitivity Analysis of Model Output-SAMO 2004.
- R. Hemmerling, O. Kniemeyer, D. Lanwert, W. Kurth, and G. Buck-Sorlin. The rule-based language xl and the modelling environment groimp illustrated with simulated tree competition. *Functional Plant Biology*, 35(10):739, 2008.

- T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52:1–17, 1996.
- G. M. Hornberger and R. C. Spear. An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 12:7–18, 1981.
- T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. In *Proceedings of the ISUMA'90, First International Symposium on Uncertainty Modelling and Analysis*, pages 398–403, University of Maryland, USA, 1990.
- M. J. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1-2):35 – 43, 1999.
- S. Julier, J. Uhlmann, and H. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3), 2000.
- M. Kang, P.-H. Cournède, P. de Reffye, D. Auclair, and B. Hu. Analytical study of a stochastic plant growth model: application to the greenlab model. *Mathematics and Computers in Simulation*, 78(1):57–75, 2008.
- M. Koda, G. J. Mcrae, and J. H. Seinfeld. Automatic sensitivity analysis of kinetic mechanisms. *International Journal of Chemical Kinetics*, 11(4):427–444, 1979.
- W. Kurth and B. Sloboda. Growth grammars simulating trees: An extension of L-systems incorporating local variables and sensitivity. *Silva Fennica*, 31(3):285–295, 1997.
- A. Lacointe. Carbon allocation among tree organs: A review of basic processes and representation in functional-structural tree models. *Ann. For. Sci.*, 57(5):521–533, 2000.
- M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450–459, 2011.
- X. Le Roux, A. Lacointe, A. Escobar-Gutiérrez, and S. Le Dizès. Carbon-based models of individual tree growth: A critical appraisal. *Annals of Forest Sciences*, 58:469–506, 2001.
- J. Lecoeur, R. Poiré-Lassus, A. Christophe, B. Pallas, P. Casadebaig, P. Debaeke, F. Vear, and L. Guillioni. Quantifying physiological determinants of genetic variation for yield potential in sunflower. sunflo: a model-based analysis. *Functional Plant Biology*, 38:246–259, 2011.
- S. Lehuger, B. Gabrielle, M. van Oijen, D. Makowski, J.-C. Germon, T. Morvan, and C. Hénault. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. *Agriculture, Ecosystems & Environment*, 133(3-4):208–222, 2009.

- V. Letort. *Adaptation of the GreenLab growth model to plants with complex architectures and multi-scale analysis of source-sink relationships for parametric identification*. PhD thesis, École Centrale Des Arts et Manufactures(Ecole Centrale Paris), 2008.
- V. Letort, P.-H. Cournède, A. Mathieu, P. de Reffye, and T. Constant. Parametric identification of a functional-structural tree growth model and application to beech trees (*fagus sylvatica*). *Functional Plant Biology*, 35(10), 2008a.
- V. Letort, P. Mahe, P.-H. Cournède, P. de Reffye, and B. Courtois. Quantitative genetics and functional-structural plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. *Annals of Botany*, 101(8), 2008b.
- C. Loi and P.-H. Cournède. Generating functions of stochastic L-systems and application to models of plant development. *Discrete Mathematics and Theoretical Computer Science Proceedings*, AI:325–338, 2008.
- A. Lurette, S. Touzeau, M. Lamboni, and H. Monod. Sensitivity analysis to identify key parameters influencing salmonella infection dynamics in a pig batch. *Journal of Theoretical Biology*, 258(1):43 – 52, 2009.
- Y. Ma, M. Wen, B. Li, P.-H. Cournède, and P. de Reffye. Calibration of greenlab model for maize with sparse experimental data. In T. Fourcaud and X. Zhang, editors, *The Second International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA06)*. IEEE (Los Alamitos, California), 2007.
- Y. Ma, M. Wen, Y. Guo, B. Li, P.-H. Cournède, and P. de Reffye. Parameter optimization and field validation of the functional-structural model greenlab for maize at different population densities. *Annals of Botany*, 101(8), 2008.
- D. Makowski, C. Naud, M.-H. Jeuffroy, A. Barbottin, and H. Monod. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. *Reliability Engineering & System Safety*, 91(10-11):1142 – 1147, 2006.
- A. Mathieu. *Essai sur la modélisation des interactions entre la croissance d'une plante et son développement dans le modèle GreenLab*. PhD thesis, Ecole Centrale Paris, 2006.
- A. Mathieu, P.-H. Cournède, D. Barthélémy, and P. de Reffye. Rhythms and alternating patterns in plants as emergent properties of a model of interaction between development and functioning. *Annals of Botany*, 101(8):1233–1242, 2008.
- A. Mathieu, P.-H. Cournède, V. Letort, D. Barthélémy, and P. de Reffye. A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition. *Annals of Botany*, 103:1173 – 1186, 2009.

- H. Monod, C. Naud, and D. Makowski. Uncertainty and sensitivity analysis for crop models. In D. Wallach, D. Makowski, and J. Jones, editors, *Working with Dynamic Crop Models*, pages 55–100. Elsevier, 2006.
- M. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174, 1991.
- J. Oakley and A. O’Hagan. Probabilistic sensitivity analysis of complex models: a bayesian approach. *J. Royal Stat. Soc*, B66:751–769, 2004.
- A. Pagano and M. Ratto. An extended use of the sobol’ estimator. In *The Fifth International Conference on Sensitivity Analysis of Model Output- SAMO 2007*, 2007.
- B. Pallas, C. Loi, A. Christophe, P.-H. Cournède, and J. Lecoecur. Comparison of three approaches to model grapevine organogenesis in conditions of fluctuating temperature, solar radiation and soil water content. *Annals of Botany*, 107(5):729–745, 2011.
- T.-B. Pathak, C. Fraisse, C. Messina, and G. Hoogenboom. Use of global sensitivity analysis for cropgro cotton model development. *Transactions of the ASABE*, 50(6): 2295–2302, 2007.
- J. Perttunen, R. Sievänen, and E. Nikinmaa. Lignum: a model combining the structure and the functioning of trees. *Ecological Modelling*, 108(1-3):189–198, 1998.
- C. Pradal, S. Dufour-Kowalski, F. Boudon, C. Fournier, and C. Godin. Openalea: a visual programming and component-based software platform for plant modelling. *Functional Plant Biology*, 35(11):751–760, 2008.
- R. Qi, Y. Ma, B. Hu, P. De Reffye, and P.-H. Cournède. Optimization of source-sink dynamics in plant growth for ideotype breeding: a case study on maize. *Computers and Electronics in Agriculture*, 71(1):96–105, 2010.
- J. C. Refsgaard and H. J. Henriksen. Modelling guidelines: terminology and guiding principles. *Advances in Water Resources*, 27(1):71 – 82, 2004.
- J. Ricard. What do we mean by biological complexity? *C.R. Biologies*, 326:133–140, 2003.
- P. Room, J. Hanan, and P. Prusinkiewicz. Virtual plants: new perspectives for ecologists, pathologists and agricultural scientists. *Trends in Plant Science*, 1(1):33–38, Jan. 1996.
- F. Ruget, N. Brisson, R. Delécolle, and Faivre.R. Sensitivity analysis of a crop simulation model, stics, in order to choose the main parameters to be estimated. *Agronomie*, 22:133–158, 2002.
- A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297, 2002.

- A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of American Statistical Association*, 97:702–709, 2002.
- A. Saltelli, T. Andres, and T. Homma. Sensitivity analysis of model output: An investigation of new techniques. *Computational Statistics & Data Analysis*, 15(2): 211 – 238, 1993.
- A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, 2004.
- A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo. Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*, 91: 1109–1125, 2006.
- A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis*. John Wiley&Sons, the primer edition, 2008.
- A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.
- H. Scholten, R. Van Waveren, and S. Groot. Improving the quality of model-based decision support: good modelling practice in water management. In *Congress regional management of water resources*, Maastricht, July 2001. IAHS Press, Wallingford.
- K. Shinozaki, K. Yoda, K. Hozumi, and T. Kira. A quantitative analysis of plant form - the pipe model theory i. basic analysis. *Japanese Journal of Ecology*, 14:97–105, 1964.
- R. Sievänen, E. Nikinmaa, P. Nygren, H. Ozier-Lafontaine, J. Perttunen, and H. Hakula. Components of a functional-structural tree model. *Annals of Forest Sciences*, 57:399–412, 2000a.
- R. Sievänen, E. Nikinmaa, P. Nygren, H. Ozier-Lafontaine, J. Perttunen, and H. Hakula. Components of functional-structural tree models. *Ann. For. Sci.*, 57(5): 399–412, 2000b.
- R. Sievänen, J. Perttunen, E. Nikinmaa, and J. M. Posada. Invited talk: Functional structural plant models-case lignum. In *Proceedings of the 2009 Plant Growth Modeling, Simulation, Visualization, and Applications*, PMA09, pages 3–9, Washington, DC, USA, 2009. IEEE Computer Society.
- H. Sinoquet and X. Le Roux. Short term interactions between tree foliage and the aerial environment: An overview of modelling approaches available for tree structure-function models. *Annals of Forest Sciences*, 57(5):477–496, 2000.

- I. Sobol. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993.
- S. Tarantola, W. Becker, and D. Zeitz. A comparison of two sampling methods for global sensitivity analysis. *Computer Physics Communications*, 183(5):1061 – 1072, 2012.
- R. Team. *R: A Language and Environment for Statistical Computing*, volume I. R Foundation for Statistical Computing, Vienna, Austria, 2009.
- R. Tomovic and M. Vukobratovic. Modern analytical and computational methods in science and mathematics. *General Sensitivity Theory.*, 35, 1972.
- E. Triboi and A.-M. Triboi-Blondel. Productivity and grain or seed composition: a new approach to an old problem-invited paper. *European Journal of Agronomy*, 16(3):163 – 186, 2002.
- R. Van Waveren, S. Groot, H. Scholten, F. Van Geer, J. Wösten, R. Koeze, and J. Noort. *Good Modelling Practice Handbook*, volume 99. STOWA report, 1st edition, 1999.
- A. Varma, M. Morbidelli, and H. Wu. Parametric sensitivity in chemical systems. *General Sensitivity Theory*, 1999.
- J. Vos, L. F. M. Marcelis, P. H. B. de Visser, P. C. Struik, and J. B. Evers. *Functional-Structural Plant Modelling in Crop Production*. Springer Publishing Company, Incorporated, 1st edition, 2007.
- J. Vos, J. B. Evers, G. Buck-Sorlin, B. Andrieu, M. Chelle, and P. H. B. de Visser. Functional-structural plant modelling: a new versatile tool in crop science. *Journal of Experimental Botany*, 61(8):2101–2115, 2010.
- D. Wallach, B. Goffinet, J. Bergez, P. Debaeke, D. Leenhardt, and J. Aubertot. The effect of parameter uncertainty on a model with adjusted parameters. *Agronomie*, 22:159–170, 2002.
- I. Wardlaw. The control of carbon partitioning in plants. *New Phytologist*, 116:341–381, 1990.
- J. Warren-Wilson. Ecological data on dry matter production by plants and plant communities. In E. Bradley and O. Denmead, editors, *The collection and processing of field data*, pages 77–123. Interscience Publishers, New York, 1967.
- A. Y. Weiße. *Global Sensitivity Analysis of Ordinary Differential Equations*. PhD thesis, Freie Universität Berlin, 2009.
- P. Wernecke, J. Müller, T. Dornbusch, A. Wernecke, and W. Diepenbrock. The virtual crop-modelling system 'vica' specified for barley. In J. Vos, L. Marcelis, P. de Visser, P. Struik, and J. Evers, editors, *Functional-structural plant modelling in crop production, Wageningen*, volume Chapter 5, pages 58–69. Springer, 2007.

- Q.-L. Wu and P.-H. Cournède. Sensitivity analysis of greenlab model for maize. In *Proceedings of the 2009 Plant Growth Modeling, Simulation, Visualization, and Applications*, PMA'09, pages 311–318, Washington, DC, USA, 2009. IEEE Computer Society.
- Q.-L. Wu and P.-H. Cournède. The use of sensitivity analysis for the design of functional structural plant models. *Procedia - Social and Behavioral Sciences*, 2(6): 7768 – 7769, 2010. Sixth International Conference on Sensitivity Analysis of Model Output.
- Q.-L. Wu, P.-H. Cournède, and A. Mathieu. An efficient computational method for global sensitivity analysis and its application to tree growth modelling. *Reliability Engineering & System Safety*, In Press, Accepted Manuscript:–, 2011.
- H. Yan, M. Kang, P. De Reffye, and M. Dingkuhn. A dynamic, architectural plant model simulating resource-dependent growth. *Annals of Botany*, 93:591–602, 2004.
- E. Zio. Importance measures of multi-state components in multi-state systems. *International Journal of Reliability, Quality and Safety Engineering (IJRQSE)*, 10(3): 289–310, 2003.
- E. Zio. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94(2):125 – 141, 2009.

# ACKNOWLEDGMENT

First and foremost, I would like to thank each member of my dissertation jury, Prof. Stefano Tarantola, Prof. Gerhard Buck-Sorlin, Prof. Erico Zio, Dr. Amélie Mathieu and Prof. Paul-Henry Cournède, for all of the ideas, feedback and opportunities you have provided to me. Your guidance has helped me a lot and I owe you my heartfelt appreciation. This work was financially supported by INRIA Saclay Île-de-France. Their support is gratefully acknowledged.

Along my doctoral journey, I have been shared it with many people. It has been a great experience to spend several years in the DIGIPLANTE team of laboratory Mathématique Appliquées aux Systèmes (MAS) at École Centrale Paris, and its members will always remain dear to me, no matter those who have left or those who are still in the group.

My first debt of gratitude must go to my advisor, Prof. Paul-Henry Cournède. Thank you for your patience of providing me the vision, encouragement and advise necessary to proceed through the thesis so that I can complete my dissertation. You have been a strong and supportive adviser to me, meanwhile, you have always given me great freedom to pursue independent work. Thank you for passing me your wisdom of being a good researcher, no matter the way of thinking as a researcher, or the way of writing, of being creative and patient when facing difficulties. When I look back of my three years and half PhD under your supervision, it was really a great pleasure to work with you. I used to be in the area of communication and information system during my master's, and I felt lucky that I found the PhD proposal under your supervision several years ago. You helped me to know the brand new but exciting world of plant modelling, what's more, we found together the new interesting points that are presented in the thesis. I believe that our work together next will also be inspiring and interesting.

During this work, I have collaborated with many colleagues for whom I have great regard. I wish I could extend my warmest thanks to all those who have helped me with the work. I own my sincere gratitude to Dr. Amélie Mathieu from AgroParisTech, UMR 1091 EGC Thiverval-Grignon, who helped me with the work of tree models and gave me valuable suggestions for the thesis. Great appreciation also goes to Dr. Jessica Bertheloot from INRA, UMR 0462 Sciences Agronomiques Appliquées à

l'Horticulture, who helped me understand the NEMA model. The work with you was so effective and pleasant. I hope that our close collaborated work can still continue after the thesis. I also want to thank Dr. Véronique Letort, Dr. Qi Rui, Dr. Wang Xiujuan who have always been patient to help me for the problems I encountered about plant knowledge and thank Benoît Bayol and Thomas Guyard for helping me about the programming problem.

I also appreciate so much for my work and life in DIGIPLANTE group during my PhD. All the members of the group deserve my sincerest thanks. Your friendship and assistance has meant more to me than I could ever express. I could not complete my work without invaluable friendly assistance from all of you. My life in France would be bleak without all the diverse activities with you. Thank you for being there when I need you.

I am deeply thankful to my parents who have raised me up and support me all the way. You always thought that you could not really support me in my academic career, but you did. You put so many efforts into making the three of us children in our family appreciate education as the privilege it is. Thank you for giving me such a warm family within which I grew up. My dear sisters, thank you for backing me up without any condition. I am always proud to say that I am the youngest daughter in my family who nearly gets spoiled by the great love all of you give to me. I hope you are proud of me now. My two cute nephews, thank you for bringing me so much happiness with your lovely smiling face. I wish that I could be there for all the important moments of you growing up.

I want to thank my husband for being supportive, encouraging and understanding especially during the tough period in my life. I also feel lucky to have such wonderful parents in law. Thank you for taking me as your daughter, thank you for always being thoughtful and helpful to us.

The list of people I want to thank can never be enough for one acknowledgment. Especially for me who pursuit my PhD as one foreigner, I want to express my thanks to all the people who have given me kind help during my stay in France.

Thanks to all.

Qiongli WU  
Ecole Centrale Paris  
July 2012

# PUBLICATIONS

- Publications in peer-reviewed journal:

Q.-L. Wu, P.-H. Cournède, and A.Mathieu. An efficient computational method for global sensitivity analysis and its application to tree growth modelling. *Reliability Engineering & System Safety*, 2011. In Press, Accepted Manuscript.

J.Bertheloot, Q.-L. Wu, P.-H. Cournède and B.Andrieu. Nema, a functional-structural model of nitrogen economy within wheat culms after flowering. II. Evaluation and sensitivity analysis. *Annals Of Botany* 2011;108(6):1097-1109.

- Publications in peer-reviewed conference/proceedings:

Q.-L. Wu, P.-H. Cournède. A methodology to study Complex Biophysical Systems with Global Sensitivity Analysis. *The European Safety and Reliability Conference (ESREL2011, Troyes, France.) Advances in Safety, Reliability and Risk Management*, CRC press, Taylor&Francis Group, London, page 339, 2011.

Q.-L. Wu, P.-H. Cournède. The use of Sensitivity Analysis for the design of Functional Structural Plant Models. *Sixth international conference on sensitivity analysis of model output(SAMO2010,Milan, Italy.)*, Procedia - Social and Behavioral Sciences, Vol.2, pages 7768-776, 2010.

Q.-L. Wu, J.Bertheloot, A.Mathieu, B.Andrieu, P.-H. Cournède. Assessment of Non-Linearity in Functional-Structural Plant Models. *Sixth International Workshop on Functional-Structural Plant Models(FSPM2010, Davis, California, USA)*, pages 195-198,2010.

Q.-L. Wu, P.-H. Cournède. Sensitivity analysis of GreenLab model for maize. *Third international symposium on Plant Growth and Applications (PMA2009, Beijing, China)*, IEEE Computer Society, pages 311-318, 2009.