



Calibration Bayésienne d'un modèle d'étude d'écosystème prairial : outils et applications à l'échelle de l'Europe

Haythem Ben Touhami

► **To cite this version:**

Haythem Ben Touhami. Calibration Bayésienne d'un modèle d'étude d'écosystème prairial : outils et applications à l'échelle de l'Europe. Autre [cs.OH]. Université Blaise Pascal - Clermont-Ferrand II, 2014. Français. <NNT : 2014CLF22444>. <tel-01136229>

HAL Id: tel-01136229

<https://tel.archives-ouvertes.fr/tel-01136229>

Submitted on 26 Mar 2015

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

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

N° d'ordre : D.U : 2444

EDSPIC : 647

Université Blaise Pascal - Clermont II
École Doctorale
Sciences pour l'Ingénieur de Clermont-Ferrand

Thèse

Présentée par

Ben Touhami Haythem

Pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ

Spécialité : Informatique

**Calibration Bayésienne d'un modèle d'étude d'écosystème
prairial : outils et applications à l'échelle de l'Europe**

Soutenue publiquement le 07/03/2014 devant le jury :

Rapporteur : Marco ACUTIS, Professeur, Université de Milan
Nicolas VIOVY, Chargé de recherche, CEA

Examineur : David HILL, Professeur, Université Blaise Pascal
David MAKOWSKI, Directeur de recherche, INRA

Directeur de thèse : Gianni BELLOCCHI, Directeur de recherche, INRA

Co-encadrent de thèse : Vincent BARRA, Professeur, Université Blaise Pascal

Table des matières

Remerciements	8
Résumé.....	10
Abstract.....	11
Introduction générale	12
Chapitre 1 : Les écosystèmes prairiaux et les changements climatiques.....	14
1.1. Introduction	14
1.2. Les écosystèmes prairiaux.....	14
1.2.1. Diffusion et rôle des prairies en Europe et en France.....	15
1.3. Réponses des écosystèmes prairiaux au changement climatique.....	17
1.3.1. Le changement climatique	17
1.3.2. Contribution de l'agriculture aux émissions de gaz à effet de serre.....	18
1.3.3. Conséquences possibles du changement climatique pour l'agriculture.....	19
1.3.4. La prairie et le changement climatique.....	20
1.4. La modélisation des écosystèmes	22
1.4.1. Modèles agro-écologiques	23
1.5. Description de PaSim	25
1.6. Problématique.....	27
1.7. Conclusion	29
Chapitre 2 : État de l'art : calibration et statistiques Bayésienne	30
2.1. Introduction	30
2.2. Sources d'incertitudes.....	30
2.3. Calibration : techniques et définitions.....	32
2.3.1. Les approches fréquentistes	32
2.3.2. L'approche Bayésienne	36
2.3.3. Méthodes Monte Carlo par Chaînes de Markov	42
2.4. Analyse de sensibilité	47
2.4.1. La méthode de Morris.....	49
2.4.2. Autres méthodes d'analyse de sensibilité.....	51
2.5. Ressources informatiques	53
2.5.1. Ressources matérielles.....	53
2.5.1. Ressources logicielles	54
2.6. Conclusion	55

Chapitre 3 : Calibration Bayésienne de PaSim en conditions de sécheresse	56
3.1. Introduction.....	56
3.2. Matériels et méthodes.....	57
3.2.1. Sites et variables de sortie utilisés	57
3.2.2. Paramètres calibrés.....	61
3.3. Résultats et discussion	65
3.3.1. L'a posteriori des paramètres	66
3.3.2. Analyse du posteriori des variables de sortie.....	71
3.3.3. Evaluation du posteriori des variables de sortie	78
3.4. Conclusion	80
Chapitre 4 : Nouveau paramétrage de PaSim à l'échelle de l'Europe (données à long terme) ...	82
4.1. Introduction.....	82
4.2. Analyse de sensibilité de PaSim	83
4.2.1. Méthodologie	83
4.2.2. Plasticité du modèle	92
4.2.3. Résultats et discussion	93
4.3. Calibration Bayésienne multi-sites à l'échelle de l'Europe	103
4.3.1. Matériels et méthodes	103
4.3.2. Résultats et discussion	106
4.4. Validation du nouveau paramétrage de PaSim.....	113
4.5. Conclusion	121
Conclusion générale	122
Perspectives	124
Bibliographie.....	126
Annexes.....	135
I. Implémentation du code principal de la calibration Bayésienne sous R	135
II. Résultats de l'analyse de sensibilité	136
III. Relevant parameters into the modelling equations.....	143

Table des figures

Figure 1. Part des prairies agricoles dans la couverture des sols en Europe (UE-23, enquête LUCAS du programme de travail d'Eurostat, initiée en 2009, http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home)	16
Figure 2. Changements de la température annuelle entre la période (2016-2035), la période (2046-2065) et la période (2081-2100) (source : IPCC, 2013)	19
Figure 3. Illustration du modèle PaSim [Graux, 2011]	25
Figure 4. Structure du modèle PaSim (entrées + modules)	26
Figure 5. Fonctionnement de l'algorithme de base d'espérance-maximisation [Kung et al., 2005]	35
Figure 6. Principe de la calibration Bayésienne	39
Figure 7. Première étape du principe Bayésien : la connaissance a priori sur les paramètres.....	40
Figure 8. Deuxième étape du principe Bayésien : le calcul de la vraisemblance	41
Figure 9. Troisième étape du principe Bayésien : l'obtention d'une probabilité a posteriori plus précise sur les paramètres	42
Figure 10. Exemple d'une chaîne de valeurs d'un paramètre θ générée par la méthode MCMC de longueur 10 000 : à droite convergence atteinte et à gauche convergence non atteinte	45
Figure 11. Interaction entre modèle, données et expertise grâce à l'approche Bayésienne pour l'amélioration des résultats du modèle et la réduction d'incertitude sur les paramètres	46
Figure 12. Classification des méthodes d'analyse de sensibilité selon la complexité et le temps de calcul	49
Figure 13. Exemple de représentation graphique (diagramme de Morris) de l'analyse de sensibilité des paramètres d'un modèle par rapport à une variable de sortie (émission de N_2O) selon deux indicateurs: la moyenne et l'écart-type de l'indice de sensibilité.	51
Figure 14. Expérimentation (réduction de précipitation à l'aide des rideaux) au site Autrichien de Stubai.....	59
Figure 15. Localisation des deux sites expérimentaux Suisse de Chamau et Frübüel	60
Figure 16. À gauche : expérimentation (réduction de précipitation à l'aide des rideaux), à droite : tour de mesure des flux au site Suisse de Chamau.....	61
Figure 17. À gauche : expérimentation (réduction de précipitation à l'aide des rideaux), à droite : des capteurs de mesures pour les variables liées au sol au site Suisse de Frübüel.....	61
Figure 18. Fonctionnement de la calibration Bayésienne avec PaSim.....	64
Figure 19. À gauche : le nombre d'itérations MCMC obtenues en 35 jours pour la calibration de Frübüel avec (T) et sans traitement (C) ; à droite : le nombre d'itérations MCMC obtenues en 39 jours pour la calibration de Chamau avec (T) et sans traitement (C)	66
Figure 20. Densités de probabilité des paramètres <i>thetasatb</i> (teneur en eau du sol à la saturation) obtenue à Chamau et <i>bb</i> (Campbell's <i>b</i> parameter) obtenue à Frübüel, et leurs valeurs correspondant au maximum de la FDP	66
Figure 21. Exemple de densité de probabilité des paramètres obtenue après calibration Bayésienne du site autrichien de Stubai	67
Figure 22. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.05 m de profondeur avec réduction de précipitation.....	71
Figure 23. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.10 m de profondeur sans réduction de précipitation.....	72

Figure 24. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la température du sol à 0.30 m de profondeur avec réduction de précipitation.....	73
Figure 25. Analyse d'incertitude des estimations du prior et du posterior de l'indice des surfaces des feuilles sans (à gauche) et avec (à droite) réduction de précipitation	73
Figure 26. Analyse d'incertitude des estimations du prior et du posterior du rendement sans réduction de précipitation (à gauche) et de la surface spécifique foliaire avec réduction de précipitation (à droite)	74
Figure 27. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.30 m de profondeur avec réduction de précipitation.....	75
Figure 28. Analyse d'incertitude des estimations du prior et du posterior de la biomasse avec réduction de précipitation	75
Figure 29. Estimations du prior et du posterior de l'indice des surfaces des feuilles sans (en haut) et avec (en bas) réduction de précipitation	76
Figure 30. Estimations du prior et du posterior de la teneur en eau du sol (à gauche) et de la température du sol (à droite) à 0.15 m de profondeur avec réduction de précipitation.....	77
Figure 31. Analyse d'incertitude des résultats de la calibration sans (à gauche) et avec LAI (à droite) de la teneur en eau du sol à 0.30 m de profondeur avec réduction de précipitation	77
Figure 32. Analyse d'incertitude des résultats de la calibration sans (à gauche) et avec LAI (à droite) de la biomasse avec réduction de précipitation	78
Figure 33. Classification des sites selon deux indices bioclimatiques (l'indice d'aridité et le bilan hydrique climatique) liés à la disponibilité de l'eau.....	85
Figure 34. Localisation géographique des sites.....	86
Figure 35. Niveaux statistiques définis à partir de deux fonctions de distribution cumulée : uniforme (gris) et Gaussienne (noire)	91
Figure 36. Diagramme du plan expérimental pour l'analyse de sensibilité	91
Figure 37. Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire brute (GPP) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres	95
Figure 38. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Pusztá en utilisant une distribution Gaussienne pour les valeurs des paramètres.....	95
Figure 39. Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire brute (GPP) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres	96
Figure 40. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Pusztá en utilisant une distribution uniforme pour les valeurs des paramètres.....	97
Figure 41. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production nette de biome (NBP) pour le site Amplero en utilisant une distribution uniforme pour les valeurs des paramètres	101
Figure 42. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production nette de biome (NBP) pour le site Laqueuille (gestion intensive) en utilisant une distribution Gaussienne pour les valeurs des paramètres	102

Figure 43. Proportionnalité entre les deux indices de sensibilité la moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Pusztta en utilisant une distribution Gaussienne pour les valeurs des paramètres.....	103
Figure 44. Localisation des sites prairiaux utilisés pour la calibration (en bleu) et la validation (en rouge)	104
Figure 45. Comparaison des valeurs mesurées (avec les barres d'écart-type), du prior et du posterior de la biomasse aérienne au site suisse d'Oensingen en utilisant une distribution uniforme.....	108
Figure 46. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes pour le site français de Laqueuille géré intensivement en utilisant une distribution uniforme	109
Figure 47. Régression linéaire des valeurs simulées de la NEE ($\text{g C m}^{-2} \text{d}^{-1}$) (prior [-] et posterior [+]) versus les valeurs observées à Laqueuille (2004-2008) avec une gestion extensive (gauche) et intensive (droite), en utilisant une distribution uniforme	110
Figure 48. Comparaison des valeurs mesurées, du prior et du posterior de la teneur en eau du sol (à gauche) et de la température du sol (à droite) à Laqueuille avec une gestion extensive (2003-2009), en utilisant une distribution uniforme.....	111
Figure 49. Estimations du prior et du posterior de l'indice de surface foliaire au site suisse de Frübüel en utilisant une distribution uniforme.....	111
Figure 50. CH_4 ($\text{g CH}_4 \text{C m}^{-2} \text{d}^{-1}$) entérique simulé (prior et posterior) et observé à Laqueuille avec gestion intensive pour la période entre mai et octobre 2010 en utilisant une distribution uniforme	112
Figure 51. Classification des sites de calibration (en bleu) et des sites de validation (en rouge) selon deux indices bioclimatiques (l'indice d'aridité et le bilan hydrique climatique).....	114
Figure 52. Analyse d'incertitude des estimations du prior et du posterior de la biomasse aérienne au site Allemand de Grillenburg en utilisant une distribution uniforme	115
Figure 53. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la production primaire brute cumulée (GPP) au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme	115
Figure 54. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes) pour le site espagnol de Vall d'Alinyà en utilisant une distribution uniforme	116
Figure 55. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la température du sol au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme	117
Figure 56. Analyse d'incertitude des estimations du prior et du posterior de la biomasse au site hollandais de Cabauw en utilisant une distribution uniforme.....	118
Figure 57. Comparaison des valeurs mesurées, du prior et du posterior de la production primaire brute cumulée (GPP) au site espagnol de Vall d'Alinyà en utilisant une distribution Gaussienne	119
Figure 58. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes) pour le site espagnol de Vall d'Alinyà en utilisant une distribution Gaussienne	120

Table des tableaux

Tableau 1. Récapitulatif comparant différentes méthodes d'analyse de sensibilité	53
Tableau 2. Liste des sites et des variables de sortie utilisés pour la calibration	57
Tableau 3. Description des variables de sortie (D_i , $i=1, \dots, 5$) utilisées pour la calibration de Stubai	59
Tableau 4. Description des paramètres de PaSim (θ_i , $i=1, \dots, 17$) utilisés pour la calibration.....	63
Tableau 5. Nouveau paramétrage de PaSim pour trois sites prairiaux incluant des conditions de sécheresse (Traitement), obtenue après calibration Bayésienne (valeurs correspondant au maximum de la FDP)	68
Tableau 6. Exemple des distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue au site autrichien de Stubai	69
Tableau 7. Exemple des distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue au site suisse de Frübüel.....	70
Tableau 8. Evaluation du prior et du posterior des estimations de quatre variables de sortie obtenues avec et sans (Control) réduction de précipitation au site autrichien de Stubai (RMSE, <i>root mean square error</i> ; CRM, <i>coefficient of residual mass</i>).....	79
Tableau 9. Evaluation du prior et du posterior des estimations de deux variables de sortie obtenues avec et sans (Control) réduction de précipitation aux deux sites suisses (RMSE, <i>root mean square error</i> ; CRM, <i>coefficient of residual mass</i>).....	80
Tableau 10. Informations et climats des sites étudiés	83
Tableau 11. La gestion des sites étudiés	84
Tableau 12. Liste des paramètres considérés pour l'analyse de sensibilité.....	88
Tableau 13. Classement des paramètres les plus influents obtenu avec la méthode Morris (μ) pour un ensemble de variables de sortie dans chaque site en utilisant une distribution Gaussienne des valeurs des paramètres.....	98
Tableau 14. Classement des paramètres les plus influents obtenu avec la méthode Morris (μ) pour un ensemble de variables de sortie dans chaque site en utilisant une distribution uniforme des valeurs des paramètres.....	99
Tableau 15. Coefficient de concordance ($TDCC$) et plasticité du modèle (L) pour chaque variable de sortie en utilisant une distribution Gaussienne et une distribution uniforme	99
Tableau 16. Paramètres utilisés pour la calibration Bayésienne multi-sites	105
Tableau 17. Variables de sortie utilisées dans la calibration Bayésienne multi-sites	106
Tableau 18. Nouveau paramétrage européen de PaSim obtenu après calibration Bayésienne multi-sites en utilisant deux distributions des paramètres (uniforme et Gaussienne)	107
Tableau 19. Exemple de distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue après calibration Bayésienne multi-sites utilisant le nouveau paramétrage (distribution uniforme)....	107
Tableau 20. Informations et climats des sites de validation	113
Tableau 21. Evaluation des estimations a priori et a posteriori de quatre variables de sortie sur l'ensemble des sites de validation avec les deux distributions uniforme et Gaussienne (RMSE, <i>root mean square error</i> ; CRM, <i>coefficient of residual mass</i>)	121

Remerciements

Je voudrais tout d'abord remercier l'ensemble des membres du jury qui m'ont fait l'honneur de bien vouloir y participer.

Merci à Gianni Bellocchi, mon directeur de thèse, qui m'a accompagné durant ces trois années et qui était toujours disponible, il a su trouver la bonne formule pour guider mon travail et surtout me motiver avec ses phrases célèbres comme « il y a une solution pour tout » ou « courage, courage ». Tant sur le plan professionnel que personnel, je suis très reconnaissant à ton soutien Gianni et je me souviendrai longtemps de nos déplacements ensemble et de toute cette période de thèse qui s'est déroulée dans des bonnes conditions et dans la bonne humeur.

Je voudrais ensuite remercier Vincent Barra, mon co-encadrant de thèse, pour son aide et sa contribution surtout sur la partie statistique de mon travail. Il a été disponible et a su me donner parfois un point de vue différent sur mon travail grâce à son expérience. Merci infiniment encore à vous deux.

Je tiens également à remercier Pascal Carrère qui était toujours à l'écoute et toute l'unité UREP pour m'avoir permis de faire cette thèse dans des bonnes conditions. Parmi mes collègues à l'INRA, j'adresse tout d'abord un merci particulier à Romain Lardy avec qui j'ai travaillé les deux premières années de ma thèse, en plus de partager le même bureau, Romain m'a fait part de son expérience du modèle et de ses compétences et m'a beaucoup aidé surtout au début de la thèse, sans oublier son talent de pâtissier bien sûr et nos discussions ouvertes. Je voudrais ensuite remercier Franck Ganzalvo pour l'ambiance qu'il a apportée au bureau et pour ses compétences en matière de préparation de café. Je remercie Shaoxiou Ma pour les parties de ping pong et Alexandre, Lionel, Raphaël et Vincent pour les parties de foot ce qui m'a permis de changer les idées de temps en temps. J'aimerais aussi remercier mes collègues doctorants et leurs souhaiter une bonne chance: Nazia, Nian-Xun, Damien, Benoît et Pauline. Je voudrais également remercier nos deux gestionnaires Annick et Laurence et tous ceux avec qui j'ai eu le plaisir de travailler ou simplement de discuter : Katja, Sébastien, Gaël, Olivier, Julien, Giorgio, Frédérique, Catherine, Sylvie pour ne citer qu'eux. Je remercie aussi tous mes collègues du LIMOS et de l'Université Blaise Pascal pour leurs soutiens.

Pour les deux êtres les plus chers sur cette terre ma mère et mon père, les mots ne seront pas suffisants pour exprimer ma reconnaissance et ma gratitude. Si je suis là aujourd'hui c'est grâce à vous, et tout ce que vous avez fait pour moi n'a pas de prix et ne peut être récompensé. Tout simplement, je vous dédie cette thèse et tous mes succès futurs.

Je remercie toute ma famille notamment mes frères : Sami, Saïef et Hassène pour leurs conseils et leurs soutiens, ma belle nièce Nour pour le bonheur qu'elle m'apporte et je n'oublie pas bien sûr mes amis présents au quotidien qui m'ont aidé sur tous les plans et qui étaient des frères pour moi. J'ai tellement d'amis qu'il faut rédiger une thèse à part pour les citer un par un, merci infiniment à vous tous mes chers amis.

Enfin, un mot pour tous ceux que j'ai pu oublier de citer mais qui ont compté tout au long de mon parcours, merci à tous.

Résumé

Les prairies représentent 45% de la surface agricole en France et 40% en Europe, ce qui montre qu'il s'agit d'un secteur important particulièrement dans un contexte de changement climatique où les prairies contribuent d'un côté aux émissions de gaz à effet de serre et en sont impactées de l'autre côté. L'enjeu de cette thèse a été de contribuer à l'évaluation des incertitudes dans les sorties de modèles de simulation de prairies (et utilisés dans les études d'impact aux changements climatiques) dépendant du paramétrage du modèle. Nous avons fait appel aux méthodes de la statistique Bayésienne, basées sur le théorème de Bayes, afin de calibrer les paramètres d'un modèle référent et améliorer ainsi ses résultats en réduisant l'incertitude liée à ses paramètres et, par conséquent, à ses sorties.

Notre démarche s'est basée essentiellement sur l'utilisation du modèle d'écosystème prairial PaSim, déjà utilisé dans plusieurs projets européens pour simuler l'impact des changements climatiques sur les prairies. L'originalité de notre travail de thèse a été d'adapter la méthode Bayésienne à un modèle d'écosystème complexe comme PaSim (appliqué dans un contexte de climat altéré et à l'échelle du territoire européen) et de montrer ses avantages potentiels dans la réduction d'incertitudes et l'amélioration des résultats, en combinant notamment méthodes statistiques (technique Bayésienne et analyse de sensibilité avec la méthode de Morris) et outils informatiques (couplage code R-PaSim et utilisation d'un cluster de calcul). Cela nous a conduit à produire d'abord un nouveau paramétrage pour des sites prairiaux soumis à des conditions de sécheresse, et ensuite à un paramétrage commun pour les prairies européennes. Nous avons également fourni un outil informatique de calibration générique pouvant être réutilisé avec d'autres modèles et sur d'autres sites. Enfin, nous avons évalué la performance du modèle calibré par le biais de la technique Bayésienne sur des sites de validation, et dont les résultats ont confirmé l'efficacité de cette technique pour la réduction d'incertitude et l'amélioration de la fiabilité des sorties.

Mots-clés : Paramétrage de modèle, écosystème prairial, changement climatique, calibration Bayésienne, analyse de sensibilité, réduction d'incertitude.

Abstract

Grasslands cover 45% of the agricultural area in France and 40% in Europe. Grassland ecosystems have a central role in the climate change context, not only because they are impacted by climate changes but also because grasslands contribute to greenhouse gas emissions. The aim of this thesis was to contribute to the assessment of uncertainties in the outputs of grassland simulation models, which are used in impact studies, with focus on model parameterization. In particular, we used the Bayesian statistical method, based on Bayes' theorem, to calibrate the parameters of a reference model, and thus improve performance by reducing the uncertainty in the parameters and, consequently, in the outputs provided by models.

Our approach is essentially based on the use of the grassland ecosystem model PaSim (*Pasture Simulation model*) already applied in a variety of international projects to simulate the impact of climate changes on grassland systems. The originality of this thesis was to adapt the Bayesian method to a complex ecosystem model such as PaSim (applied in the context of altered climate and across the European territory) and show its potential benefits in reducing uncertainty and improving the quality of model outputs. This was obtained by combining statistical methods (Bayesian techniques and sensitivity analysis with the method of Morris) and computing tools (R code -PaSim coupling and use of cluster computing resources). We have first produced a new parameterization for grassland sites under drought conditions, and then a common parameterization for European grasslands. We have also provided a generic software tool for calibration for reuse with other models and sites. Finally, we have evaluated the performance of the calibrated model through the Bayesian technique against data from validation sites. The results have confirmed the efficiency of this technique for reducing uncertainty and improving the reliability of simulation outputs.

Keywords: Model parameterization, Grassland ecosystem, climate change, Bayesian calibration, sensitivity analysis, reducing uncertainty.

Introduction générale

Les prairies sont des écosystèmes très complexes et dynamiques. Les nombreuses interactions entre les herbivores, la végétation, le sol et l'atmosphère font que notre capacité à expérimenter ces systèmes est limitée. Ainsi, tester l'impact des scénarios de changement climatique en utilisant des modèles d'écosystèmes simulant les processus physiques, chimiques et biologiques dans les moindres détails est un impératif. De telles projections doivent tenir compte de l'incertitude qui en résulte en cascade à partir de sources multiples, y compris les scénarios climatiques, les modèles d'impact, et les conditions locales de climat, végétation et sol.

Dans le cadre des activités de notre Unité de Recherche sur les Ecosystèmes Prairiaux (UREP) de l'INRA, un intérêt particulier est porté à la modélisation des écosystèmes prairiaux ainsi que sur la problématique de l'augmentation des émissions de gaz à effet de serre qui accompagne ses écosystèmes.

Un problème majeur avec l'utilisation de modèles écologiques complexes est la connaissance incomplète des paramètres en entrée du modèle. Les estimations d'un modèle étant fortement dépendantes des valeurs attribuées aux paramètres, cela conduit à une incertitude importante due aux incertitudes liées aux valeurs des paramètres, qui s'ajoutent à celles liées aux variables d'entrée (climat, sol et gestion) et à la structure même du modèle. Il est donc nécessaire de calibrer les paramètres pour obtenir une bonne cohérence entre les sorties des modèles et les observations. Tout d'abord, les incertitudes sont quantifiées en les exprimant en tant que fonctions de distribution de probabilité. Ensuite, des échantillons représentatifs sont prélevés à partir de ces distributions pour mettre en avant l'incertitude des paramètres à travers les calculs du modèle. Pour cela, nous appliquons une méthode basée sur le modèle probabiliste Bayésien qui permet d'intégrer autant d'informations que possible dans les distributions des paramètres pour réduire leurs incertitudes et, par conséquent, améliorer les sorties du modèle. Certains paramètres sont plus pertinents que d'autres dans un modèle, ayant un impact plus important sur les variables de sortie. L'analyse de sensibilité est une solution qui permet d'identifier les paramètres les plus influents sur les sorties afin de les calibrer.

Le premier chapitre de ce manuscrit aborde les concepts des écosystèmes prairiaux, leurs rôles, leurs réponses aux changements climatiques ainsi que leurs contributions aux émissions de gaz à effet de serre.

Le deuxième chapitre présente un état de l'art sur la calibration et l'analyse de sensibilité en se penchant sur les différentes méthodes et outils permettant de répondre à la problématique de la thèse, notamment le besoin en puissance de calcul pour les simulations de modèles complexes.

Dans le troisième chapitre, nous présentons notre proposition de développement d'une méthode de calibration des paramètres du modèle de simulation d'écosystème prairial PaSim (*Pasture Simulation model*) basée sur une approche Bayésienne et appliquée sur des sites d'expérimentations à court et moyen termes incluant des manipulations de conditions de sécheresse.

Dans le quatrième chapitre, nous menons une démarche plus globale couvrant divers sites prairiaux européens, où une calibration Bayésienne multi-sites à l'échelle de l'Europe a été mise en œuvre dans le but d'obtenir un nouveau paramétrage de PaSim valable pour les prairies européennes. Ce chapitre commence par une analyse de sensibilité multi-sites pour déterminer les paramètres de la végétation les plus influents qui sont soumis ensuite à une procédure de calibration Bayésienne simultanée à plusieurs sites. Les sorties du modèle, renseigné avec le nouveau paramétrage obtenu à l'issue de cette calibration, seront également évaluées sur des sites non inclus dans la calibration.

Enfin, le bilan des résultats est présenté ainsi que des perspectives de recherches, à la fois complémentaires et liées à cette thèse, sont proposées dans la conclusion générale.

Chapitre 1 : Les écosystèmes prairiaux et les changements climatiques

1.1. Introduction

Dans ce chapitre nous allons présenter le contexte général de la thèse et définir notre champ d'application, c'est-à-dire les écosystèmes prairiaux, leurs rôles et leurs liens avec les émissions des gaz à effet de serre et les changements climatiques. Ensuite, nous détaillerons le besoin de la modélisation pour simuler les écosystèmes et les problèmes associés, et nous présenterons le modèle de prairie utilisé (PaSim, *Pasture Simulation model*). Enfin, nous posons la problématique de notre étude ainsi que les objectifs et les questions scientifiques qui en découlent.

1.2. Les écosystèmes prairiaux

Un écosystème est défini en écologie comme un ensemble formé d'êtres vivants en interaction avec l'environnement physique. Les écosystèmes liés aux activités agricoles sont appelés agro-écosystèmes. Les écosystèmes prairiaux, notre champ d'étude, sont composés d'un couvert herbacé, un sol et éventuellement des herbivores.

Les prairies (ou herbages ou cultures fourragères) sont des cultures de plantes herbacées, principalement composées de graminées et de légumineuses, destinées à être pâturées ou fauchées. On distingue deux types principaux de prairies : d'un côté les prairies permanentes et de l'autre les prairies temporaires et artificielles.

Les prairies permanentes sont des prairies naturelles ou des prairies semées depuis au moins cinq ans (d'après la définition donnée par la Commission européenne, Régulation EU No 796/2004) et composées de plusieurs espèces. Par contre, les prairies temporaires et artificielles sont des prairies semées et périodiquement retournées, entrant souvent dans une rotation de cultures. Ils existent aussi différents modes de gestion de la prairie. On trouve les fauches et les pâturages, la fertilisation et parfois l'irrigation. En complément au pâturage, l'herbe fauchée sert à nourrir les animaux à d'autres périodes de l'année. La fertilisation sert à compenser les pertes de la prairie en azote (N), phosphore (P), potassium (K) et soufre (S). Selon la quantité de la fertilisation et l'intensité des fauches ou des pâturages, on peut distinguer plusieurs niveaux d'intensification des prairies. Une

classification des modes de gestion est proposé par [Agreste 2000] suivant les pratiques : intensif, semi-intensif, modéré et extensif.

1.2.1. Diffusion et rôle des prairies en Europe et en France

En Europe (UE-27), les prairies prennent une place majeure dans l'occupation du territoire [Peyraud 2013]¹, avec 67 millions d'hectare (soit 40% de la Surface Agricole Utile [SAU]), gérées par 5,4 millions d'agriculteurs (soit 40% du total de l'UE-27). Pour la plupart (85%) des prairies permanentes, elles sont à la base de l'alimentation de 78 millions d'UGB² herbivores dont le lait et la viande représentent 25% environ de la production agricole communautaire. Les prairies couvrent également une surface importante du territoire et du secteur agricole en France (11 millions d'hectares en 2008, soit 24% de la superficie du territoire et 45% de la SAU) [Agreste 2008]. En outre, les prairies ont un rôle essentiel dans la production de nombreux biens et services environnementaux, la fourniture de ces services étant de façon générale plus importante pour les prairies permanentes que les prairies temporaires. Parmi les atouts environnementaux de la prairie, on peut noter : la séquestration du C (compensant au moins partiellement les émissions de CH₄ entérique), la protection du sol contre l'érosion et l'amélioration de sa fertilité, le support d'une diversité spécifique considérable et la fourniture d'habitats pour les invertébrés et la faune sauvage, la participation à la recharge en eau des nappes phréatiques et la production d'une eau de qualité. Enfin, la prairie est le mode privilégié d'occupation des surfaces agricoles dans des zones aux conditions climatiques et topographiques particulièrement défavorables, où la production de ruminants y constitue un secteur économique à part entière, garantissant l'entretien de l'espace rural et le maintien des paysages ouverts et diversifiés. Pour la France, les bénéfices rendus par les prairies permanentes extensives ont été estimés à environ 600 € par hectare et par an [Chevassus-au-Louis et al., 2009].

Bien que moins présentes dans la SAU, les prairies temporaires fournissent également des biens et services écologiques. Elles constituent les têtes de rotation pour de nombreuses

¹ Synthèse de données statistiques de l'analyse des bases FAOSTAT (<http://faostat.fao.org>) et Eurostat réalisées dans le cadre du projet UE-FP7 MultiSward (<http://www.multisward.eu>), ainsi que par [Agreste 2012].

² Unité de Gros Bétail, c'est-à-dire l'équivalent pâturage d'une vache laitière produisant 3 000 kg de lait par an, sans complément alimentaire concentré. Cette unité de référence permet d'agréger des effectifs de bétail à partir de coefficients attribués selon la catégorie des animaux (http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Livestock_unit_%28LSU%29/fr).

cultures, permettant ainsi de réduire l'utilisation des intrants sur celles-ci et participent (notamment celles à base de légumineuses) à la régulation des flux d'N et des adventices. C'est à ce titre, par exemple, que l'augmentation dans la SAU de la part de prairies temporaires constitue, en France, un levier dans le cadre du plan Ecophyto 2018 (<http://agriculture.gouv.fr/ecophyto>) issu des travaux du Grenelle Environnement (<http://www.developpement-durable.gouv.fr/-Le-Grenelle-de-l-environnement-de-.html>) menés en 2008 et repris par le second Plan national santé environnement (<http://www.sante.gouv.fr/plan-national-sante-environnement-pnse,3480.html>) en 2009.

La Figure 1 montre que les herbages utilisés à des fins agricoles en Europe sont principalement concentrés dans les régions où les sols sont moins fertiles et où les forêts ont disparu au cours des siècles au profit de la croissance économique ou pour des raisons climatiques. En France, les régions Auvergne et Limousin, productrices de fromage et de viande, sont très largement dominées par les prairies. Les zones ayant une part relativement faible d'herbages à usage agricole sont soumises à des conditions climatiques plus dures, aussi bien dans le nord que dans le sud de l'Europe, soit où les températures sont moins élevées (Scandinavie) ou des conditions arides prévalent (Méditerranée).

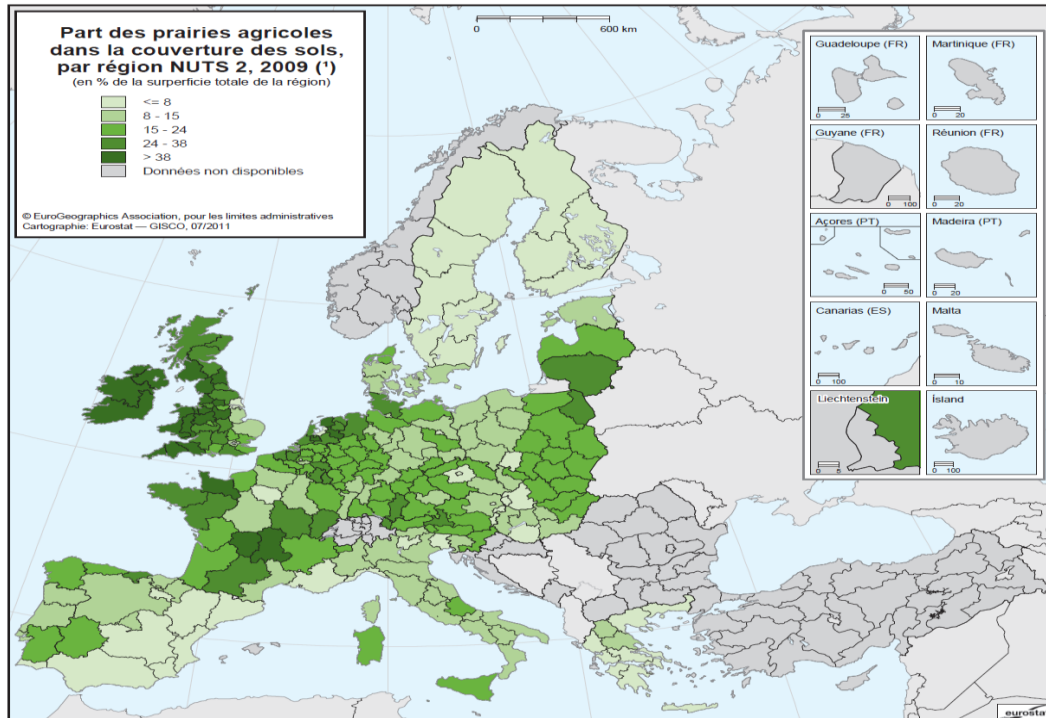


Figure 1. Part des prairies agricoles dans la couverture des sols en Europe (UE-23, enquête LUCAS du programme de travail d'Eurostat, initiée en 2009, <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>)

1.3. Réponses des écosystèmes prairiaux au changement climatique

1.3.1. Le changement climatique

Le changement climatique est une modification durable du climat. Dans les années 1970, les interrogations sur l'influence possible des activités humaines sur le climat de la planète ont fait surface, en lien avec les observations sur l'augmentation de la concentration en dioxyde de carbone (CO₂) dont le niveau actuel avoisine les 400 ppm (<http://co2now.org>), contre 260 ppm à l'époque préindustrielle. Depuis 1750, la concentration atmosphérique de CO₂ a augmenté d'un tiers. D'autres gaz, présents naturellement à l'état de traces, comme le méthane (CH₄) et le protoxyde d'azote (N₂O), s'accumulent également dans l'atmosphère et contribuent au renforcement de l'effet de serre naturel. Ces gaz sont appelés gaz à effet de serre (GES). Parmi les principaux GES, on trouve une contribution à hauteur de 76,7% pour le dioxyde de carbone (CO₂), de 14,3% pour le méthane (CH₄) et de 7,9% pour le protoxyde d'azote (N₂O) [IPCC, 2007]. L'accroissement moyen de la température de surface est estimé devoir être de 1,5 à 6 °C entre 1990 et 2100. Il est presque certain que toutes les surfaces continentales se réchaufferont plus rapidement que la moyenne, particulièrement celles situées à haute latitude, en saison froide. Les prédictions sur la pluviométrie sont plus incertaines, compte tenu de la complexité du cycle de l'eau, mais elles font état en général d'une légère augmentation de la pluviométrie, avec une tendance à la diminution de la pluviométrie estivale dans les zones tempérées de moyenne latitude. Par ailleurs, en dehors de ces variations du climat moyen, il est vraisemblable que le changement climatique s'accompagne d'un accroissement de la variabilité et d'une intensification des événements extrêmes.

Concernant les évolutions des niveaux et de la répartition, spatiale et temporelle, des précipitations, les modèles sont moins assurés que pour l'évolution des températures. Les scénarios disponibles prévoient plutôt une accentuation des différences régionales (Sud encore plus sec) et saisonnières (pluies plus abondantes en période hivernale et moindres en été), ainsi que des risques accrus d'événements extrêmes.

Par définition, les sécheresses sont des événements peu fréquents, pour lesquels les tendances sont plus difficiles à établir que pour des phénomènes continus comme l'augmentation des températures. Cependant, d'après Météo-France, il apparaît une légère

tendance en termes d'événements extrêmes (nombre de jours consécutifs sans pluie) dans le sud de la France, confortée par une diminution de la pluviométrie estivale (de 10% sur les 20 dernières années). L'hypothèse qui apparaît maintenant la plus probable, si l'on considère l'horizon 2010-2020, est celle d'une persistance des épisodes secs plutôt que l'inverse.

1.3.2. Contribution de l'agriculture aux émissions de gaz à effet de serre

Le secteur agricole est à la fois une source et un puits de GES. Les échanges de GES entre l'atmosphère et les agroécosystèmes concernent les flux de CO₂, N₂O et CH₄. Avec 5.1-6 Gt CO₂-eq. par an d'émissions de GES estimés en 2005 ce secteur contribue à hauteur de 14% environ aux émissions mondiales de GES [Smith et al., 2007].

Les émissions directes d'oxyde nitreux (N₂O) par les sols agricoles et de méthane (CH₄) par la fermentation entérique des ruminants sont les deux autres principales sources de GES, à hauteurs respectives de 2 128 et 1 792 Mt CO₂-eq. par an [Smith et al., 2007]. Les émissions de GES associées aux cycles de vie des intrants peuvent représentées jusqu'à 50% du bilan final et ne doivent donc pas être négligées dans l'évaluation des bilans GES des produits agricoles. De plus, les estimations globales des flux de GES du secteur agricole sont très incertaines car les connaissances scientifiques actuelles ne permettent pas de quantifier de manière fiable les flux de GES à large échelle [Smith et al., 2007]. Les émissions de CH₄ et de N₂O de la part du secteur agricole ont augmenté globalement de 18% entre 1990 et 2010. Les projections à 2030 prévoient une augmentation de 35 à 60% des émissions de N₂O principalement à cause de l'augmentation de l'utilisation des engrais azotés minéraux et organiques d'origine animale et une augmentation de 60% des émissions de CH₄ directement proportionnelles à la croissance du troupeau mondial.

Les GES sont considérés comme les principaux responsables des changements climatiques et de l'augmentation de la température annuelle de notre planète (Figure 2).

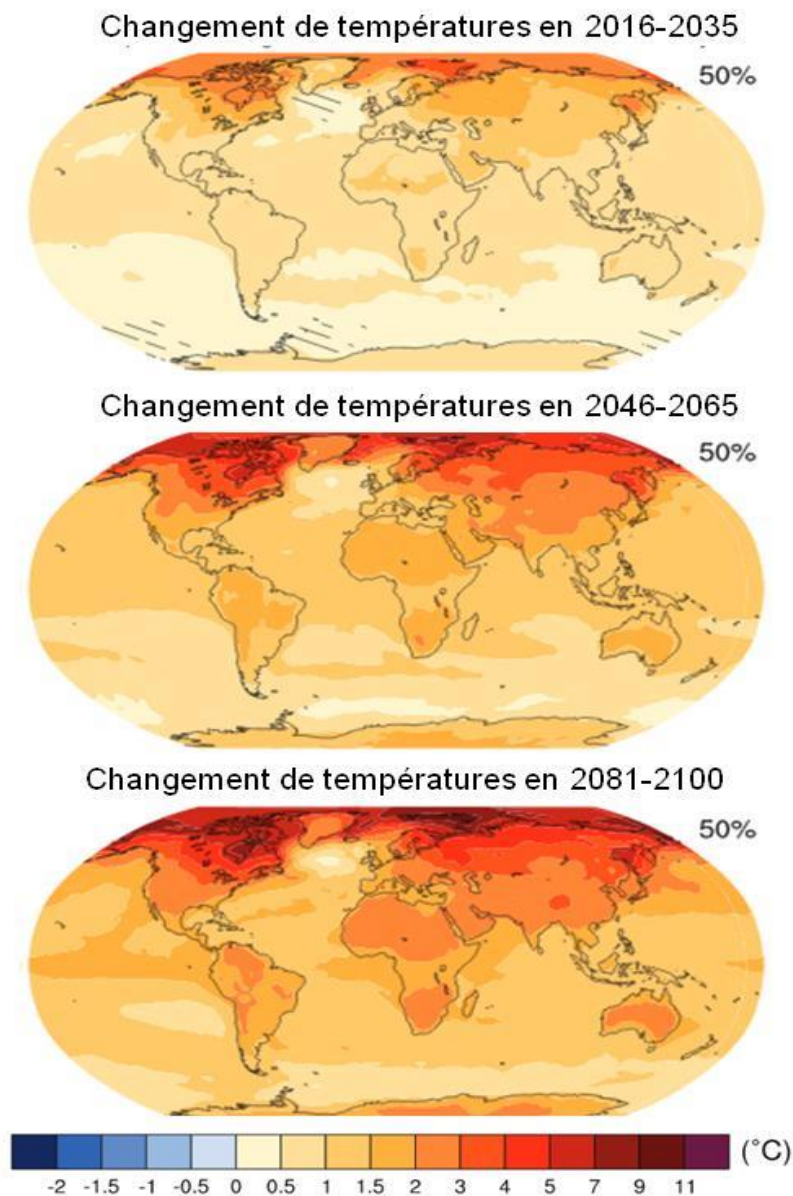


Figure 2. Changements de la température annuelle entre la période (2016-2035), la période (2046-2065) et la période (2081-2100) (source : IPCC, 2013)

1.3.3. Conséquences possibles du changement climatique pour l'agriculture

Les effets du réchauffement sont déjà observables en agriculture à travers, par exemple, l'avancement des dates de floraison des arbres fruitiers ou des dates de moisson et de fauchage. [Seguin, 2007]. L'augmentation des températures a des effets directs sur la demande en eau (augmentation de l'évaporation par accroissement du déficit de saturation de l'air), et des effets indirects par des impacts sur le cycle de développement des cultures (décalage, allongement ou raccourcissement de la période de végétation des espèces...). Le climat actuel, significativement différent de celui des années 1940-1970, est donc très

vraisemblablement en cours d'évolution sous l'action de l'augmentation de la concentration des GES dans l'atmosphère [Seguin, 2007].

Le changement climatique va modifier les données d'entrée en créant des conditions, souvent plus favorables, parfois plus défavorables, suivant les productions. L'augmentation du CO₂ va potentiellement stimuler la photosynthèse des couverts végétaux et améliorer l'efficacité de l'eau par l'augmentation de la résistance, en créant des conditions permettant d'augmenter la production de biomasse de l'ordre de 20%. Les scénarios "fin de siècle" font cependant intervenir des effets de compensation, avec le raccourcissement du cycle de végétation [Amigues et al., 2006].

À priori, les grandes cultures et les prairies européennes devraient être plutôt favorisées, sauf dans le sud où apparaît le risque de sécheresses accentuées, accompagnées de températures élevées. Les conséquences de ces changements sur les écosystèmes prairiaux en France sont encore incertaines, mais elles peuvent être approchées à partir des effets d'un réchauffement sur la prairie, des effets de l'enrichissement en CO₂, des interactions de l'enrichissement en CO₂ avec les différents facteurs climatiques et édaphiques tels que la température, la photopériode, le régime des vents et la disponibilité en eau et nutriments et les agents responsables de maladies, tout en considérant aussi l'incidence d'événements extrêmes tels que les vagues de chaleur, la sécheresse, et les pluies et vents violents).

1.3.4. La prairie et le changement climatique

À travers la production, les systèmes prairiaux modifient les émissions de GES émis par la prairie (CO₂, N₂O, CH₄). A travers les cycles du C et de l'N, on peut comprendre comment la prairie peut impacter l'environnement.

Le cycle du C commence par l'absorption du CO₂ atmosphérique par la plante par photosynthèse. Le CO₂ est alloué aux différents organes de la plante, que ce soit dans la partie souterraine (racines) ou aérienne (tiges, gaines, épis, limbes, ...). Une partie du C est réémise par respiration, tandis qu'une autre va au sol sous forme de litière. Le C de la litière rentre alors dans la matière organique du sol et il est stabilisé ou réémis dans l'atmosphère par respiration en fonction de l'activité des microbes. La fertilisation organique se traduit par un apport de C au niveau de la matière organique du sol. Quand il y a des fauches, le carbone représente 30 à 40% de la biomasse fauchée. Lors du pâturage,

ce C contenu dans l'herbe et consommé par les animaux sera réémis dans l'atmosphère par la respiration et surtout sous forme de CH₄ entérique.

L'article 3.4 du protocole de Kyoto reconnaît le rôle joué par le stockage de C par les prairies dans la lutte contre les émissions de GES. Une récente synthèse de l'INRA montre que, sur les 30 premiers centimètres de sol, la quantité de C est de 70 tonnes par hectare pour les prairies [Institut de l'Élevage 2007]. Cela s'explique par un flux de C important dans les systèmes prairiaux (davantage de racines et débris), une décomposition lente de la matière organique en l'absence d'aération du sol, et enfin, une dégradation faible des racines prairiales.

Le cycle de l'N dans la prairie se fait principalement par la fertilisation minérale (sous forme d'ammonium ou de nitrate) ou organique (fumier, lisier,...), la déposition atmosphérique d'ammoniac (vers la plante ou le sol), ou bien par fixation symbiotique des légumineuses à partir du diazote (N₂) atmosphérique. L'N sous forme de nitrate (NO₃⁻) et d'ammonium (NH₄⁺) est assimilé par les plantes au niveau des racines dans le sol. De même que pour le C, une partie de l'N de la plante peut être perdue par sénescence et entrer dans la matière organique du sol. En fonction de la dynamique de la matière organique du sol, de l'N peut être libéré sous forme d'ammonium ou de nitrate dans le sol (minéralisation), ou inversement être immobilisé dans la matière organique. Au sein de l'N inorganique (minéral) du sol, des réactions de nitrification³ et de dénitrification⁴ ont lieu, entraînant la formation de protoxyde d'azote (N₂O) et de diazote (N₂). Lors de la fauche une partie de l'N est exportée au sein de la biomasse. De même lors du pâturage, une fraction de l'N est assimilée par l'animal ou exportée dans le lait, tandis que le reste est rendu au système par les déjections [Lardy, 2013].

Le rôle des prairies est crucial dans le système d'élevage, la France est le premier éleveur bovin en Europe avec 19 millions de têtes en 2010 soit 20% du troupeau européen [Agreste 2010], d'où l'importance de l'étude de l'impact du changement climatique sur les prairies. Le changement climatique se manifeste par l'augmentation de la température, par les changements des précipitations (éventuellement des événements extrêmes tels que les sécheresses) et l'augmentation de la concentration en CO₂ atmosphérique. Parmi les impacts du changement climatique sur les prairies, on trouve une augmentation de la

³ Dénitrification : réaction permettant de passer des nitrates au diazote (N₂).

⁴ Nitrification : réaction permettant de passer de l'ammoniac au nitrate en passant par l'ammonium.

production annuelle et des baisses de la qualité du fourrage [Graux et al., 2012a]. Ces impacts sont de plusieurs ordres, souvent progressifs mais parfois brutaux, et touchent de manière très inégale les différentes régions du monde. A long terme, une diminution globale de production risque d'être provoquée par i) la forte augmentation de température et ii) les effets négatifs de la diminution des précipitations et de l'augmentation de l'évapotranspiration [Ruget et al., 2013].

Tout cela montre que les prairies, en même temps, impactent et sont impactées par le changement climatique et jouent un rôle important que ce soit dans la contribution aux émissions des GES ou dans l'aide à la réduction de leurs émissions. En outre, l'importance de différents enjeux environnementaux tournants autour des prairies : la qualité de l'eau, la protection contre l'érosion et les inondations, le maintien de la biodiversité et de la qualité des paysages et des produits.

Les prairies sont des écosystèmes dynamiques et complexes comptants différentes interactions avec leur environnement. Le fait d'expérimenter ces systèmes en nature et en conditions réelles peut s'avérer coûteux en termes de temps et de moyens, ce qui rend notre capacité à les étudier extrêmement limitée. Ainsi, il est pratiquement indispensable d'avoir recours à la modélisation pour tester différents scénarios climatiques à travers des modèles d'écosystèmes qui simulent les processus physiques, chimiques et biologiques.

1.4. La modélisation des écosystèmes

Pour répondre aux questions de recherche concernant l'étude de l'impact du changement climatique, la modélisation apparaît comme un moyen indispensable, du fait que la simulation peut nous fournir des informations que l'expérimentation seule n'aurait pas pu offrir en raison des contraintes de temps et de ressources nécessaires. En effet, la simulation donne la possibilité de modifier les variables et le paramétrage du modèle et ainsi de contrôler l'environnement de simulation.

De plus, parallèlement au progrès technologique d'un point de vue logiciel et matériel, le besoin de la modélisation ne cesse d'augmenter dans différents domaines et en particulier dans le domaine de l'agronomie et de l'environnement.

Un modèle peut être défini comme une représentation de la réalité qui simplifie le système étudié, construite sur la base d'un ensemble d'hypothèses relatives à un

phénomène observable et mesurable. Le but étant de reproduire le comportement du système réel étudié, en fonction de la problématique posée et des aspects qui intéressent le modélisateur [Coquillard et Hill., 1997].

Une autre définition, fournie par [Parent et Bernier., 2007], présente un modèle comme une construction mentale qui a pour but la traduction opérationnelle d'un ensemble de connaissances à des fins de déduction.

Un modèle fait généralement intervenir trois éléments clés. Tout d'abord, il explicite des hypothèses fondatrices qui organisent une représentation simplifiée du phénomène adaptée aux objectifs. On y trouve ensuite une formulation quantifiée des relations possibles entre grandeurs d'intérêts, la valeur attribuée à chacun des paramètres doit nécessairement se trouver à l'intérieur de la gamme permise par les hypothèses sur lesquelles le modèle a été fondé. Enfin, tout modèle statistique comprend une partie probabiliste qui sert à représenter les influences qui échappent au savoir actuel.

1.4.1. Modèles agro-écologiques

L'écologie étudie les conditions d'existence d'un être vivant et les rapports entre ce dernier et son environnement, la notion d'écosystème qu'en découle reflète l'ensemble constitué par un milieu (sol, eau, etc.) et des êtres vivants (plantes, animaux, etc.), ainsi que les relations énergétiques et trophiques entre eux (qui se rapportent à la nutrition des tissus) [Gotelli, 1998]. Les écosystèmes prairiaux, qui seront notre milieu d'application, sont formés de plusieurs composantes, en interaction entre elles et avec leur environnement.

En fait, le comportement d'un système n'est pas équivalent à la somme des comportements des parties. Seule une modélisation mettant en interaction les différentes parties du système peut faire apparaître les comportements émergents. Il n'est donc pas surprenant que les modèles en matière d'écologie soient de plus en plus utilisés pour une meilleure compréhension des écosystèmes [Hill, 2009]. [Jorgensen, 1994] résume en quatre points les avantages de la modélisation :

- Les modèles ont leur utilité dans la surveillance de systèmes complexes.
- Les modèles peuvent être utilisés pour révéler les propriétés des systèmes écologiques.

- Les modèles peuvent montrer des carences dans nos connaissances et être utilisés pour définir des priorités dans la recherche.
- Les modèles sont utiles pour tester des hypothèses scientifiques, dans la mesure où le modèle peut simuler les réactions de l'écosystème, qui peuvent être comparées aux observations.

Face à la complexité du fonctionnement des écosystèmes, seules les techniques de modélisation liées aux performances croissantes de l'outil informatique permettent d'utiliser des aides fiables à la gestion raisonnées de notre environnement et nous fournissent des éléments d'aide à la prise de décision.

Il existe différents types de modèles, les modèles peuvent être déterministes ou stochastiques, dynamiques ou statiques, empiriques ou mécanistes.

Un modèle déterministe est un modèle où l'activité peut être déterminée entièrement par ses entrées sans faire intervenir aucune variable aléatoire et donc dont les prédictions ne reposent pas sur des distributions de probabilité, contrairement au modèle stochastique dont les prédictions sont basées sur l'aléatoire [Thornley, 2001].

Un modèle dynamique utilise des équations différentielles en fonctions du temps, alors qu'un modèle analytique est indépendant du temps et ne l'intègre pas comme une variable.

Un modèle empirique est construit sur la base d'équations prédictives statistiques (dont les modèles de régression) ou mathématiques sans contenir des données scientifiques ou explicatives des processus modélisés, tandis qu'un modèle mécaniste est basé sur la compréhension du fonctionnement du système étudié et l'explication des mécanismes modélisés. Ceci à travers des équations et des interactions entre les processus modélisés à un niveau d'organisation donné et les phénomènes observés à un niveau d'organisation supérieur [Thornley, 2001].

Cela n'empêche qu'un seul modèle peut regrouper différents aspects des types qu'on vient d'aborder ci-dessus, comme par exemple le modèle PaSim que nous allons présenter dans la section suivante et qui est un modèle mécaniste, déterministe et dynamique à la fois.

1.5. Description de PaSim

PaSim (Pasture Simulation Model) initialement développé par [Riedo et al., 1998] (<https://www1.clermont.inra.fr/urep/modeles/pasim.htm>) est un modèle biogéochimique de simulation des prairies à l'échelle de la parcelle. PaSim est inspiré du *Hurley Pasture Model* de [Thornley et al., 1998] pour sa simulation du couvert en réponse à la lumière, à la température et aux ressources C, N et eau. Ce modèle simule les flux de matière et d'énergie au sein du système (sol, végétation, animal) en réponse aux conditions climatiques et aux pratiques de gestion, à un pas de temps proche de la demi-heure (1/50^{ème} du jour) sur plusieurs années. Il prend en compte la gestion pratiquée de fertilisation minérale ou organique, d'irrigation, de retournement du sol et d'exploitation de l'herbe par la fauche et le pâturage. PaSim prend plusieurs entrées liées au climat, sol, végétation et herbivores, et génère plusieurs sorties liées aux flux des GES, à la production végétale et animale (Figure 3). PaSim est structuré en six modules (sous-modèles) interdépendants, simulant le fonctionnement du couvert végétal, des herbivores, de la gestion, du microclimat, de la biologie ou de la physique du sol (Figure 4).

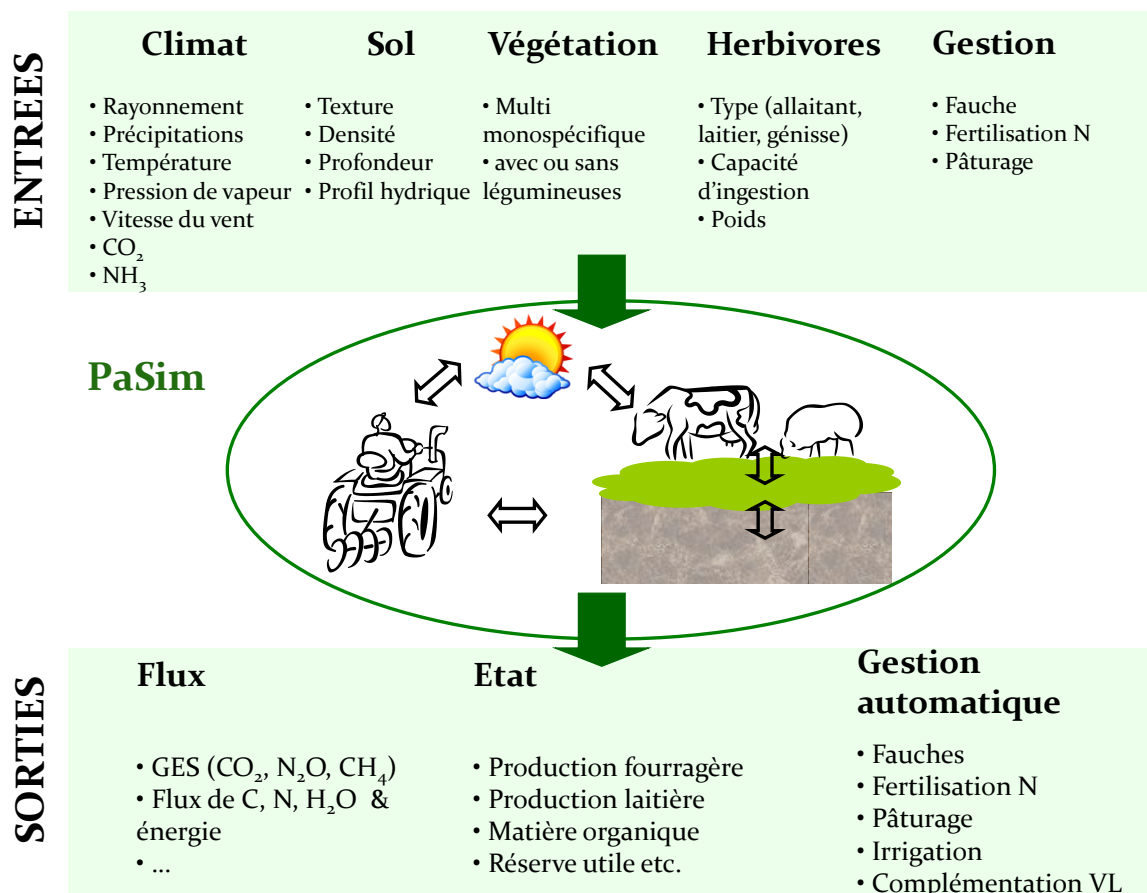


Figure 3. Illustration du modèle PaSim [Graux, 2011]

Actuellement PaSim est traduit en Fortran 90 après une première version codée en ACSL (*Advanced Continuous Simulation Language*), et intègre une nouvelle interface en JAVA récemment créée par l'équipe informatique de l'INRA-UREP simplifiant le renseignement du paramétrage d'entrée du modèle pour des simples utilisateurs. En plus, un guide utilisateur détaillé est disponible en ligne (https://www1.clermont.inra.fr/urep/modeles/PasimUserGuide-pasim_v5-3_1septembre_2011_online_version.pdf) expliquant comment faire tourner le modèle. Pour le développement de PaSim, l'INRA-UREP collabore depuis plusieurs années avec le LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes ; UMR 6158 CNRS, <http://limos.isima.fr>) de l'Université Blaise Pascal de Clermont-Ferrand. Récemment, un brevet de la version 5.3 du modèle a été déposé à l'Agence pour la Protection des Programmes (APP) sous le numéro IDDN.FR.001.220024.000.R.P.2012.000.10000.

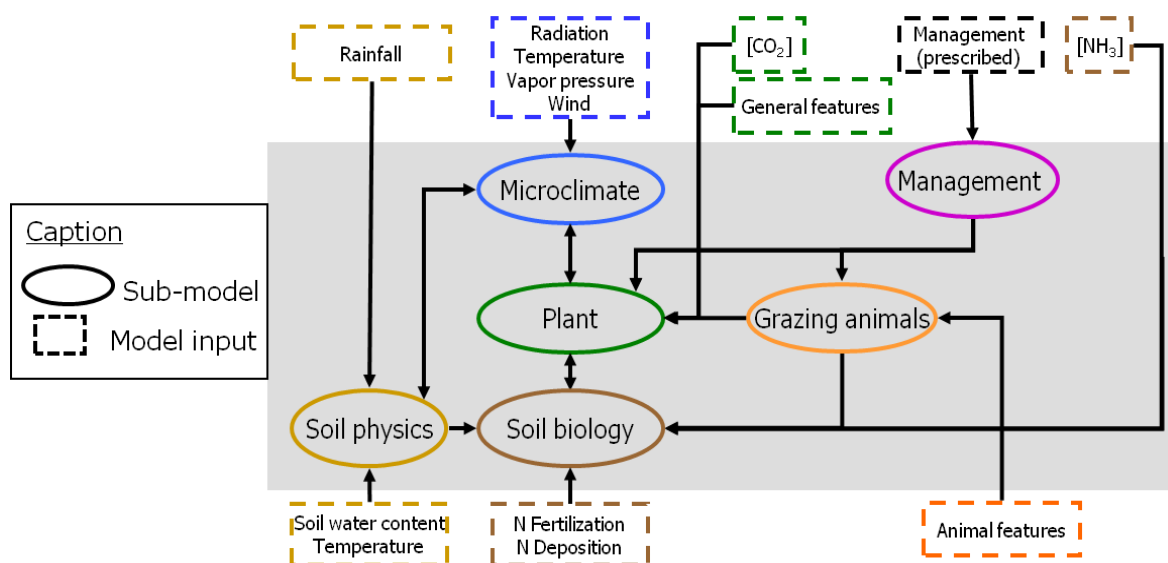


Figure 4. Structure du modèle PaSim (entrées + modules)

PaSim, qui avait déjà été développé et évalué au sein de notre équipe [Vuichard et al., 2007 a, b], est utilisé dans plusieurs projets de recherche européens (NitroEurope IP, CARBO-Extreme, ANIMALCHANGE, MODEXTREME) et français (CLIMATOR, VALIDATE). Le but global de ces différents projets est d'étudier les cycles du C et de l'N et les émissions de GES des prairies, ainsi que les effets du changement climatique.

Contrairement à d'autres modèles à base mécaniste incluant des fonctionnalités relatives à des applications sur la prairie, comme le modèle de grandes cultures STICS ([Brisson et al., 1998], [Brisson et al., 2002], [Ruget et al., 2006]) ou les modèles de biomes LPJ [Sitch et al., 2003], [Gerten et al., 2004], PaSim est un modèle spécifique pour la prairie. Il intègre également un module des interactions herbe-animal permettant de simuler des sorties (production, émissions) liées aux animaux au pâturage. PaSim possède une représentation détaillée des processus végétaux pour permettre des simulations sous changement climatique. En plus, PaSim est l'un des modèles les plus complets en termes de disponibilité de variables de sortie (~475) couvrant sol, flux de GES, production végétale et animale.

1.6. Problématique

Les modèles et les simulations constituent à l'heure actuelle des outils puissants et performants. Mais généralement, les valeurs simulées à l'aide des modèles sont dépendantes de nos connaissances sur le système étudié, et des hypothèses de simulation adoptées. Elles s'accompagnent donc d'incertitudes souvent importantes et qu'il est nécessaire de considérer et d'analyser.

De telles projections doivent tenir compte de l'incertitude liées aux sources multiples, y compris des scénarios climatiques, le modèle d'impact, le climat local, la végétation et les conditions du sol [Olesen et al., 2007; Soussana et al., 2010].

Le problème majeur avec l'utilisation des modèles écologiques complexes est la connaissance incomplète des variables d'entrée ainsi que les paramètres du modèle. Cela signifie que les estimations du modèle sont fortement dépendantes des paramètres, ce qui conduit à une incertitude importante due aux incertitudes liées aux valeurs des paramètres, des variables (climat, sol et gestion) et de la structure du modèle lui-même [Gabrielle et al., 2006]. Il est donc nécessaire de calibrer (estimer) les valeurs des paramètres pour obtenir une bonne cohérence entre les résultats des modèles et les observations [Wallach et al., 2011].

Dans notre cas, PaSim implique un nombre très important de paramètres (environ 150 rien que pour la végétation) pour simuler les nombreux processus en jeu dans l'écosystème prairial. Chaque paramètre contient une part d'incertitude qui lui est propre et souvent mal connue, ce qui rend, par conséquent, les sorties des modèles très incertaines.

En cas de changement climatique accompagné par une intensification d'évènements extrêmes (sécheresse, forte précipitation...) l'incertitude liée aux paramètres des modèles est accentuée et se propage aux sorties des modèles.

D'où la nécessité, d'un côté, d'une méthode pour la quantification de l'incertitude sur les paramètres et la calibration de ces paramètres dans le but de réduire cette incertitude.

Les incertitudes sont quantifiées en les exprimant en tant que fonctions de distribution de probabilité (FDP). Ensuite, des échantillons représentatifs sont prélevés à partir des FDP pour mettre en avant l'incertitude des paramètres à travers les calculs du modèle. Pour cela, nous avons opté pour la calibration Bayésienne qui se base sur le théorème de Bayes [Bayes, 1763] et qui permet d'intégrer autant d'informations que possible dans des FDP [van Oijen et al., 2005] et d'utiliser cette distribution des paramètres pour obtenir des nouvelles valeurs plus correctes des paramètres et moins incertaines. Nous allons revenir sur la calibration Bayésienne et les arguments justifiant ce choix dans le chapitre suivant.

Les questions qui se posent à ce niveau sont alors :

- Jusqu'à quel point la calibration Bayésienne peut-elle réduire l'incertitude sur les paramètres et améliorer les résultats du modèle ?
- La calibration Bayésienne est-elle en mesure de réduire l'incertitude même en cas d'évènements extrêmes ?
- À l'aide de la calibration Bayésienne, est-il possible d'arriver à un paramétrage commun valide à une échelle élargie, notamment l'Europe ?

De l'autre côté, il est coûteux (en terme de temps et de calcul) et pas nécessaire d'estimer simultanément un nombre important de paramètres, une étape de sélection d'un sous-ensemble de paramètres avant l'étape de calibration est donc nécessaire. Pour cela, nous avons eu recours à l'analyse de sensibilité, qui est l'estimation des paramètres les plus influents sur les sorties, pour déterminer les paramètres du modèle à calibrer.

Enfin, la question qui se pose concerne le coût en terme de temps de calcul pour la calibration Bayésienne et l'analyse de sensibilité et quelles sont les moyens techniques et matériels pour réduire le temps de calcul.

Où se situe ce travail ?

Cette thèse est effectuée dans le cadre d'un package scientifique de l'INRA et financée par la Région Auvergne et le Département EFPA-INRA (<http://www.efpa.inra.fr>). Elle est soutenue par deux projets européens EU-FP7 CARBO-Extreme (2009-2013) et ANIMALCHANGE (2011-2014) et se situe dans un contexte applicatif d'un "work package" (WP) dans chacun des deux projets. Nous avons déjà assisté à la réunion annuelle du projet CARBO-Extreme (Montpellier, 26-29 septembre 2011) où nous avons présenté les résultats préliminaires de notre travail. Pour le projet ANIMALCHANGE, nous avons d'abord fourni un livrable (deliverable 15.2) de 22 pages, ensuite nous avons donné un cours « training session » et une communication orale lors de la réunion annuelle (Dublin, 26-28 juin 2013). Récemment, dans le cadre d'une collaboration INRA/Embrapa (<http://www.cnptia.embrapa.br/english.html>) nous avons assuré deux cours de calibration Bayésienne et d'analyse de sensibilité, et nous avons présenté nos travaux de recherche à Campinas, Brésil du 12 au 21 novembre 2013.

1.7. Conclusion

Nous avons vu dans ce chapitre que les changements climatiques provoquant des modifications du climat actuel, risquent d'entraîner des impacts importants sur les écosystèmes prairiaux qui, eux aussi, contribuent aux changements climatiques à travers les émissions de gaz à effet de serre. Le recours à la modélisation des écosystèmes s'avère alors indispensable afin de comprendre la nature et l'ampleur de ces impacts. Pour cela, nous utilisons PaSim, un modèle simulant les prestations productives et environnementales des prairies que nous allons viser à améliorer dans ce qui suit en s'intéressant à son paramétrage.

Dans le chapitre suivant, nous dresserons un état de l'art sur les techniques utilisées pour l'étude du paramétrage des modèles, notamment la calibration des paramètres des modèles en général et la calibration Bayésienne plus en particulier, ainsi que l'analyse de sensibilité des sorties des modèles aux modifications des paramètres.

Chapitre 2 : État de l'art : calibration et statistiques Bayésienne

2.1. Introduction

Dans le chapitre précédent nous avons introduit et défini le contexte où se place notre travail et nous avons abordé la problématique et les principales attentes de cette thèse. Au cours de ce chapitre, nous allons décrire les domaines clés sur lesquels s'appuie notre thèse. Nous allons d'abord introduire les principales sources d'incertitudes liées aux modèles, en particulier les incertitudes liées aux paramètres du modèle, ensuite nous présenterons la calibration des paramètres comme une solution possible pour palier à ce problème et réduire ainsi les incertitudes. Nous montrons, en particulier, que la calibration Bayésienne est adaptée aux modèles complexes et permet d'obtenir un paramétrage moins incertain à travers une distribution à posteriori mise à jour à partir de données mesurées. La deuxième partie de ce chapitre sera consacrée aux méthodes d'analyse de sensibilité des modèles aux paramètres, permettant de ressortir les paramètres les plus influents sur les sorties des modèles afin de les calibrer ensuite.

2.2. Sources d'incertitudes

“All models are wrong, but some are useful” [Box et Draper., 1987]. Bien que les modèles soient nécessaires et utiles, ils s'accompagnent cependant d'une part d'erreur due aux différentes sources d'incertitudes. En effet, les estimations d'un modèle sont fortement dépendantes de ses paramètres, des variables d'entrée et de la structure du modèle. Ces trois facteurs représentent les principales sources d'incertitude en modélisation:

- **Incertitude liée aux paramètres :** tout d'abord il est important de signaler ce qu'un paramètre. Un paramètre est généralement une valeur numérique qui peut être constante ou variable qui représente une information apportée au modèle et permettant d'ajuster ce dernier. Dans le cas de notre modèle PaSim, les paramètres sont des valeurs numériques constantes durant la simulation. Les seules informations quant aux valeurs de ces paramètres sont les résultats d'expériences antérieures et de la littérature, et dans les meilleurs des cas nous avons une plage de variabilité pour chaque paramètre (un minimum et un maximum). Les paramètres sont donc entachés de manque de connaissance et

représentent une source importante d'incertitude qui peut affecter les résultats du modèle, et plus le nombre de paramètres d'un modèle est important plus l'incertitude est importante. C'est sur cette source d'incertitude que nous nous sommes focalisé durant ce travail pour essayer d'estimer des valeurs plus précises pour les paramètres afin de réduire leurs incertitudes et par conséquent l'incertitude liée aux variables de sorties du modèle.

- **Incertitude liée aux données :** en modélisation, la qualité des données a été toujours pointée du doigt par les modélisateurs. Ce sont des données observées ou mesurées qu'elles soient des données d'entrée (météo, gestion...) nécessaires pour faire tourner le modèle, ou des variables de sortie nécessaires pour la calibration du modèle (température du sol, biomasse, humidité...). Par définition, une variable en modélisation est une quantité qui n'est pas constante et qui varie selon l'espace et le temps auquel elle a été mesurée, elle peut être une variable d'entrée du modèle ou une variable de sortie du modèle ou les deux. Ces données sont mesurées à l'aide de plusieurs dispositifs de mesures où chaque mesure réalisée comporte des erreurs. Les sources d'erreur peuvent donc venir des instruments de mesure, des protocoles choisis, de la variation du phénomène mesuré ou tout simplement de l'erreur humaine. Ce qui peut engendrer parfois des données manquantes et/ou aberrantes nécessitant des procédures de remplissage (appelées « gap-filling ») pour combler les vides. Mais généralement les valeurs des données sont renseignées par une moyenne et un écart type ou une erreur standard.
- **Incertitude liée à la structure du modèle :** les modèles servent à simuler les différents phénomènes et scénarios mis en jeu. Ainsi, l'incertitude liée au fonctionnement du modèle représente un enjeu majeur. En effet, les variables de sortie d'un modèle sont la conséquence d'une multitude d'équations et de processus corrélés entre eux et dont la traduction informatique ne fait qu'augmenter l'incertitude liée au modèle. Tout cela représente non seulement une source d'incertitude mais aussi une source de propagation d'incertitude qui affectera certainement les résultats du modèle. Même si le fait d'étudier et de réduire l'incertitude liée à la structure du modèle ne fait pas partie de notre champ d'étude, notre travail nous a permis aussi de mettre le doigt sur

d'éventuelles limites des simulations du modèle que nous verrons plus précisément dans la partie consacrée aux résultats.

2.3. Calibration : techniques et définitions

Comme évoqué ci-dessus les paramètres représentent une source importante d'incertitude d'où le besoin de leur attribuer des valeurs fiables. Pour cela, en cas d'absence de mesures expérimentales ou de connaissances préalables sur les valeurs des paramètres, la calibration est alors une étape majeure avant la mise en œuvre des modèles. En effet, les modèles impliquent un nombre très important de paramètres pour simuler des nombreux processus. Chaque paramètre contient une part d'incertitude qui lui est propre et souvent mal connue, ce qui rend les sorties des modèles très incertaines. La mise au point de méthodes de calibration des paramètres avec une quantification de l'incertitude des paramètres et des sorties du modèle est un défi primordial [Sutton et al., 2007]. L'utilisation des données observées joue un rôle indispensable dans la calibration et l'évaluation des modèles [Beck, 1987]. La calibration est un aspect du concept de "fusion modèle-données" (*model-data fusion*) défini par [Williams et al., 2009] comme une série de procédures combinant des données avec un modèle en quantifiant l'incertitude respective des deux parties.

L'objectif de l'étape de calibration (ou d'estimation) des paramètres est de proposer une valeur approchée et vraisemblable des paramètres du modèle à partir des données observées. Deux grandes familles de méthodes d'estimation se distinguent selon une approche fréquentiste ou une approche Bayésienne.

2.3.1. Les approches fréquentistes

Les approches fréquentistes utilisent des méthodes d'estimation qui permettent d'approcher une valeur fixe du vecteur de paramètres en utilisant un échantillon de données [Makowski et Jeuffroy, 2002 ; Makowski et al., 2006]. Ces approches tirent leur nom du fait de considérer la tendance d'un système à produire un événement donné selon sa fréquence d'apparition. Elles utilisent des fonctions appelées estimateurs qui permettent de relier les valeurs estimées des paramètres aux données observées. Voici quelques exemples de méthodes d'estimation:

La méthode des moindres carrés

C'est la méthode la plus connue et probablement la plus utilisée pour la calibration des modèles. La méthode des moindres carrés [Legendre, 1805] est un algorithme classique visant à obtenir une meilleure approximation des valeurs des paramètres. Son principe consiste à minimiser la somme des carrés des écarts entre les observations et les prédictions du modèle, la quantité à minimiser peut s'écrire ainsi [Wallach et al., 2011]:

$$Res = \arg \min \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$

Où $f(x_i; \theta)$ représente le résultat du modèle, y_i les mesures et θ le vecteur des paramètres.

Ils existent différentes formes d'application de cette méthode : la méthode des moindres carrés ordinaires, par exemple, est utilisée pour calculer l'estimation du vecteur des paramètres quand on a un seul type de mesure disponible. Cette méthode est également appliquée pour des modèles linéaires et non-linéaires avec des algorithmes itératifs. La méthode des moindres carrés pondérés, quant à elle, permet d'estimer les paramètres en présence de plusieurs types de mesures disponibles. L'estimateur des moindres carrés pondérés est utilisé quand on connaît la matrice de variance-covariance des paramètres car il permet de minimiser la somme des carrés des écarts pondérés par les variances sur les mesures obtenues par les itérations. On peut aussi citer la méthode des moindres carrés élagués où il s'agit uniquement de minimiser la somme des carrés des erreurs ayant les valeurs les plus petites.

Cependant, ces méthodes présentent des limites en cas où le nombre de paramètres à estimer est élevé car, plus le nombre de paramètres augmente, plus la dimension de l'espace à parcourir pour trouver le minimum augmente. À cela s'ajoute la forte dépendance du résultat à la valeur de départ du vecteur des paramètres du modèle ainsi que la possibilité de la convergence de l'algorithme vers un minimum local.

L'algorithme espérance-maximisation

En statistique, l'algorithme espérance-maximisation (EM, en anglais *expectation-maximisation* algorithm) [Dempster et al., 1977]) est un algorithme utilisé pour

l'estimation des paramètres quand les données dont on dispose sont incomplètes. D'une manière générale, l'EM vise à fournir un estimateur lorsque l'on a des données manquantes et que le fait de connaître ces données permettrait d'estimer les paramètres. Pour cela, l'EM se base sur la vraisemblance maximale de l'ensemble des données pour compléter les séries des observations recueillies.

Tel que son nom l'indique, l'algorithme EM nécessite l'exécution de deux étapes distinctes à chaque itération dont voici une brève description :

- L'étape d'espérance ou « étape E » permet d'estimer les données inconnues étant données les données observées et la valeur des paramètres estimée à l'itération précédente. Cette phase permet d'obtenir une quantité $Q(\theta, \theta^{i-1})$ qui représente l'espérance de la vraisemblance de l'ensemble des données connaissant les paramètres θ du modèle.
- L'étape de maximisation ou « étape M » utilise l'estimation des données inconnues effectuée à l'étape d'Espérance, vise à maximiser la vraisemblance et met à jour la valeur des paramètres pour la prochaine itération.

$$\theta_i = \arg \max Q(\theta, \theta^{i-1})$$

- Itération de ces étapes jusqu'à ce que la variation des paramètres soit presque nulle entre deux itérations (Figure 5).

Le défaut principal de l'algorithme EM tient aux problèmes de convergence vers un maximum local de vraisemblance, ce qui explique le recours à d'autres algorithmes complémentaires comme le « Stochastic » algorithme (SEM) et le « Generalized » algorithme (GEM) [Dempster et al., 1977] qui simplifient la convergence et la complexité de l'algorithme de base.

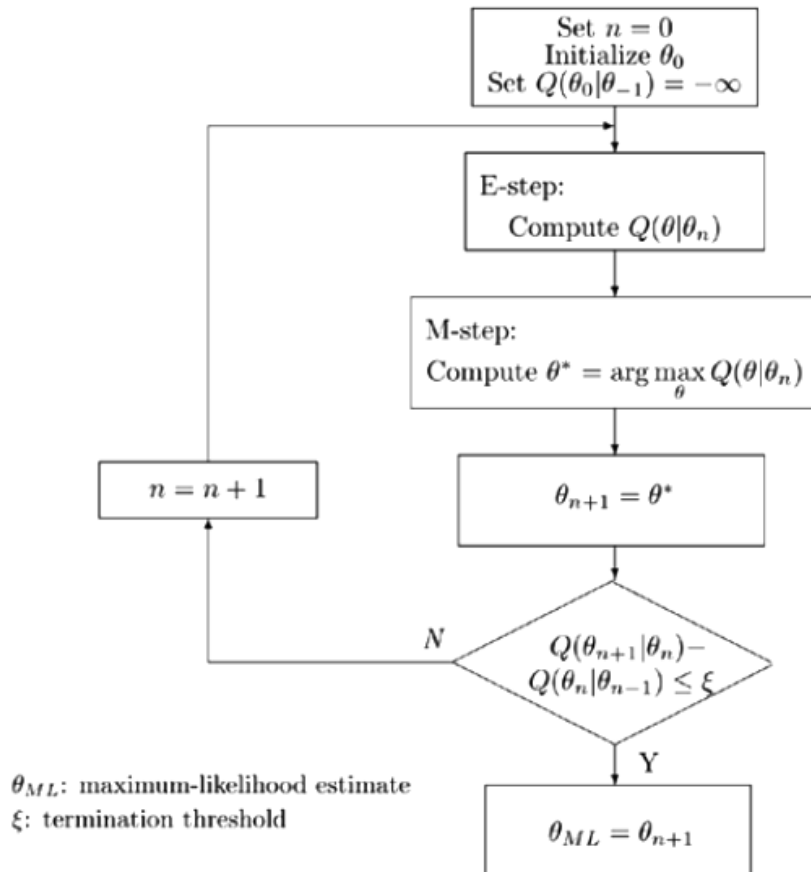


Figure 5. Fonctionnement de l’algorithme de base d’espérance-maximisation [Kung et al., 2005]

L’algorithme de Levenberg-Marquart

L’algorithme de Levenberg-Marquart a été développé par Kenneth Levenberg [Levenberg, 1944] et repris et publié par Donald Marquardt [Marquardt, 1963]. Il permet de résoudre des problèmes numériques de minimisation pour des fonctions non-linéaires et est adapté à des problèmes de taille minime ou moyenne. C’est un algorithme itératif dont le principe consiste à partir d’une valeur initiale d’un paramètre P à lui rajouter une quantité Q et à essayer de déterminer la valeur de Q en minimisant l’équation utilisée dans le calcul.

Cette méthode présente un problème de convergence car si le vecteur de départ des paramètres est trop éloigné de la solution, l’algorithme risque alors de ne pas converger [Transtrum et Sethna, 2012] ou au meilleur des cas la convergence risque d’être très lente.

En plus, cette méthode nécessite un volume de calculs important à chaque itération suivant le nombre de paramètres.

Méthode du simplexe

La méthode du simplexe est une technique d'optimisation élaborée par [Nelder et Mead, 1965]. Elle est basée sur un concept géométrique appelé simplexe qui consiste en $n+1$ sommets (interconnectés) dans un espace de n dimensions (ex. dans un espace de 2 dimensions le simplexe est un triangle). L'algorithme qui parcourt les sommets du simplexe se déplace dans un espace multidimensionnel et à chaque itération le but est d'obtenir un nouveau simplexe permettant de se rapprocher de l'optimum des valeurs des paramètres considérés.

L'inconvénient de cette méthode c'est que les données observées ne sont pas prises en compte directement dans les calculs, il faut donc à chaque fois calculer des indices de comparaison entre les observations et les simulations (ex. Root Mean Square Error [RMSE]) pour déterminer si les valeurs des paramètres obtenus sont acceptables [Confalonieri et Acutis, 2006].

Il y a aussi d'autres concurrents de l'approche Bayésienne qui peuvent inclure des observations et analyser les incertitudes. On peut citer par exemple, la logique floue qui tient compte des imprécisions des données et des incertitudes [Kaufman, 1973] et qui est appliquée dans des nombreux domaines comme la robotique, le traitement d'image et la médecine, par contre cette technique se base sur des variables qualitatives et ne peut donc pas être adapté à la modélisation des systèmes quantitatifs complexes. D'autres techniques suscitent l'intérêt, comme les techniques de réseaux de neurones qui permettent d'établir des prévisions et de construire des modèles, néanmoins les réseaux de neurones rendent la calibration des modèles lourde et complexe [Dreyfus et al., 2002].

2.3.2. L'approche Bayésienne

Historique

«Bayesian statistics is not a branch of statistics, it was a way of looking at the whole of statistics» [Lindley, 1971].

Dans la littérature du début du 20^{ème} siècle, la statistique Bayésienne était considérée comme une approche statistique basée sur l'arbitraire dans le choix des lois de probabilité et n'a pas été fréquemment utilisée dans différents domaines. Ce n'est qu'à partir des deux dernières décennies que la statistique Bayésienne commence à regagner du terrain et à susciter l'intérêt, notamment grâce à l'amélioration des moyens de calcul coté matériels et coté logiciels.

Plus récemment, l'approche Bayésienne fut adoptée considérablement par plusieurs disciplines que ce soit en finance, en médecine ou en économie. La présence renforcée des statistiques Bayésiennes peut se constater à travers les publications des grands journaux de statistique comme *Annals of Statistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society* ou *Biometrika*. Dernièrement, il est devenu courant de voir paraître des articles Bayésiens en écologie où la statistique Bayésienne commence à devenir un outil puissant et efficace pour l'estimation et la modélisation des différents processus écologiques [Clark, 2005].

L'adoption de l'approche Bayésienne par les praticiens des différents domaines est justifiée par des motifs pratiques comme l'intégration des informations et de l'incertitude sur ces informations, l'obtention de meilleurs résultats en sortie et la facilité de calcul. Deux livres récents [Gelman et al., 2003] et [Carlin et Louis, 2009] montrent l'intérêt de l'analyse de données Bayésienne et évoquent même l'aspect philosophique derrière. Tout cela prouve la pertinence de l'approche Bayésienne qui a fini par s'imposer dans nombreuses disciplines notamment grâce à sa capacité de résoudre plus facilement et plus naturellement les problèmes [Robert, 2001].

Dans cette approche, la probabilité représente une mesure de crédibilité en tenant compte des informations disponibles. Cette probabilité s'appuie sur le théorème de Bayes [Bayes, 1763] où les probabilités attribuées initialement à un événement peuvent différer d'un observateur à un autre mais elles évolueront ensuite de manière convergente en intégrant les nouvelles informations selon les critères donnés par la formule de Bayes [Carranza et Kuzniak, 2007]. Cette formule que nous allons détailler dans la section suivante est importante car elle représente aussi la base des approches Bayésiennes.

Le principe de la calibration Bayésienne

Dans un contexte Bayésien, les paramètres du modèles (noté θ) prennent le statut de variables aléatoires. C'est-à-dire que le paramètre n'a plus une valeur fixe inconnue mais il est considéré comme une variable aléatoire dont la distribution caractérise notre degré de connaissance à son égard (c'est l'incertitude) et son caractère plus ou moins fixé (c'est la variabilité). L'approche Bayésienne permet donc de reprendre les mêmes paramètres en les considérant comme aléatoires du fait de la connaissance imparfaite que nous avons. Avant d'intégrer les données expérimentales, une distribution de probabilité *a priori* (« prior »), noté $P(\theta)$, est attribuée pour chaque paramètre inconnu. Cette distribution reflète la connaissance sur θ et représente le degré de croyance sur les valeurs possibles de θ .

Les distributions *a priori* sont ensuite mises à jour avec les données expérimentales [Makowski et al., 2006; Gallagher et Doherty, 2007], pour obtenir à la fin une distribution de probabilité *a posteriori* (« posterior ») incluant l'information contenue dans les observations [Van Oijen et al., 2005]. La mise à jour est réalisée à travers le théorème de Bayes [Bayes, 1763] qui exprime la distribution de probabilité *a posteriori* notée $P(\theta/D)$ comme une distribution conditionnelle des paramètres par rapport aux données (Figure 6). Le posterior est alors proportionnel au produit de tous les prior P_θ , et de la vraisemblance des données notée $P(D/\theta)$ et qui est en fait la distribution des données conditionnées par les valeurs des paramètres:

$$P(\theta/D) = \frac{P(\theta) \times P(D/\theta)}{P(D)}$$

Dans cette formule, $P(D)$ représente une constante nécessaire pour normaliser la distribution de probabilité *a posteriori* $P(\theta/D)$. La démarche statistique Bayésienne est donc la synthèse des deux sources d'information sur les paramètres d'intérêt: d'une part celle apportée par le prior (censée représenter notre état de connaissances généralement apporté par la littérature, les expériences et les observations antérieures) et d'autre part celle convoyée par les données à travers la modélisation probabiliste que représente la vraisemblance.

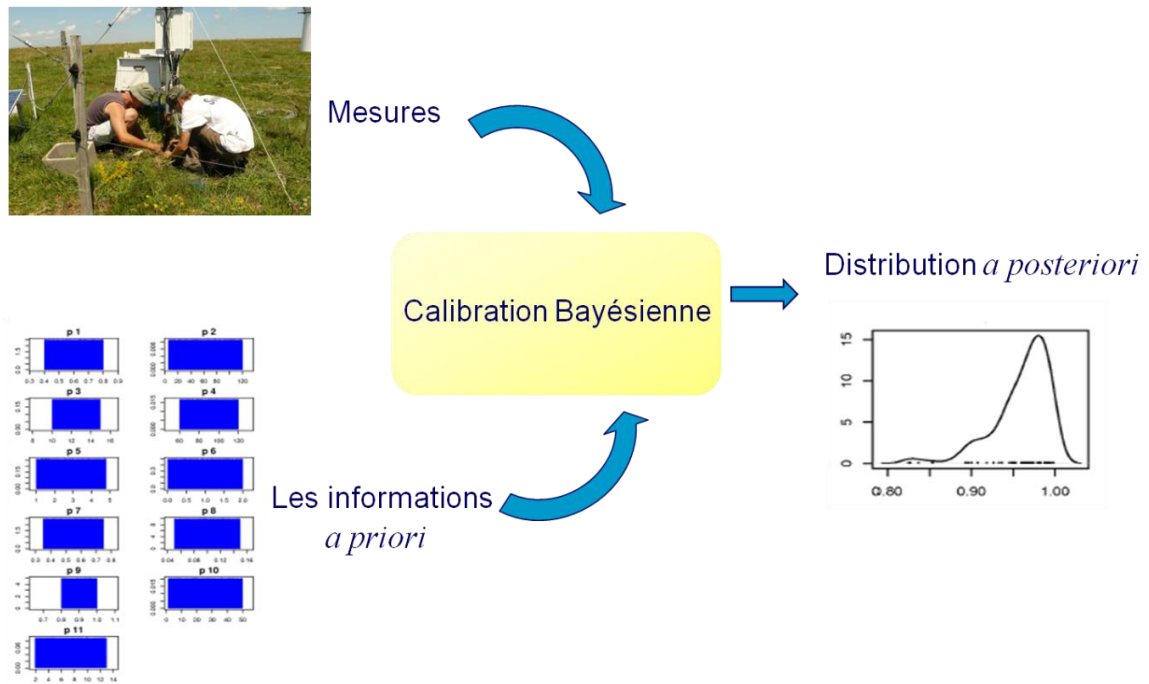


Figure 6. Principe de la calibration Bayésienne

Selon le principe du théorème de Bayes, la mise en œuvre de cette approche nécessite donc deux éléments en entrée: la distribution *a priori* des paramètres et la définition de la vraisemblance à partir des données mesurées, pour l'obtention d'un nouveau élément en sortie : la distribution *a posteriori* (Figure 6).

Distribution a priori

La distribution *a priori* des paramètres n'intègre pas les informations relatives aux données, elle en est donc indépendante. Cette distribution peut être déduite à la lumière des connaissances collectées dans la littérature et/ou à travers des expériences antérieures (Figure 7). Il n'y a pas de règles d'or dans le choix de cette distribution, tout dépendra en fait de ce que nous avons comme informations entre les mains, d'où la classification des distributions *a priori* en deux types: les distributions informatives et les distributions non-informatives [Brochot, 2004].

Les distributions dites non-informatives s'imposent dans le cas où nous n'avons pas assez d'informations sur la valeur des paramètres (ex. plage de variabilité), alors les lois uniformes par exemple sont privilégiées.

Réciproquement, les distributions dites informatives reposent sur des informations plus précises et plus disponibles, ce qui permet de fournir une distribution a priori plus développée, dans ce cas les distributions normales ou log-normales sont privilégiées.

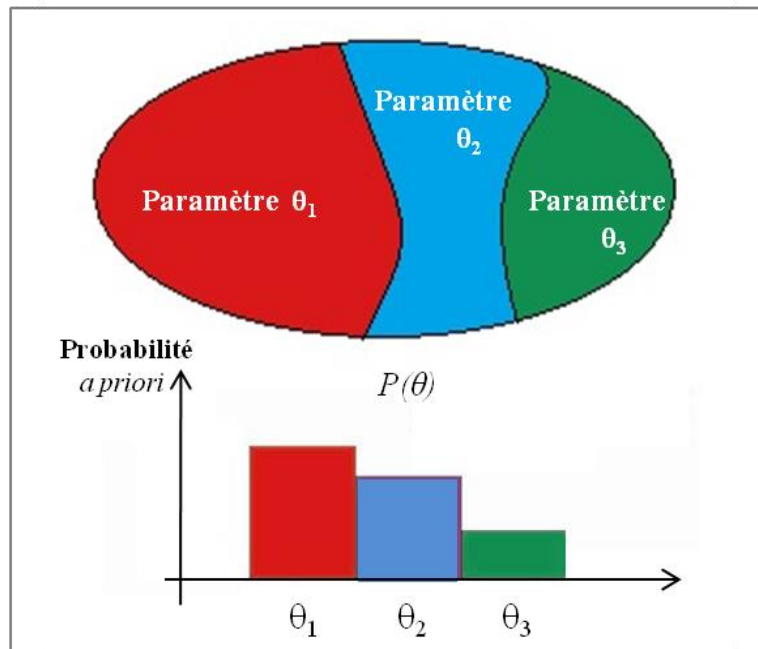


Figure 7. Première étape du principe Bayésien : la connaissance a priori sur les paramètres

La vraisemblance

La vraisemblance $P(D/\theta)$ est la distribution des données sachant les paramètres. Pour chaque ensemble de données D_i , le logarithme de la vraisemblance des données L_i , qui est la probabilité des données étant donné les paramètres, est calculé comme suit :

$$\log L_i = \sum_{j=1}^k \left(-0.5 \left(\frac{f_i(\omega; \theta) - D_j}{\sigma_j} \right)^2 - 0.5 \log(2\pi) - \log(\sigma_j) \right)$$

Où D_j est la valeur des données mesurées à la date d'échantillonnage j dans l'ensemble de données D_i , σ_j est l'écart type des données à cette date et qui représente l'erreur sur les mesures qu'il faut bien prendre en compte, ω est le vecteur de données d'entrée (« inputs ») du modèle à la même date, $f_i(\omega; \theta)$ est la simulation du modèle de D_j avec le vecteur de paramètres θ (généré par l'algorithme d'échantillonnage), et K est le nombre total des dates d'observations dans l'ensemble de données.

À travers la vraisemblance et à l'aide des méthodes d'échantillonnage que nous allons détailler dans la section suivante, seules les valeurs des paramètres les plus vraisemblables sont conservées après les simulations. Ce qui permet d'obtenir des valeurs de paramètres plus précises en intégrant les observations et les erreurs qui les entachent dans le calcul (Figure 8).

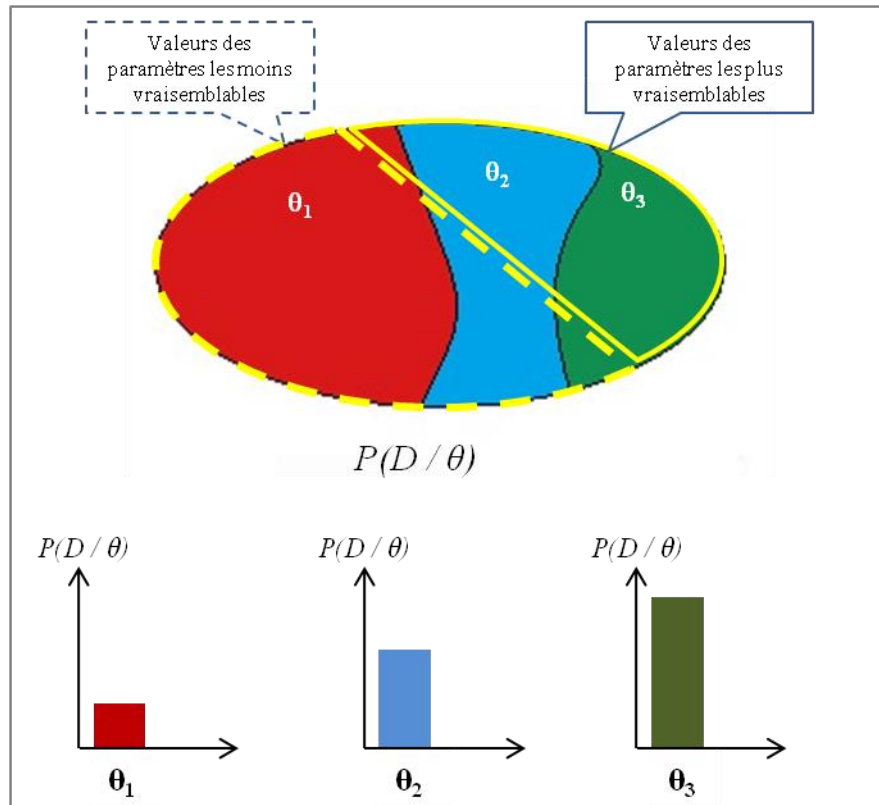


Figure 8. Deuxième étape du principe Bayésien : le calcul de la vraisemblance

Distribution a posteriori

La loi *a posteriori* $P(\theta/D)$ donne le degré de crédibilité des valeurs possibles de l'état de la nature θ sachant que le résultat expérimental D s'est produit (Figure 9).

La formule de Bayes est donc une formule de mise à jour de la connaissance *a priori* en une connaissance *a posteriori* grâce aux données D introduite à travers le calcul de la vraisemblance.

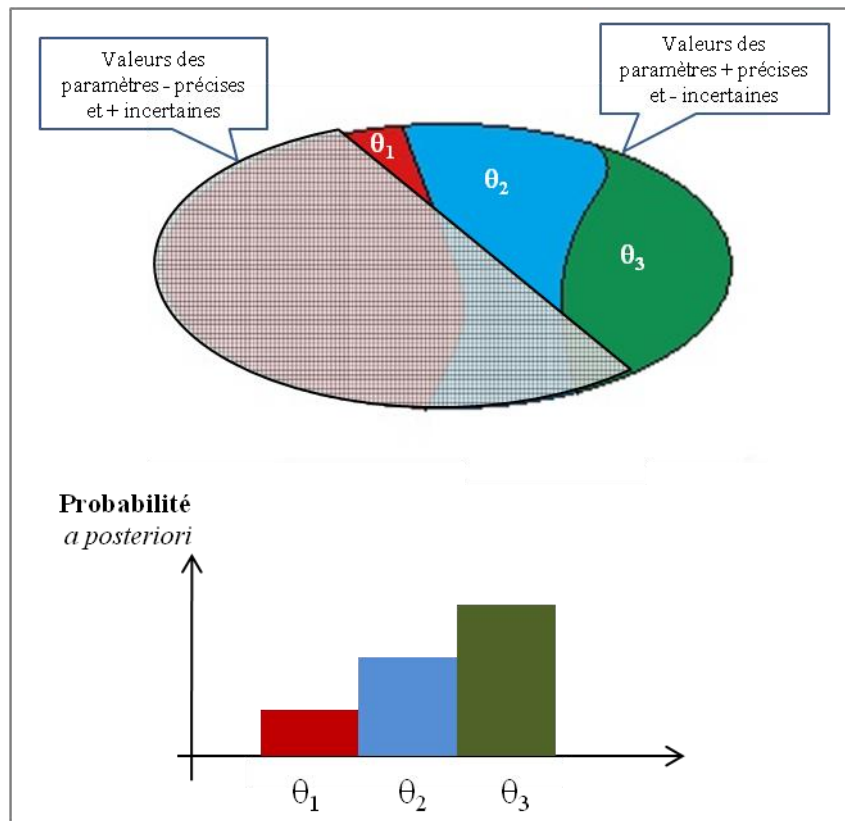


Figure 9. Troisième étape du principe Bayésien : l'obtention d'une probabilité a posteriori plus précise sur les paramètres

Pour pouvoir simuler la nouvelle loi a posteriori, il est nécessaire d'avoir recours à des techniques d'échantillonnage telles que les méthodes Monte Carlo par Chaînes de Markov (MCMC) qui permettent d'effectuer des tirages aléatoires et de générer ainsi des échantillons tirés de la distribution *a posteriori* des paramètres du modèle.

2.3.3. Méthodes Monte Carlo par Chaînes de Markov

Les Méthodes Monte Carlo par Chaînes de Markov (« Monte Carlo Markov Chain ») sont des méthodes d'échantillonnage permettant d'effectuer des tirages à partir des distributions de probabilité. L'apparition de ce genre de moyens de calcul puissant s'est avérée efficace pour l'application de l'approche Bayésienne, et l'association de ces deux méthodes permet l'estimation de paramètres au sein des modèles complexes [Pinheiro et Bates, 2000].

Face à la complexité des modèles et des calculs nécessaires pour la calibration, les méthodes MCMC basées sur l'algorithme de Metropolis-Hastings [Metropolis et al., 1953]

présentent une solution pour contourner ces difficultés et pour générer un échantillon représentatif de la distribution *a posteriori* à partir du vecteur des paramètres.

Parmi les algorithmes permettant d'appliquer les méthodes MCMC, l'algorithme Metropolis-Hastings qui est l'un des algorithmes les plus simples et les plus performants en raison de son adaptation aux modèles complexes basés sur les processus et de toutes tailles (plus particulièrement les modèles de ≥ 10 paramètres), et parce qu'il ne nécessite pas de connaissances préalables sur la forme de la distribution *a posteriori* à générer.

L'algorithme de Metropolis est séquentiel dans le sens où il crée une « marche » à travers l'espace de paramètres de telle sorte que la chaîne des points visités tend vers l'échantillon recherché à partir du posterior. C'est ce qu'on appelle une marche aléatoire où chaque nouveau point de la chaîne est trouvé en générant aléatoirement un vecteur de paramètres candidats, et chaque paramètre candidat peut être accepté ou rejeté. Le vecteur de paramètres candidats est généré à travers un « pas » d'une distribution normale multivariée à partir du vecteur actuel. Comme pour cela, nous avons besoin d'une matrice de variance pour définir cette distribution normale multivariée, la matrice proposée est une simple matrice diagonale [Gelman et al., 1996]. L'acceptation d'un paramètre candidat proposé ou son rejet dépend du rapport de Metropolis, qui représente le rapport de deux produits: la vraisemblance du candidat et la vraisemblance du point courant (voir Algorithme 1). Si le ratio de Metropolis est supérieur à 1, c'est-à-dire que la probabilité *a posteriori* du candidat est supérieure à la probabilité du point actuel alors il est accepté. Si le ratio de Metropolis est inférieur à 1, c'est-à-dire que le candidat est « moins probable » que le vecteur actuel, le candidat peut encore être accepté mais seulement avec une probabilité égale au ratio de Metropolis, si non le candidat est rejeté. Le taux d'acceptation optimal est entre 25% et 45% selon le nombre de paramètres calibrés [Gelman et al., 2003]. La chaîne s'arrête quand il y a « convergence », c'est à dire qu'elle a exploré l'espace des paramètres de façon adéquate. La convergence peut être confirmée visuellement en utilisant les tracés de courbes des différents paramètres qui montrent comment la chaîne se déplace à travers l'espace de chaque paramètre (Figure 10). Pour assurer la convergence, il existe plusieurs solutions comme, par exemple, augmenter la taille de la chaîne c'est-à-dire le nombre d'itérations ou bien lancer plusieurs chaînes en parallèles mais avec des points de départ différents.

À la fin de la calibration, nous obtenons donc une distribution *a posteriori* des paramètres, c'est-à-dire un ensemble de valeurs pour chaque paramètre. L'approximation par maximum de la FDP nous permet alors d'obtenir la nouvelle valeur unique pour chaque paramètre que nous pouvons utiliser ensuite dans les simulations du modèle.

- Requis :**
- (1) Un modèle: M
 - (2) Des données: D
 - (3) Une distribution de probabilité *a priori* pour les paramètres: $P(\theta)$
 - (4) La fonction de vraisemblance: $P(D/\theta)$ qui est une fonction de la différence $M(\theta)-D$

- Implémentation:**
- (1) Sélectionnez un point de départ dans l'espace des paramètres, $\theta^{(0)}$
 - (2) Calculer le produit du *prior* et de la vraisemblance:

$$\pi^{(0)} = P(\theta^{(0)}) \cdot P(D/\theta^{(0)})$$
 - (3) Initialiser $i=0$ (nombre de candidats acceptés)
 - (4) Exécutez la boucle suivante jusqu'à ce que nous ayons suffisamment d'échantillon de points:
 - (i) Générer un nouveau point aléatoire, $\theta^{(candidat)}$
 - (ii) Calculer $\pi^{(candidat)} = P(\theta^{(candidat)}) \cdot P(D/\theta^{(candidat)})$
 - (iii) Calculer le ratio de *Metropolis*: $\pi^{(candidat)} / \pi^{(i)}$
 - (iv) Accepter le candidat ayant une probabilité égale à *min (ratio Metropolis, 1)*
 - (v) Si le candidat est accepté: $\theta^{(i+1)} = \theta^{(candidat)}$, sinon: $\theta^{(i+1)} = \theta^{(i)}$
 - (vi) Incrémenter $i=i+1$

- Résultat:**
- (1) Un échantillon représentatif $\{\theta^{(0)} \dots \theta^{(n)}\}$ à partir de la fdp du *posterior*, $P(\theta/D)$

Algorithme 1. Description de l'algorithme de Metropolis-Hastings

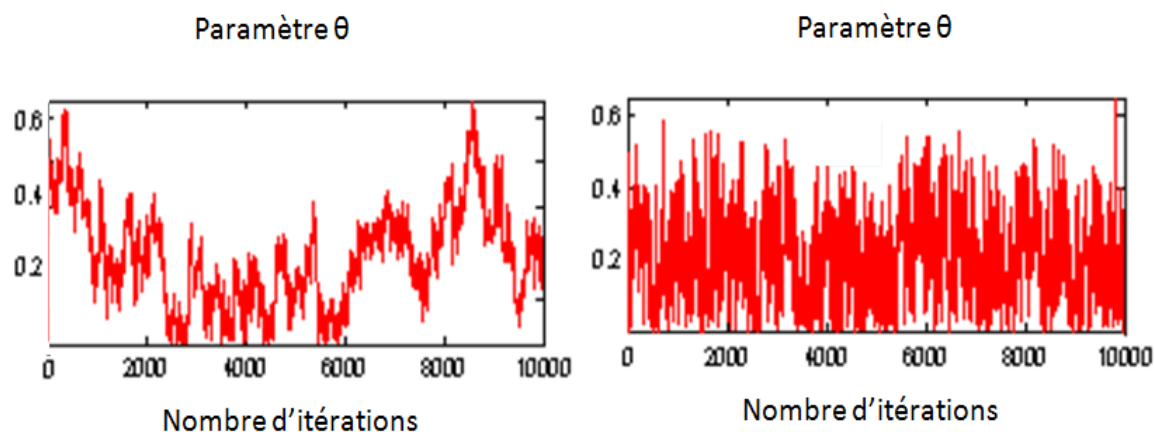


Figure 10. Exemple d'une chaîne de valeurs d'un paramètre θ générée par la méthode MCMC de longueur 10 000 : à droite convergence atteinte et à gauche convergence non atteinte

Choix de la méthode Bayésienne

La calibration des modèles éco-systémiques constitue un problème complexe vu la grande variabilité des processus à reproduire, le nombre de paramètres qui est souvent important avec une forte incertitude ainsi que les incertitudes liées aux données et au modèle lui-même. Ce qui est le cas, par exemple, de notre modèle d'application PaSim dont le code brut est écrit en plus que 60 000 lignes comptant approximativement 150 paramètres pour la végétation et quasiment 500 variables de sortie. Cela témoigne effectivement de la complexité du modèle et de l'importance du nombre des paramètres qui y sont inclus.

Le choix de la calibration Bayésienne vient donc d'abord du fait que cette technique s'applique à tout type de modèle quelle que soit la taille et surtout pour les modèles comptant un nombre important de paramètres. En plus, cette méthode permet de combiner le paramétrage du modèle avec l'analyse d'incertitude, elle permet d'intégrer et de synthétiser plus facilement diverses sources d'information (données, modèles, expertises) [Figure 11]. Toutefois, la principale caractéristique de l'approche Bayésienne est qu'elle quantifie les paramètres et les sorties du modèle sous la forme de distributions de probabilité, et applique les règles de la théorie des probabilités à travers la vraisemblance pour mettre à jour ces distributions à travers les données disponibles [Sivia, 1996]. Bien que quelques logiciels spécifiques de calibration Bayésienne existent comme, par exemple, Winbugs [Lunn et al., 2000], la méthode Bayésienne n'a pas vraiment besoin d'un logiciel

ou outil spécifique pour la faire tourner, pouvant être développée sous plusieurs environnements et outils génériques de calcul scientifique (ex. R, MATLAB...) ou codé en langages informatiques ayant des bibliothèques statistiques (ex. C/C++, Java...). Ces atouts majeurs de la calibration Bayésienne la distinguent des autres approches permettant principalement d'utiliser la nouvelle distribution *a posteriori* des paramètres pour obtenir des nouvelles valeurs plus précises et moins incertaines [van Oijen et al., 2005].

Toutes ces raisons confortent le choix de la méthode, qui était déjà en partie suggéré par le cadre applicatif de cette thèse notamment par les projets européens CARBO-Extreme et ANIMALCHANGE. En effet, même si une bonne marge de liberté était accordée dans le choix du modèle et des sites à utiliser, les données d'entrée utilisées dans la thèse sont en grande partie fournies par les projets européens et certaines simulations ont été réalisées avec PaSim pour répondre aux différents objectifs mis en jeu dans le cadre de ces mêmes projets.

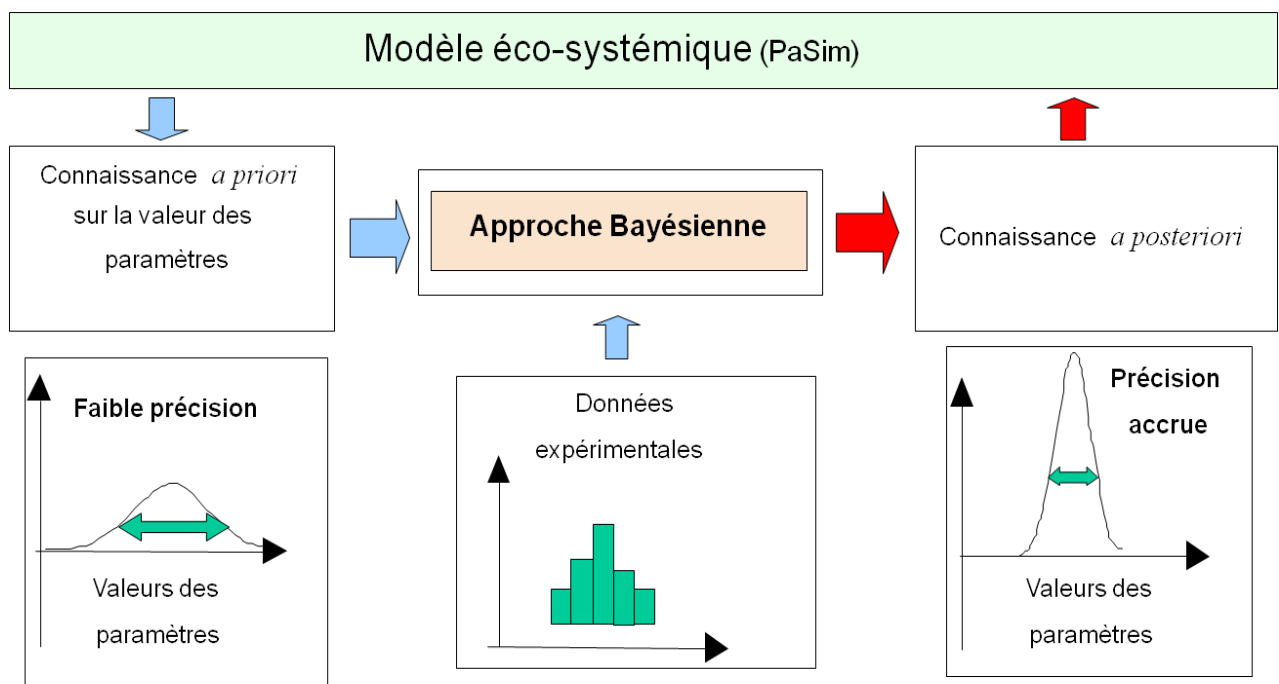


Figure 11. Interaction entre modèle, données et expertise grâce à l'approche Bayésienne pour l'amélioration des résultats du modèle et la réduction d'incertitude sur les paramètres

2.4. Analyse de sensibilité

De plus en plus complexes, les modèles de simulations impliquent un nombre très important de paramètres pour simuler les nombreux processus en jeu. Mais il est difficile et coûteux d'estimer simultanément un nombre important de paramètres, ce qui ne s'avère pas toujours nécessaire. Donc, une étape de sélection d'un sous-ensemble de paramètres avant l'étape de calibration est primordiale.

Cette étape c'est l'analyse de sensibilité qui est la détermination des paramètres les plus influents sur les sorties du modèle. L'analyse de sensibilité permet d'étudier le comportement d'un modèle à travers ses paramètres. Dans notre cas, le but de l'analyse de sensibilité est d'identifier les paramètres à calibrer, ce qui permet de réduire et cibler l'effort de paramétrage du modèle [Rabitz., 1989]. Jusqu'ici, les études de sensibilité menées sur des modèles complexes tel que PaSim [Graux., 2011] ou des modèles moins complexes tel que STICS [Varella et al., 2010] sont rares et n'étaient basées que sur une analyse relativement locale et limitée à un nombre restreint de sites (climat, type de gestion) et de variables de sortie.

L'analyse de sensibilité permet d'évaluer le changement dans les sorties due au changement dans les valeurs des paramètres du modèle (ces derniers sont générées par échantillonnage à partir de la distribution des paramètres). Par conséquent, l'analyse de sensibilité fournit une méthode utile pour identifier les propriétés et renforcer la compréhension du système étudié [Saltelli et al., 2000]. La distinction entre paramètres influents (pertinents) et non influents se base généralement sur les résultats de l'analyse de sensibilité [Cariboni et al., 2007]. En classant les paramètres des modèles (« ranking ») par ordre d'importance [Cryer et Havens, 1999], l'analyse de sensibilité offre un outil efficace pour la compréhension des modèles dans la perspective de la calibration. Face au grand nombre de paramètres requis par les modèles et au temps de calcul nécessaire, généralement trop coûteux, pour les simulations, l'analyse de sensibilité permet de n'extraire que les paramètres les plus influents afin de les calibrer.

L'importance de l'analyse de sensibilité ne vient pas seulement des informations sur la sensibilité d'un modèle à ses paramètres mais aussi à travers les informations sur leurs interactions.

En analyse de sensibilité, un paramètre, ou une entrée, est appelé un facteur. Il peut être qualitatif ou quantitatif [Kleijnen et al., 2005]. Chaque facteur peut prendre différentes valeurs que l'on appelle niveaux. Dans les simulations expérimentales, on change les valeurs de ces facteurs dans le but d'analyser les effets d'un tel changement [Kleijnen et van Groenendaal, 1992].

Différentes méthodes d'analyse de sensibilité existent [Confalonieri et al., 2010], le choix de la méthode se fait en fonction du cas d'étude et des propriétés du modèle, notamment le nombre de paramètres et le temps d'exécution des simulations (Figure 12). On peut distinguer des méthodes d'identification et des méthodes de quantification de l'influence des paramètres, si le nombre de paramètres est élevé, les méthodes de quantification peuvent être appliquées dans un deuxième temps, seulement sur les paramètres identifiés comme les plus influents

Par exemple, pour un modèle où le nombre de paramètres à analyser est < 20 et le temps d'exécution est inférieur à une minute, il est préférable d'utiliser une méthode basée sur la variance [Makowski et al., 2006b]. Pour un nombre de paramètres compris entre 20 et 100 et un temps d'exécution compris entre une minute et une heure, une méthode de régression ou de Morris [Morris, 1991; Campolongo et al., 2007] semble la plus adaptée, et tel est notre cas avec PaSim. Lorsque le nombre de paramètres à estimer est > 100 et/ou que le temps d'exécution est supérieur à une heure, des méthodes locales de type plan factoriel fractionnel [Varma et al., 1999], sont plutôt envisageables (Tableau 1).

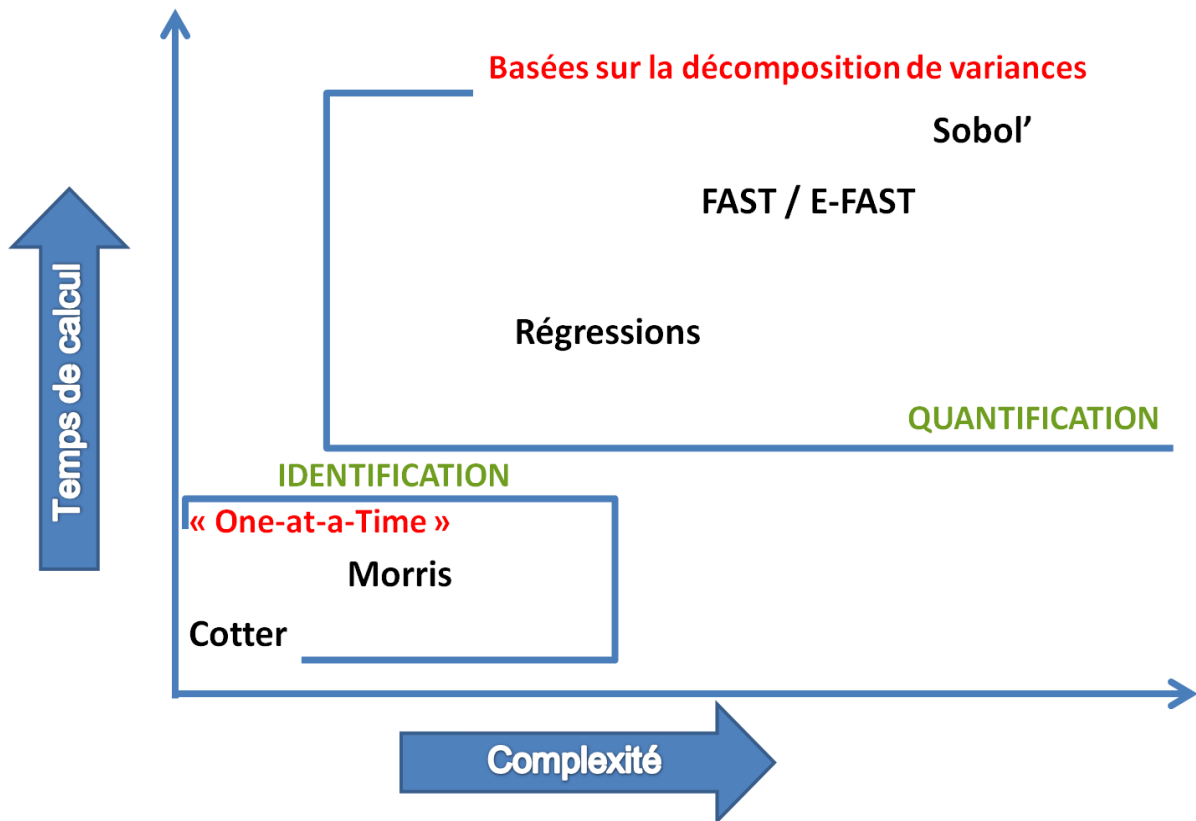


Figure 12. Classification des méthodes d'analyse de sensibilité selon la complexité et le temps de calcul

2.4.1. La méthode de Morris

La méthode de Morris initialement développée par [Morris, 1991] et modifiée par [Campolongo et al., 2007], est connue sous le nom de méthode à effet élémentaire. Elle permet l'identification des facteurs importants d'un modèle, y compris ceux qui sont impliqués dans les interactions. Elle s'appuie sur le principe OAT (« one-factor-at-a-time ») qui se caractérise par le fait que l'on ne change qu'un facteur à la fois (autrement dit on ne se déplace que dans une seule direction) par rapport à la position précédente. Ce plan d'expérience OAT est généralement utilisé lorsque le nombre de paramètres du modèle est assez élevé, ce qui exige des simulations coûteuses. Pour chaque paramètre, deux indicateurs de sensibilité sont obtenus, tous deux basés sur le calcul du ratio incrémental sur différents points dans l'espace d'entrée des paramètres. Ce ratio incrémental est le rapport entre la variation des sorties du modèle en deux points différents de l'espace d'entrée (où on fait varier un seul paramètre à la fois) et l'amplitude de la variation du paramètre lui-même. La méthode de Morris calcule l'effet élémentaire (R_i) associé à chaque facteur d'entrée en utilisant l'équation suivante:

$$R_i(x_1, \dots, x_n, \Delta) = \frac{y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_n) - y(x_1, \dots, x_i, \dots, x_n)}{\Delta}$$

Où $y(X)$ est la sortie, $X = (x_1, x_2, \dots, x_n)$ est le vecteur à n dimensions de facteurs étudiés, Δ est un pas entre $1/(p-1)$ and $1-1/(p-1)$, p est le nombre de niveaux (nombre discret de valeurs, qui sont choisis dans la plage d'incertitude des paramètres).

La méthode permet d'échantillonner des valeurs de X à partir de l'hyperespace Ω (identifié par une grille de n -dimensions et de p -niveaux) et calcule enfin la moyenne (μ , évaluant l'influence globale du facteur sur $y(X)$) et l'écart-type (σ , estimant l'ensemble des effets, i.e. non-linéarité ou interactions avec d'autres facteurs) de tous les R_i obtenus pour chaque facteur. Le nombre total d'évaluations du modèle nécessaire est $t(n+1)$, où t est le nombre de trajectoires. Les valeurs de μ et σ sont calculées sur des trajectoires différentes qui consistent en des expériences OAT individuelles. La première mesure (μ) est obtenue en calculant le ratio incrémental à différents points de l'espace des entrées, et en prenant ensuite la moyenne de leurs valeurs absolues [Campolongo et al., 2007; Pujol, 2009]. La seconde mesure (σ), l'écart-type de la distribution du ratio incrémental est importante pour la détection de paramètres en éventuelles interactions avec d'autres, ou dont l'effet est non linéaire [Saltelli et al., 2004]. Une valeur élevée de μ indique une entrée avec une influence importante sur la sortie (effet total), tandis qu'une valeur élevée de σ indique soit un paramètre à effet non linéaire sur la sortie, soit un paramètre impliqué dans l'interaction avec d'autres paramètres. Les paramètres les plus influents alors sont ceux situés dans la partie supérieure droite du graphe représentant σ (la propagation) par rapport à μ (la force) [Figure 13], où les deux mesures de sensibilités sont élevées (zone rouge).

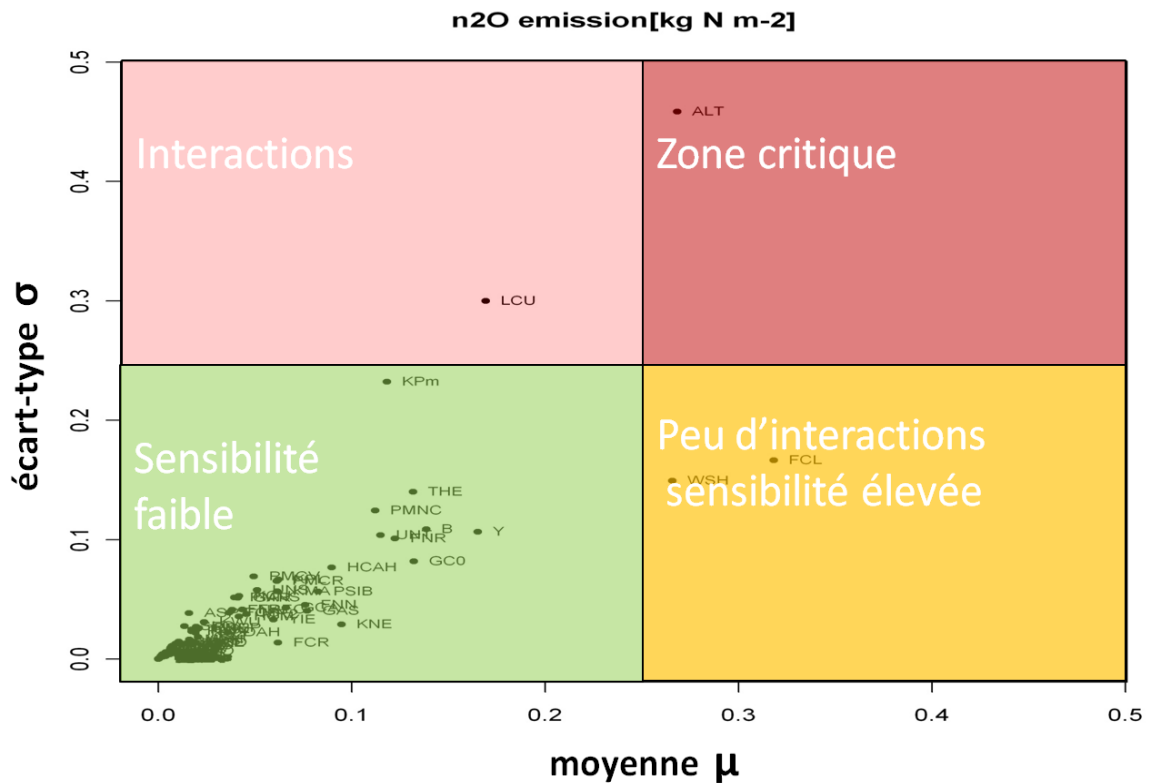


Figure 13. Exemple de représentation graphique (diagramme de Morris) de l'analyse de sensibilité des paramètres d'un modèle par rapport à une variable de sortie (émission de N_2O) selon deux indicateurs: la moyenne et l'écart-type de l'indice de sensibilité.

2.4.2. Autres méthodes d'analyse de sensibilité

La méthode de Cotter

Cette méthode [Cotter, 1979] permet d'ordonner les paramètres selon l'importance en utilisant les bornes de variation des paramètres d'entrée. Elle fait partie des méthodes utilisant le plan d'expérience OAT et fait varier un seul facteur à la fois de son minimum au maximum (et inversement) pendant que les autres facteurs restent tous au même niveau. Un indice de sensibilité est alors calculé à partir des niveaux haut et bas d'un paramètre d'entrée permettant de quantifier l'influence de ce facteur.

Son coût de calcul ($2n+2$ pour n facteurs) reste raisonnable même pour un nombre relativement élevé de facteurs d'entrée. Toutefois, cette méthode manque de précision et l'effet important d'un facteur peut ne pas être détecté. En plus, elle n'est pas directement implémentée sous R.

Méthodes basées sur la variance

Des méthodes comme Sobol' [Sobol', 1993], FAST (Fourier Amplitude Sensitivity Test) [Cukier et al., 1978] et Extended-FAST [Saltelli et al., 1999] utilisent une décomposition de la variance de la variable de sortie suivant l'influence des paramètres d'entrée. La méthode de Sobol' utilise des variances et espérances conditionnelles alors que la méthode Extended-FAST utilise des outils mathématiques plus complexes comme les transformées de Fourier permettant de transformer l'intégrale multidimensionnelle servant à calculer la variance de la variable de sortie en une intégrale unidimensionnelle. Ces méthodes basées sur la décomposition de la variance permettent de calculer les effets d'influences et d'interactions de tous les paramètres d'entrée sur la réponse. Ainsi une classification des facteurs sensibles est obtenue en se basant sur les différents effets.

La limite de ce genre de méthodes est qu'elles sont coûteuses en terme de temps de calcul et en terme de complexité (Figure 12) à fur et à mesure que le nombre de paramètres étudiés augmente.

Plans factoriels fractionnaires

Cette méthode utilise les bornes de variations des paramètres d'entrée en considérant un niveau max et un niveau min, ce sont donc des plans d'expériences de 2^n où n est le nombre de paramètres d'entrée à analyser [Kobilinsky, 1990]. Cela permet de calculer les effets et les interactions entre les paramètres à partir des sorties du modèle.

Cette méthode permet le calcul des effets principaux et de tous les effets d'interaction des paramètres. Mais, face à un nombre important de paramètres, l'inconvénient de cette méthode est que le nombre d'exécutions peut rapidement devenir conséquent et l'interprétation des résultats devient de plus en plus complexe. En outre, il n'y a pas une implémentation de cette méthode sous R.

Tableau 1. Récapitulatif comparant différentes méthodes d'analyse de sensibilité

Méthodes	Nombre de paramètres nécessaires	Nombre de simulations du modèle (pour n paramètres)	Implémentation sous R
Morris	Grand	$(n+1) \times r$ où r est le nombre de trajectoires	Oui
Cotter	Grand	$2 \times n + 2$	Non
FAST / Extended FAST	Petit	$n \times k$ où k est la taille d'échantillon	Oui
Sobol'	Petit	$(n+2) \times k$ où k est la taille d'échantillon	Oui
Plans factoriels fractionnaires	Grand	2^n	Non

2.5. Ressources informatiques

2.5.1. Ressources matérielles

La recherche représente 15% de l'utilisation des ressources matérielles informatiques, le reste étant partagé sur d'autres secteurs comme la finance, la géophysique et les services de l'information. En effet, étudier le changement climatique, réduire les incertitudes, simuler des événements extrêmes et étudier les impacts sur les écosystèmes, nécessitent une très grande puissance de calcul ou ce qu'on appelle du calcul intensif.

Le calcul intensif

Le calcul intensif est un domaine rassemblant les outils matériels et logiciels nécessaires à l'exécution d'applications et de techniques complexes [Turton et Openshaw, 1998 ; Fiore et Aloisio, 2011]. Les supercalculateurs, les clusters, les grilles de calcul et les processeurs graphiques sont des infrastructures de calcul intensif. Les supercalculateurs sont des ordinateurs dédiés à l'exécution de tâches nécessitant du calcul intensif et peuvent exécuter une même tâche simultanément. Les supercalculateurs sont dotés de plusieurs processeurs travaillant en parallèle [Hoffman et Traub, 1989]. En 2013, Tianhe-2 est le supercalculateur le plus puissant au monde avec plus de 3 100 000 processeurs. Cependant

pour utiliser correctement ces énormes machines il faut savoir gérer les communications entre processus et l'architecture complexe des supercalculateurs.

On distingue aussi les « clusters » ou fermes de calcul qui sont des ensembles d'unités de calcul regroupées physiquement en un même lieu afin d'additionner les puissances de calcul et de stockage distribué. C'est ce genre de machines qui sera utilisé dans cette thèse pour lancer les différentes tâches (ou « jobs ») de calibration Bayésienne et d'analyse de sensibilité. L'accès au cluster par l'utilisateur se fait via un serveur nommé passerelle. Il s'agit d'un serveur d'accès depuis l'extérieur, on peut être éloigné géographiquement tel est mon cas, puisque je suis principalement basé à l'INRA et je lance mes jobs à distance sur le cluster de l'Université Blaise Pascal de Clermont-Ferrand à travers une passerelle.

Ensuite, face aux demandes massives de calcul et de stockage telles que demandées par des grands projets internationaux comme le « Large Hadron Collider » (LHC, <http://home.web.cern.ch/about/accelerators/large-hadron-collider>) ou le télescope « Large Scale Spacio-Telescope » (LSST) il existe ce qu'on appelle les grilles de calcul « grid computing ». Cela consiste à mettre en commun et en réseau les ressources de nombreux centres de recherche. On passe alors d'une puissance de calcul de l'ordre de grandeur de quelques centaines de processeurs pour des cluster à des dizaines et des centaines de milliers de processeurs dans le cadre d'une grille de calcul [Foster et Kesselman, 1999]. Plus récemment, les processeurs graphiques GPU « Graphical Processing Unit » commencent à être utilisés pour le traitement des données à la place des processeurs CPU « Central Processor Unit » classiques. Néanmoins, malgré leur efficacité en termes de puissance de calcul, l'utilisation des GPU nécessite un travail d'adaptation des algorithmes à ces nouvelles architectures de calcul.

2.5.1. Ressources logicielles

Pour optimiser les calculs, avoir des machines de calcul intensif ne suffit pas, il faut bien gérer la répartition de ses jobs sur ces machines et leurs nœuds. La parallélisation des simulations permet alors de réduire le temps de calcul en le distribuant sur différentes machines ou processeurs. Dans notre cas, nous utilisons le langage Perl qui est un langage combinant des fonctionnalités du langage C avec d'autres langages de scripts, et à travers le code Perl nous lançons nos jobs en faisant appel à des commandes systèmes (Linux).

En ce qui concerne le code de la calibration Bayésienne, il a été écrit en langage R (<http://www.r-project.org>). Dans la littérature, certains livres traitent le calcul Bayésien à l'aide de R comme [Albert, 2009]. R est à la fois un outil statistique et un langage de programmation. L'efficacité de R réside dans les nombreuses fonctions et bibliothèques destinées aux analyses statistiques. En outre le côté analytique, R est doté de plusieurs bibliothèques graphiques permettant la représentation graphique des résultats sous différentes formes voir même en 3D.

En plus, R est flexible car il permet de stocker les résultats dans des « objets » qu'on peut y accéder à tout moment. Il est aussi un outil portable existant sous de nombreuses plateformes : Windows, MacOS, Linux... Son syntaxe est simple, il est bien documenté sur internet et surtout gratuit et distribué librement contrairement à d'autres outils comme MATLAB.

Tout cela, nous a permis le couplage entre R et PaSim où le code de la calibration Bayésienne écrit en R appelle l'exécutable du modèle et les simulations sont lancées à l'aide de Perl d'une façon parallélisée sur le cluster qui lui tourne sous un système Linux. Aussi, nous nous sommes familiarisés avec le langage Fortran puisque PaSim est écrit en Fortran et cela pour l'extraction des équations et la documentation du modèle.

2.6.Conclusion

Nous avons fait un tour d'horizon dans la littérature concernant les termes clés de cette étude, notamment la calibration Bayésienne des paramètres de modèles de simulation dans le but d'améliorer les résultats et réduire en même temps l'incertitude liée à ses paramètres et par conséquent aux sorties. Ensuite, nous avons présenté des méthodes d'analyse de sensibilité, une étape nécessaire pour la sélection des paramètres à calibrer. Enfin, nous avons justifié et argumenté nos choix que ce soit pour la technique de calibration ou la méthode d'analyse de sensibilité qui seront utilisées.

Le chapitre suivant décrit la démarche suivie pour l'application de la calibration Bayésienne au modèle PaSim en termes de choix de sites, données et paramètres. Ensuite, nous présentons et discutons les résultats obtenus à l'issue de cette étude.

Chapitre 3 : Calibration Bayésienne de PaSim en conditions de sécheresse

3.1. Introduction

En dépit de la pertinence de la calibration Bayésienne, la technique a été rarement appliquée au domaine de recherche sur les relations entre prairie et climat, en particulier en utilisant les modèles qui simulent les prestations productives et environnementales des prairies sous conditions climatiques modifiées comme, par exemple, les vagues de chaleur et les sécheresses estivales, ou des crues brutales plus nombreuses et des tempêtes [Field et al., 2012].

Dans ce chapitre, nous proposons des éléments de réponse à la problématique de la calibration de modèles complexes que nous avons exposée dans les chapitres précédents. Tout d'abord, nous cherchons à montrer comment l'approche Bayésienne fournit une méthode logique pour la calibration des modèles de prairies, en utilisant le modèle de simulation PaSim comme un modèle de référence combinant, de façon mécaniste, les caractéristiques biophysiques de l'écosystème prairial (y compris la présence éventuelle d'animaux au pâturage).

La principale question est de savoir jusqu'à quel point la calibration Bayésienne peut améliorer les estimations du modèle et réduire leurs incertitudes lorsque le modèle est utilisé pour représenter l'impact de divers phénomènes associés aux changements climatiques. Pour répondre à cette question, nous avons appliqué la calibration Bayésienne pour simuler des écosystèmes prairiaux typiques en Europe centrale incluant des traitements sous conditions de sécheresse induite. En effet, les systèmes basés sur l'exploitation de l'herbe sont directement dépendant des effets de la sécheresse qui détermine à la fois la croissance de l'herbe au pâturage et la constitution des stocks fourragers d'où l'importance d'étudier ce phénomène.

Ce travail a fait l'objet de deux publications avec deux présentations dans deux manifestations scientifiques internationales : la conférence 2012 Probabilistic Safety Assessment & Management [Ben Touhami et al., 2012a] et le douzième congrès de l'European Society for Agronomy [Ben Touhami et al., 2012b].

La première partie de ce chapitre présente, d'abord, la description des sites et des variables de sortie utilisées dans la calibration et, ensuite, les paramètres du modèle à calibrer et le fonctionnement de la calibration Bayésienne avec PaSim.

Dans la deuxième partie, nous montrons et analysons les résultats de la calibration Bayésienne des trois sites avec et sans conditions de sécheresse, et nous discutons l'impact de cette calibration sur les résultats du modèle.

3.2. Matériels et méthodes

3.2.1. Sites et variables de sortie utilisés

Dans cette étude, nous avons considéré trois sites prairiaux (Tableau 2) impliquant un traitement sous conditions de sécheresse (réduction de précipitation) et pour lesquels nous avons les informations nécessaires pour faire fonctionner le modèle et des données disponibles pour comparer les résultats. Ces informations ont été fournies directement par les propriétaires des données suite à deux-quatre ans d'expérimentation. Dans les sections suivantes, nous allons présenter les sites utilisés, les types d'expérimentation, les variables de sortie et les paramètres du modèle considérés pour la calibration.

Tableau 2. Liste des sites et des variables de sortie utilisés pour la calibration

Sites (S_i)	Nombre d'années	Variables de sortie				
		Rendement (g MS m ⁻²)	Indice de surface foliaire (m ² m ⁻²)	Teneur en eau du sol (m ³ m ⁻³)	Température du sol (K)	Surface foliaire spécifique (m ² kg ⁻¹)
Stubai (Autriche)	2	X	X	X	X	X
Chamau (Suisse)	4	X	X	X	X	-
Früebüel (Suisse)	4	X	X	X	X	-

Le rendement de la prairie (Y) est la quantité de matière sèche (MS) des parties aérienne récoltée à chaque fauche, ce qui implique la sortie simultanée de carbone et azote d'une prairie fauchée ou pâturée et fauchée. L'indice de surface foliaire (LAI), calculé comme le

ratio de la surface totale supérieure des feuilles à la surface du sol sur laquelle la végétation se développe, est une mesure de la capacité d'utilisation de la lumière de la part de la plante, de base pour le calcul de la production photosynthétique, l'évapotranspiration et la croissance végétale. La surface foliaire spécifique (SLA) correspond à la surface de feuille réalisée par masse de matière sèche, positivement corrélée à la capacité photosynthétique et à la teneur en azote foliaire par unité de masse et négativement corrélée à la durée de vie foliaire. La connaissance biophysique de l'écosystème prairial est accompagnée par deux variables du sol, la teneur en eau et la température (SWC and T_{SOIL}), dont la distribution tout le long du profil du sol dépend du bilan d'énergie à la surface du sol. L'estimation des valeurs des variables du sol de la part de PaSim se font à partir d'un schéma de division du profil en six couches d'épaisseur différent.

Site Autrichien : Stubai

Stubai est une prairie de montagne typique des Alpes centrales orientales, localisée en Autriche (47° 05' Nord, 11° 11' Est) à 1850 m d'altitude. Elle est fauchée (coupée) une fois par an, légèrement pâturée en fin d'été, et fertilisée avec du fumier / lisier environ tous les trois ans [Schmitt et al., 2010]. Les données utilisées sont collectées et disponibles pour deux ans sur la période 2008-2009.

Le dispositif expérimental comprend deux types de traitements : une parcelle témoin dans laquelle aucun des facteurs n'a été manipulé, donc sans traitement spécifique (Control), et une parcelle avec un traitement de réduction de précipitations en utilisant des rideaux de tailles appropriées pour la protection de la parcelle contre la pluie (Figure 14).



Figure 14. Expérimentation (réduction de précipitation à l'aide des rideaux) au site Autrichien de Stubai

Nous avons utilisé un ensemble de cinq variables de sorties ($D_1..D_5$) mesurées et disponibles (Tableau 3) qui sont le rendement en matière sèche (yield, g m^{-2}), la teneur en eau du sol (soil moisture, $\text{m}^3 \text{m}^{-3}$), la température du sol (soil temperature K), l'indice de surface foliaire (leaf area index $\text{m}^2 \text{m}^{-2}$) et la surface spécifique des feuilles (specific leaf area $\text{m}^2 \text{kg}^{-1}$), sachant que la teneur en eau du sol et la température du sol ont été mesurées à trois profondeurs (0,05 m, 0,10 m et 0,20 m) et moyennées à un pas de temps de cinq jours. Ces données sont utilisées pour la mise à jour de la distribution a priori des paramètres du modèle à travers le calcul de la vraisemblance, et utilisées après la fin de la calibration pour la comparaison avec les simulations du modèle et l'analyse des résultats.

Tableau 3. Description des variables de sortie (D_i , $i=1, \dots, 5$) utilisées pour la calibration de Stubai

Variable (D_i)	Unité	Dates de mesure	
		Control	Traitement
D_1 Leaf area index	$\text{m}^2 \text{m}^{-2}$	15.08.2008, 29.07.2009	29.07.2009
D_2 Yield	g m^{-2}	15.08.2008, 29.07.2009	29.07.2009
D_3 Specific leaf area	$\text{m}^2 \text{kg}^{-1}$	15.08.2008, 29.07.2009	29.07.2009
D_4 Soil moisture	$\text{m}^3 \text{m}^{-3}$	2008-2009 (mesures journalière)	2008-2009 (mesures journalière)
D_5 Soil temperature	K	2008-2009 (mesures journalière)	2008-2009 (mesures journalière)

Sites Suisse : Chamau & Frübüel

Chamau et Frübüel sont deux prairies expérimentales Suisse (Figure 15), la première est une prairie artificielle située dans une vallée à Chamau (47° 12' Nord, 8° 24' Est) à 393 m d'altitude et la deuxième sur un plateau préalpin à Frübüel (47° 06' Nord, 8° 32' Est) à 982 m d'altitude.

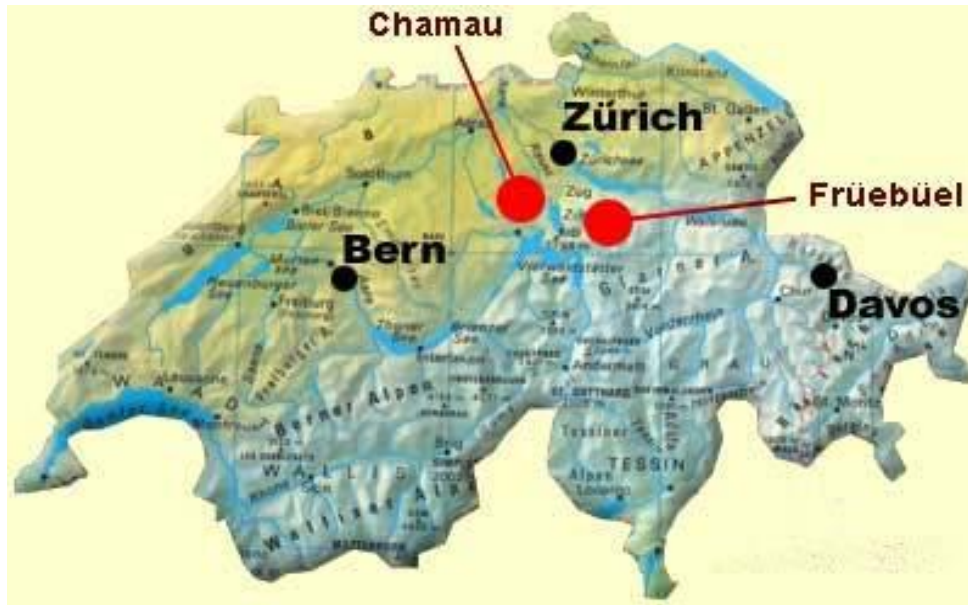


Figure 15. Localisation des deux sites expérimentaux Suisse de Chamau et Frübüel

Le site de Chamau est géré intensivement avec six coupes par année, tandis que celui de Frübüel est géré de façon modérément intensive avec généralement une à deux coupes par an et du pâturage en automne [Zeeman et al., 2010]. Sur les deux sites, des manipulations expérimentales de précipitations ont été effectuées, incluant des parcelles témoins (pas de manipulation) et des parcelles de traitement où les précipitations ont été réduites en utilisant des rideaux (Figure 16 et Figure 17). Quatre variables de sortie ($D_1..D_4$), mesurées pendant quatre ans entre (2006-2009), étaient considérées pour la calibration Bayésienne pour Chamau et Frübüel. Ces variables de sortie sont l'indice de surface foliaire (leaf area index $m^2 m^{-2}$), le rendement en matière sèche (yield/harvested biomass $g m^{-2}$), la teneur en eau du sol (soil moisture $m^3 m^{-3}$) et la température du sol (soil temperature K) mesurées à trois profondeurs (0,05 m, 0,15 m et 0,30 m) avec un pas de temps journalier.



Figure 16. À gauche : expérimentation (réduction de précipitation à l'aide des rideaux), à droite : tour de mesure des flux au site Suisse de Chamau



Figure 17. À gauche : expérimentation (réduction de précipitation à l'aide des rideaux), à droite : des capteurs de mesures pour les variables liées au sol au site Suisse de Frübüel

3.2.2. Paramètres calibrés

PaSim est un modèle complexe qui compte un nombre important de paramètres. Dans cette étude, un ensemble de 17 paramètres liés à la végétation, au sol et à la gestion, et spécifiques aux sites ont fait l'objet de la calibration Bayésienne. Ce choix des paramètres à calibrer s'est basé sur des études antérieures [Graux, 2011] en tenant compte des informations a priori disponibles. Le Tableau 4 donne, pour chaque paramètre, la description et l'information a priori disponible à partir des expertises précédentes et de la littérature, à la fois des sources publiées [Johnson et Thornley, 1985; Thornley et Verberne, 1989; Riedo et al, 1998, 1999; Graux, 2011] et des sources non publiées. L'ensemble des paramètres utilisés se compose de cinq paramètres de végétation ($\theta_1.. \theta_5$), 10 paramètres

de sol ($\theta_6.. \theta_{15}$) considérés comme spécifiques au site, et deux paramètres liés aux pratiques de gestion du site ($\theta_{16}.. \theta_{17}$). Les paramètres ($\theta_1.. \theta_4$) sont, en effet, des paramètres composés de 26 paramètres liés à la végétation et qui caractérisent le type fonctionnel de prairie (types A, B, C et D en fonctions des facteurs végétaux). Le paramétrage de PaSim pour la simulation des types fonctionnels de prairie a été défini par [Cruz et al., 2002]. En l'absence d'informations sur la forme des fonctions de distribution de probabilité (FDP) des paramètres, une distribution uniforme a été considérée en tant que distribution a priori avec des bornes fixées au minimum (θ_i min) et au maximum (θ_i max) du rang de valeurs des paramètres connues a priori (Tableau 4).

Nous avons lancé des chaînes de Markov de longueur 50.000-100.000 pour chacun des trois sites et pour chaque type de traitement (Control et Traitement) en utilisant une FDP normale multivariée pour générer les vecteurs des valeurs candidats de chaque paramètre. La matrice de covariance de cette distribution gaussienne a été construite de telle sorte que les chaînes de Markov explorent tout l'espace des paramètres au cours de la marche aléatoire. Pour chaque nouveau vecteur de paramètres θ_i nous calculons le logarithme de la vraisemblance L_i de chaque donnée D_i , qui est la probabilité des données étant donné les valeurs des paramètres (dont la formule a été présentée dans le chapitre précédent). En se basant sur la valeur de la vraisemblance on calcule le rapport de Metropolis dont dépend l'acceptation d'un paramètre candidat proposé ou son rejet, puis on passe à l'itération suivante dans la chaîne de Markov. On peut souligner ici, que le fait d'utiliser tout l'ensemble des données D_i au même temps ou par étapes revient au même, en effet si on a deux données D_1 et D_2 on peut les utiliser ensemble pour le calcul de la FDP du posterior qui sera de la forme $P(\theta/D_1, D_2)$, ou bien calculer premièrement $P(\theta/D_1)$ et ensuite l'utiliser comme prior pour D_2 . Dans les deux cas nous obtiendrons la même FDP finale du posterior $P(\theta/D_1, D_2)$, cela est l'un des avantages de la calibration Bayésienne. La corrélation entre les paramètres est prise en compte par la calibration Bayésienne du fait que la FDP des paramètres représente une distribution de probabilité de toutes les combinaisons possibles des valeurs des paramètres. En d'autres termes, la FDP obtenue est une distribution de probabilité jointe et non pas un ensemble de distributions indépendantes pour des paramètres individuels.

Tableau 4. Description des paramètres de PaSim (θ_i , $i=1, \dots, 17$) utilisés pour la calibration

Paramètre				Valeur par default	Distribution de probabilité a priori	
θ_i	Symbole	Description	Unité		θ_i min	θ_i max
θ_1	TypeA	Type fonctionnel* de prairie avec surface spécifique des feuilles élevée, haute digestibilité, courte durée de vie des feuilles et croissance reproductive prématuré	-	0.1	0.01	1
θ_2	TypeB	Type fonctionnel de prairie avec surface spécifique des feuilles moyenne, haute digestibilité, longue durée de vie des feuilles et croissance reproductive moyenne	-	0.1	0.01	1
θ_3	TypeC	Type fonctionnel de prairie avec faible surface spécifique des feuilles, digestibilité moyenne, longue durée de vie des feuilles et croissance reproductive moyenne	-	0.1	0.01	1
θ_4	TypeD	Type fonctionnel de prairie avec faible surface spécifique des feuilles, faible digestibilité, très longue durée de vie des feuilles et croissance reproductive tardive	-	0.95	0.01	1
θ_5	slaMax	Surface foliaire spécifique maximale	m ² kg ⁻¹	28	21	35
θ_6	zsb	Profondeur du sol	mm	700	500	900
θ_7	thetassatb	Teneur en eau du sol à la saturation	m ³ m ⁻³	0.6	0.4	0.7
θ_8	thetasfc	Fraction de la teneur en eau du sol à saturation à la capacité au champ	-	0.56	0.5	0.9
θ_9	thetasfpwp	Fraction de la teneur en eau du sol à saturation au point de flétrissement	-	0.4	0.2	0.49
θ_{10}	psieb	Valeur seuil du potentiel auquel les plus grands pores remplis d'eau commencent à se vider	mm	-228	-20	-400
θ_{11}	bb	Paramètre b de Campbell	-	7.9	2	14
θ_{12}	ksatb	Conductivité hydraulique en milieu saturé	mm d ⁻¹	165	20	400
θ_{13}	b_root_profile	Paramètre de la distribution de la matière sèche des racines par rapport aux différentes couches de sol	-	10	4	20
θ_{14}	thetasbspring	Teneur en eau de la couche limite du sol au printemps	m ³ m ⁻³	0.35	0.15	0.3
θ_{15}	thetasbautomn	Teneur en eau de la couche limite du sol en automne	m ³ m ⁻³	0.35	0.2	0.39
θ_{16}	wshtotcutinit	Matière sèche de la partie aérienne après fauche	kg m ⁻²	0.15	0.09	0.25
θ_{17}	lcutinit	Indice de surface foliaire après fauche	m ² m ⁻²	0.4	0.3	1.7

* $F_A = \text{TypeA} / (\text{TypeA} + \text{TypeB} + \text{TypeC} + \text{TypeD})$

Environ 5% à 10% du nombre total d'itérations au début de la chaîne ne sont pas prises en compte car considérées comme non représentatives des chaînes, cela étant une étape de stabilisation du modèle (étape de « burn-in ») [van Oijen et al., 2005].

La Figure 18 représente le fonctionnement de la calibration Bayésienne avec PaSim. D'abord, il y a le prior des paramètres à calibrer θ_i à partir duquel on peut simuler le prior des variables de sortie D_i (avant calibration). Ensuite, le modèle a besoin des entrées liées à chaque site S_i (gestion, météo et conditions du sol) et des données observées correspondantes aux variables de sortie considérées pour la calibration. Enfin, après calibration une nouvelle FDP des paramètres a posteriori est obtenue et par conséquent un nouveau paramétrage posterior (nouvelle valeur pour chaque paramètre obtenue par maximum de vraisemblance à partir de la FDP du posterior), ce qui permet de simuler le posterior des variables de sorties (après calibration). La comparaison des données avant calibration (prior), des données observées et des données après calibration (posterior) nous permettra de déduire l'impact de la calibration Bayésienne sur les résultats du modèle.

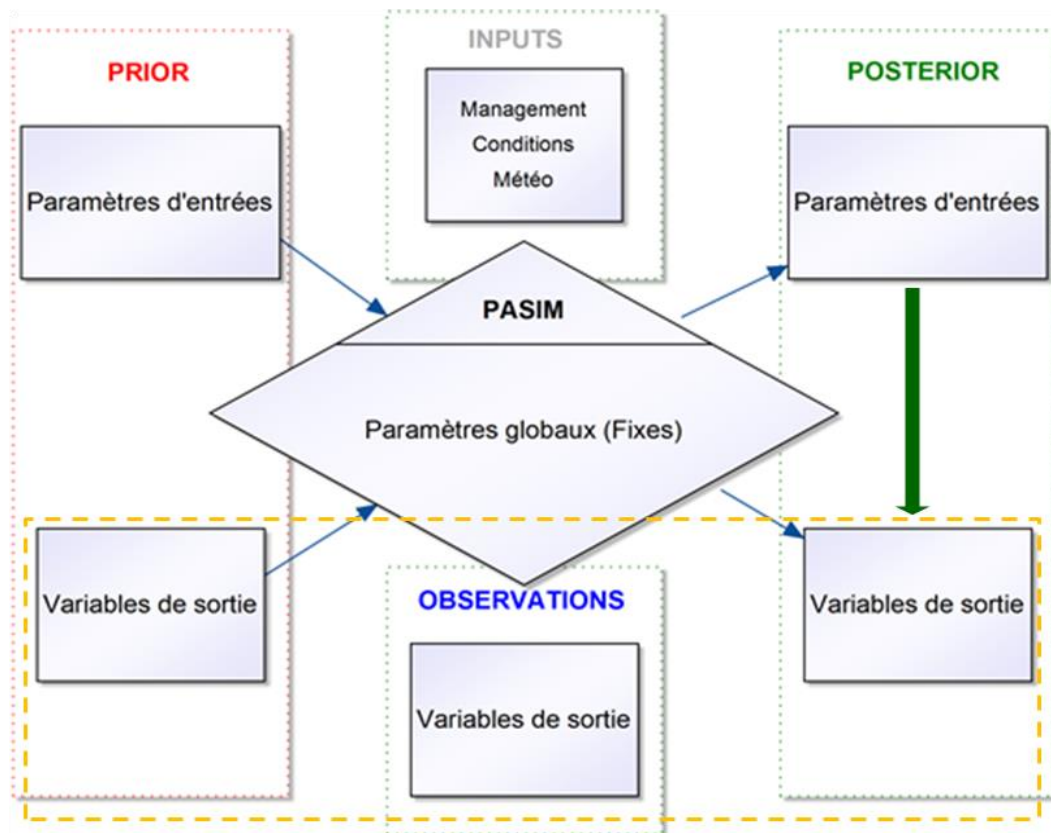


Figure 18. Fonctionnement de la calibration Bayésienne avec PaSim

Le code de la calibration Bayésienne, la préparation des fichiers de données et l'analyse des résultats étaient tous codés en R. L'outil R⁵ nous a permis d'appliquer l'algorithme Metropolis des méthodes MCMC en effectuant les tirages des échantillons des paramètres dans les distributions (voir *Annexe I : Implémentation du code principal de la calibration Bayésienne sous R*). La calibration est testée localement, ensuite elle est lancée sur un cluster qui se trouve à l'Université Blaise Pascal de Clermont-Ferrand dans les locaux du LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes). Puisque ma thèse est en co-encadrement INRA/LIMOS⁶, j'ai pu donc bénéficier de ces machines à calcul haute performance qui comportent 22 nœuds biprocesseurs « AMD Opteron Quadcore », avec 12 Go de RAM⁷ par machine, sous CentOS 5.5. Ce cluster est utilisé pour le calcul intensif parallèle ou séquentiel à 64 bits. La connexion au cluster se fait à travers un serveur passerelle, les jobs sont ensuite lancés à travers du code Perl en faisant appel à des commandes systèmes (Linux) tel que « qsub » qui permet de soumettre des jobs. Les options de la commande « qsub » permettent la spécification des attributs pour déterminer le comportement de la tâche soumise en termes de nœuds et de nombre de processeurs utilisés, de priorité des jobs et d'autres options permettant d'optimiser le temps de calcul.

3.3. Résultats et discussion

La calibration Bayésienne du site autrichien de Stubai a duré 25 jours pour les deux types de traitements en parallèle, tandis que la durée moyenne de la calibration pour les deux sites suisses de Chamau et Frübüel a été de 37 jours pour les deux types de traitements en parallèle (Figure 19). Les ressources matérielles et logicielles, malgré leur robustesse, peuvent encourir des bugs, des pannes ou des coupures électriques pour le cluster par exemple, ce qui est arrivé de nombreuses fois durant la thèse. Pour cela, nous avons opté pour une solution qui consiste à l'enregistrement de l'état courant et des résultats toutes les 20 itérations de la chaîne ce qui permet de récupérer le travail même en cas de problème. Pour les chaînes longues ($\geq 30\,000$ itérations), cela représente une manière efficace d'optimiser l'obtention des résultats.

⁵ <http://www.r-project.org>

⁶ <http://limos.isima.fr>

⁷ La mémoire vive (*Random Access Memory*)

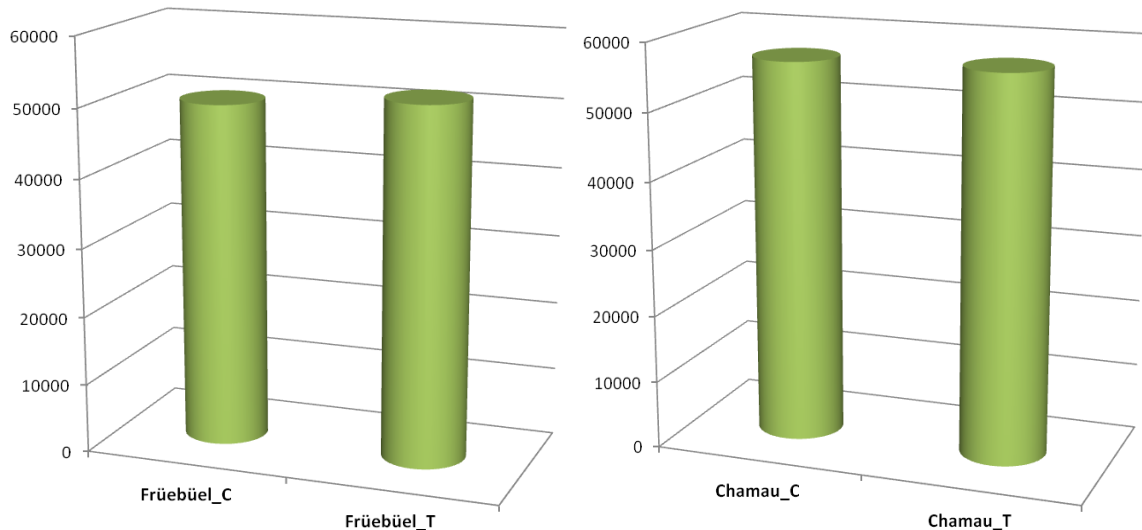


Figure 19. À gauche : le nombre d'itérations MCMC obtenues en 35 jours pour la calibration de Fruebüel avec (T) et sans traitement (C) ; à droite : le nombre d'itérations MCMC obtenues en 39 jours pour la calibration de Chamau avec (T) et sans traitement (C)

3.3.1. L'a posteriori des paramètres

Le taux d'acceptation moyen des paramètres candidats dans la chaîne MCMC a été de 40% pour les trois sites de cette étude. Il y a eu certains paramètres dont la convergence n'a pas été optimale, mais globalement la convergence a été atteinte pour la majorité des paramètres. Les Figure 20 et Figure 21 montrent quelques exemples de densités de probabilité des paramètres obtenues après calibration Bayésienne. En effet, après calibration nous obtenons une distribution donc un ensemble de valeurs pour chaque paramètre (selon la longueur de la chaîne MCMC). La nouvelle valeur retenue pour un paramètre est la valeur du maximum de la FDP (Figure 20), c'est-à-dire la projection de la valeur maximale de la densité d'un paramètre (axe des Y) sur le rang des valeurs possibles du paramètre (axe des X).

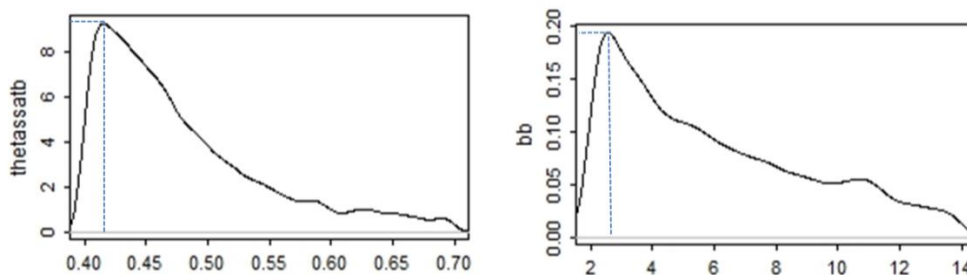


Figure 20. Densités de probabilité des paramètres *thetasatb* (teneur en eau du sol à la saturation) obtenue à Chamau et *bb* (Campbell's *b* parameter) obtenue à Fruebüel, et leurs valeurs correspondant au maximum de la FDP

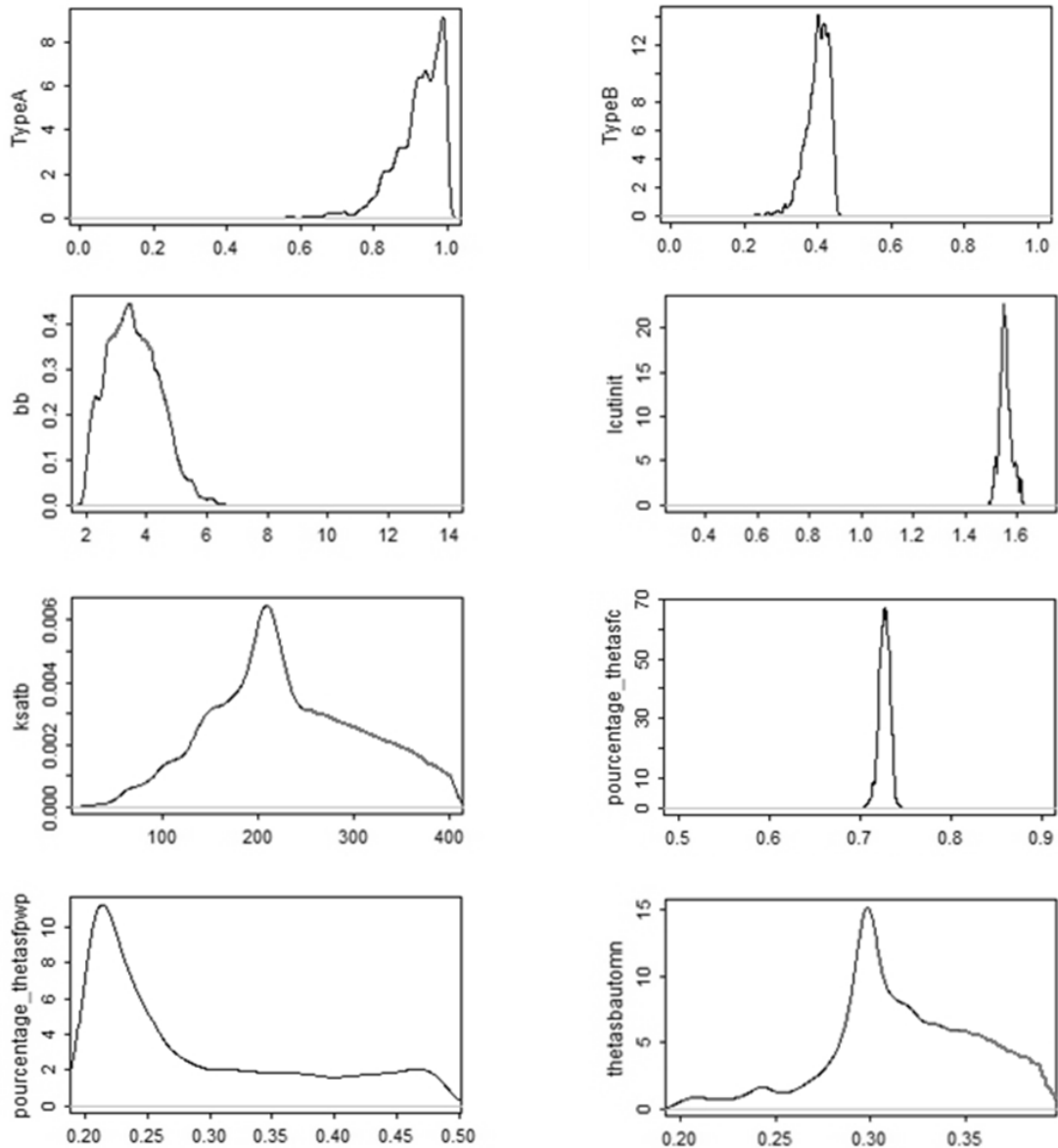


Figure 21. Exemple de densité de probabilité des paramètres obtenue après calibration Bayésienne du site autrichien de Stubai

La dernière étape de la calibration Bayésienne consiste donc à extraire les nouvelles valeurs pour les paramètres à partir de la nouvelle distribution a posteriori. Le modèle est ensuite utilisé avec ces nouvelles valeurs pour obtenir les simulations a posteriori (après calibration). Le Tableau 5 dresse la liste des nouveaux paramétrages obtenus après calibration Bayésienne site par site et pour chaque type de traitement. Ce nouveau paramétrage est à utiliser pour toute étude ultérieure impliquant l'un de ces trois sites.

Tableau 5. Nouveau paramétrage de PaSim pour trois sites prairiaux incluant des conditions de sécheresse (Traitement), obtenue après calibration Bayésienne (valeurs correspondant au maximum de la FDP)

Paramètre	Stubai		Chamau		Früebüel	
	Control	Traitement	Control	Traitement	Control	Traitement
TypeA	0.988	0.963	0.924	0.333	0.940	0.150
TypeB	0.401	0.734	0.794	0.850	0.917	0.707
TypeC	0.011	0.018	0.397	0.782	0.120	0.803
TypeD	0.011	0.012	0.916	0.825	0.181	0.234
zsb	899.641	500.057	1095.715	1445.365	1223.942	1438.925
thetassatb	0.504	0.400	0.415	0.417	0.404	0.404
thetasfc	0.727	0.897	0.785	0.635	0.663	0.615
thetasfpwp	0.231	0.481	0.409	0.470	0.465	0.469
psieb	-146.393	-27.888	-57.375	-352.919	-253.415	-371.122
bb	3.433	2.129	13.152	2.638	2.552	2.558
ksatb	209.420	45.954	75.647	262.636	364.571	278.088
b_root_profile	8.455	19.937	18.259	10.181	8.655	5.160
wshtotcutinit	0.090	0.092	0.119	0.178	0.125	0.103
lcutinit	1.549	1.697	1.589	1.482	1.536	1.589
thetasbspring	0.212	0.292	0.214	0.199	0.268	0.163
thetasbautomn	0.298	0.381	0.215	0.362	0.366	0.231
slaMax	34.998	34.995	25.829	27.051	22.277	22.854

L'un des points forts de la calibration Bayésienne est qu'elle fournit une distribution a posteriori pour chaque paramètre et non pas seulement une valeur unique. Cette FDP nous permet de mieux étudier et de quantifier l'incertitude liée aux paramètres du modèle. Pour cela nous avons calculé pour chaque paramètre la moyenne et le coefficient de variation qui est le rapport de l'écart-type à la moyenne arithmétique ($CV = \frac{\sigma}{\mu}$). Plus la valeur du coefficient de variation est faible, plus l'estimation est précise et donc moins incertaine.

Le Tableau 6 montre un exemple des valeurs de la moyenne et du CV des paramètres calculés pour le prior et le posterior, obtenues pour le site autrichien de Stubai sans traitement. Pour calculer la moyenne et le CV, nous avons considéré un échantillon de taille 1000 à partir de la distribution a priori et de celle a posteriori. À l'exception d'un paramètre qui est le *psieb* (Air-entry water potential), des faibles valeurs du CV pour la distribution a posteriori des paramètres montrent que l'information apportée sur les paramètres est plus accrue et moins incertaine que celle de la distribution a priori. Tel est le cas par exemple pour le CV de la profondeur du sol (*zsb*) qui était presque totalement réduit, la teneur en eau du sol à la saturation (*thetassatb*), la surface spécifique foliaire maximale (*slaMax*) et la matière sèche de la partie aérienne après fauche (*wshtotcutinit*).

Enfin, à travers le changement du CV nous pouvons constater que la réduction d'incertitude globale du posterior est de ~66% par rapport au prior pour les paramètres étudiés.

Tableau 6. Exemple des distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue au site autrichien de Stubai

Paramètre du modèle	Prior		Posterior		% changement CV
	Moyenne	CV	Moyenne	CV	
TypeA	0,18	0,62	0,92	0,07	-88,68
TypeB	0,10	1,60	0,40	0,07	-95,31
TypeC	0,10	1,58	0,01	0,30	-81,11
zsb	697	0,10	897	~0,00	-96,65
thetassatb	0,60	0,08	0,51	~0,00	-95,68
thetasfc	0,56	0,12	0,73	0,01	-93,54
thetasfpwp	0,40	0,12	0,23	0,10	-12,77
psieb	-231	0,27	-251	0,34	29,16
ksatb	165	0,38	245	0,34	-12,82
b_root_profile	9,88	0,27	13	0,24	-11,48
wshtotcutinit	0,15	0,19	0,09	~0,00	-98,49
lcutinit	0,40	0,58	1,55	0,02	-97,36
slaMax	28	0,09	35	~0,00	-99,56

Également, pour les deux sites suisses, la calibration Bayésienne a permis de réduire l'incertitude liée aux paramètres. À Frübüel, par exemple (Tableau 7), la réduction d'incertitude apportée par la distribution a posteriori a été de l'ordre de ~40% par rapport à la distribution a priori pour l'ensemble des paramètres.

Globalement, les valeurs du CV ont été réduites pour le posterior. Cela a été plus remarquable à travers les paramètres composés des types fonctionnels des prairies (types A et B) et la teneur en eau du sol à la saturation (*thetassatb*) avec une réduction d'incertitude avoisinant les 70%. Les paramètres de la surface spécifique foliaire maximale (*slaMax*) et de la fraction de la teneur en eau du sol à saturation à la capacité au champ (*thetasfc*) affichent, quant à eux, une réduction d'incertitude de plus de 50% comparée à la distribution du prior.

Tableau 7. Exemple des distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue au site suisse de Frübüel

Paramètre du modèle	Prior		Posterior		% changement CV
	Moyenne	CV	Moyenne	CV	
TypeA	0.16	0.73	0.50	0.56	-24.01
TypeB	0.08	1.92	0.52	0.54	-71.74
TypeC	0.08	1.94	0.47	0.59	-70.02
zsb	1197	0.30	1162	0.17	-41.55
thetassatb	0.55	0.30	0.42	0.08	-73.30
thetasfc	0.51	0.31	0.65	0.15	-52.17
thetasfpwp	0.36	0.32	0.36	0.22	-28.72
psieb	-216	-0.40	-201	-0.53	30.80
wshtotcutinit	0.14	0.34	0.18	0.26	-24.50
lcutinit	0.37	0.66	1.03	0.39	-40.86
thetasbspring	0.32	0.30	0.23	0.19	-35.99
thetasbautomn	0.32	0.30	0.29	0.18	-39.09
slaMax	25.89	0.30	27.82	0.14	-52.59

3.3.2. Analyse du posteriori des variables de sortie

Après la calibration Bayésienne site par site des trois sites de cette étude, nous avons utilisé le nouveau paramétrage de PaSim obtenu dans le Tableau 5 pour simuler le posterior des variables de sortie étudiées afin de les comparer avec le prior (avant calibration) et les observations pour pouvoir évaluer les résultats de la calibration Bayésienne site par site. Pour analyser graphiquement et analytiquement les incertitudes liées aux sorties, nous avons effectué 1000 simulations en utilisant une fois la distribution a priori et une fois la distribution a posteriori pour chaque site et pour chaque type de traitement. Ensuite, nous avons tracé les graphiques avec les bandes/barres d'incertitude selon la variable de sortie et son pas de temps tout en comparant entre prior, posterior et observation. Dans ce qui suit nous allons présenter et analyser quelques exemples de résultats obtenus dans chaque site.

Stubai

La Figure 22 compare l'incertitude des estimations du prior (avant calibration) et du posterior (après calibration) de la teneur en eau du sol à 0.05 m de profondeur sous conditions de sécheresse (réduction de précipitation).

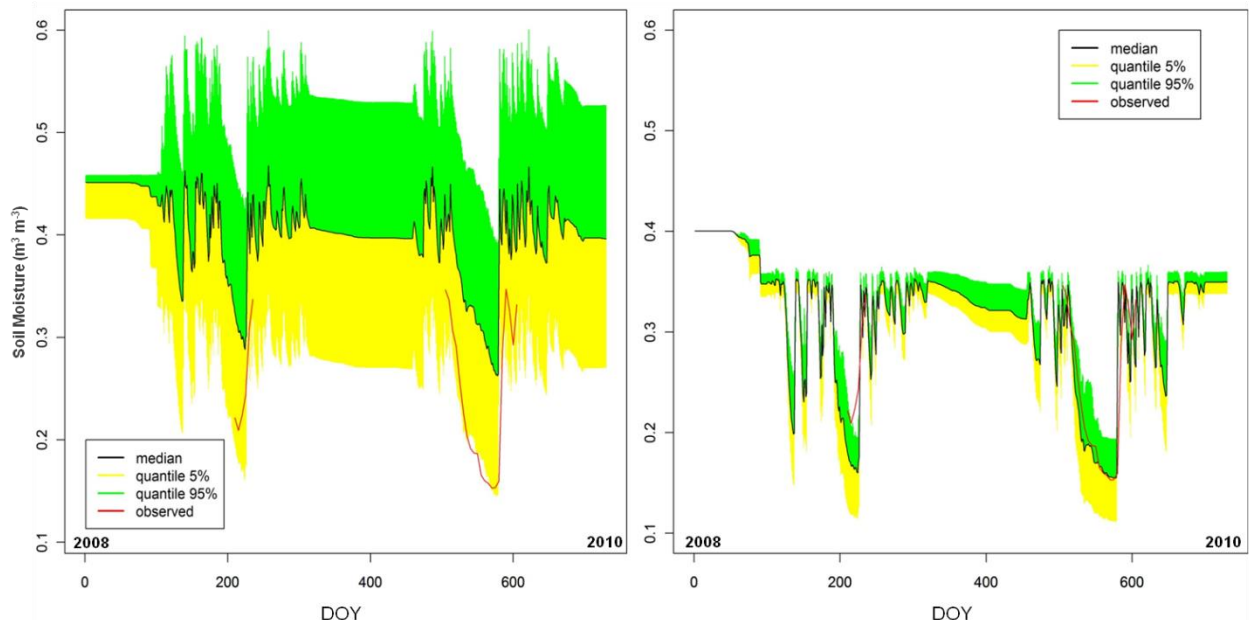


Figure 22. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.05 m de profondeur avec réduction de précipitation

Malgré le manque des données observées sur certaines périodes, la courbe du posterior s'approche nettement mieux des observations et surtout l'incertitude a été fortement réduite par rapport à l'incertitude du prior. Des résultats similaires ont été obtenus pour la même variable de sortie à une profondeur de sol de 0.10 m sans réduction de précipitation (Figure 23). Dans ce cas, on peut remarquer au début de la simulation certaines inexactitudes pour le prior et le posterior, dues à l'initialisation de la teneur en eau du sol et au temps nécessaire pour la stabilisation du modèle.

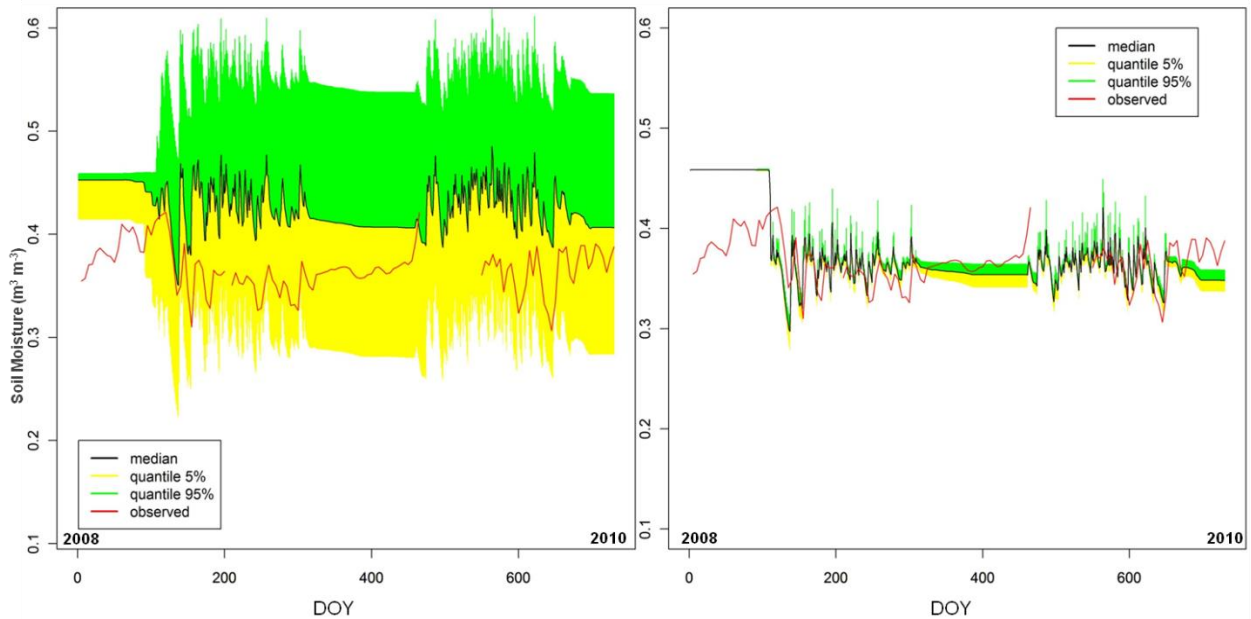


Figure 23. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.10 m de profondeur sans réduction de précipitation

En ce qui concerne la variable de sortie température du sol (Figure 24), le modèle fournit relativement un bon résultat avec le prior, et malgré ça la calibration Bayésienne a permis d'améliorer d'avantage les résultats de simulation en réduisant presque totalement l'incertitude et en correspondant mieux à la courbe des observations. L'analyse des simulations de la température du sol nous a permis, entre autre, de mettre le doigt sur l'une des limites de PaSim. En effet, quand la température du sol avoisine les 0 °C (273.15 K), le modèle montre des difficultés à la simuler. Et même après calibration Bayésienne, les simulations du modèle ne montrent pas d'améliorations sur ces périodes hivernales. Cela peut être donc une piste pour revoir les équations du modèle responsable du calcul de la température du sol pour la prise en compte des étapes neigeuses.

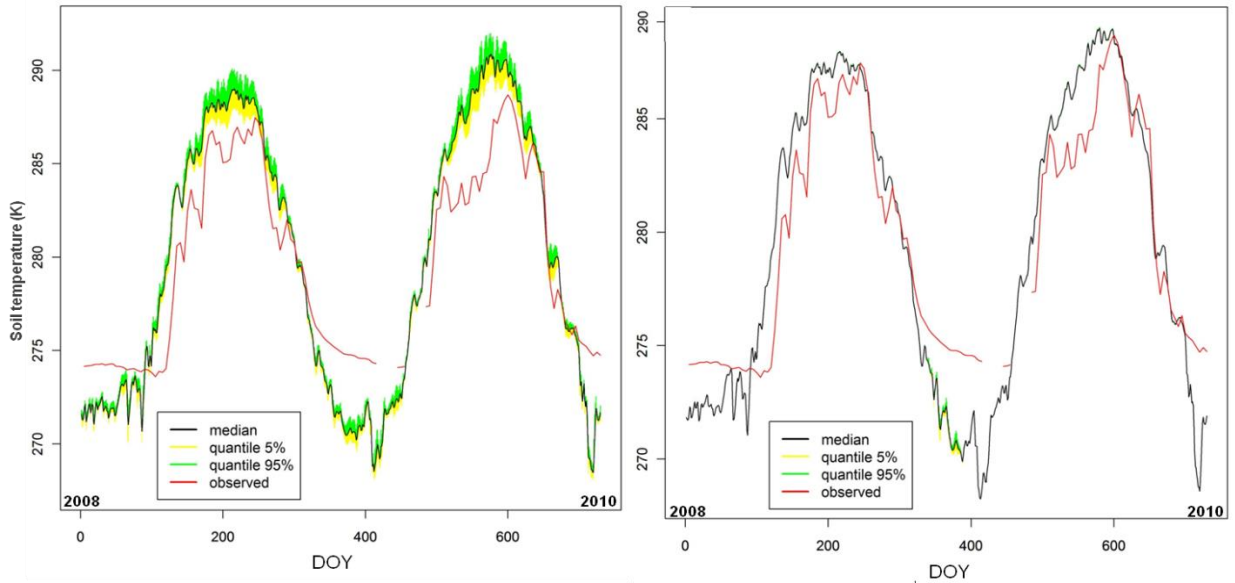


Figure 24. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la température du sol à 0.30 m de profondeur avec réduction de précipitation

Pour l'indice de surface foliaire (Figure 25), bien que nous n'ayons pas assez de données (seulement deux dates pour le contrôle et une pour le traitement), le contrôle et le traitement ont montré tous les deux une amélioration globale dans les estimations a posteriori même si l'incertitude (barres d'écart-type) n'a pas été réduite avec le traitement.

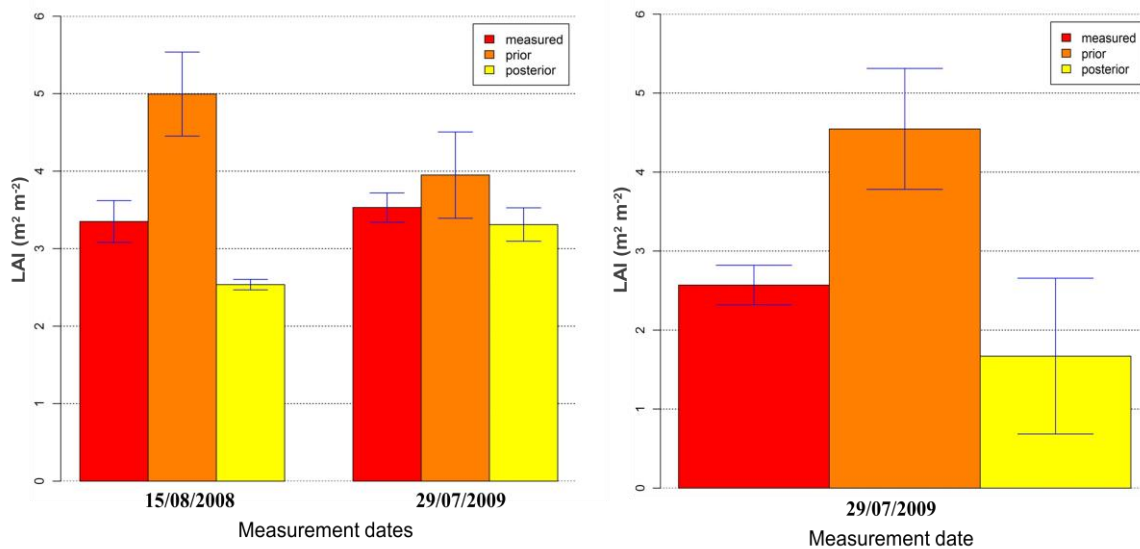


Figure 25. Analyse d'incertitude des estimations du prior et du posterior de l'indice des surfaces des feuilles sans (à gauche) et avec (à droite) réduction de précipitation

La Figure 26 présente les résultats obtenus pour le rendement et la surface spécifique foliaire. Ici aussi nous ne possédons pas assez de données, cependant on peut constater que pour le rendement la calibration Bayésienne a permis de réduire l'incertitude en 2008 mais avec moins de précision, alors qu'elle a permis d'améliorer la précision du résultat en 2009 sans réduction de l'incertitude par rapport au prior. Et pour la seule date disponible pour la surface foliaire spécifique, le posterior s'approche plus de l'observation et l'incertitude a également été réduite par rapport au prior.

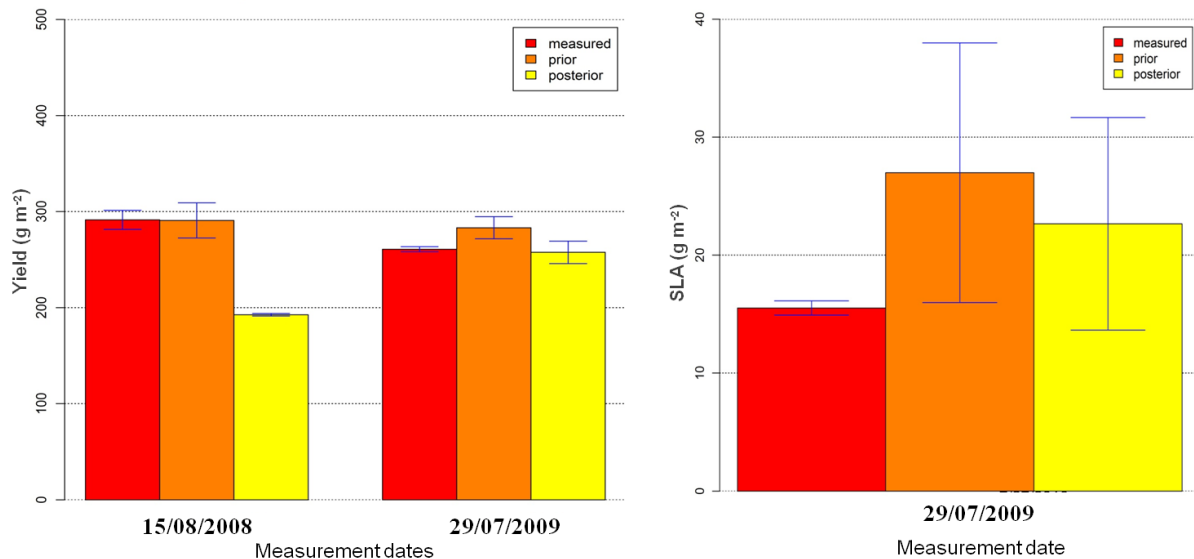


Figure 26. Analyse d'incertitude des estimations du prior et du posterior du rendement sans réduction de précipitation (à gauche) et de la surface spécifique foliaire avec réduction de précipitation (à droite)

Früebüel

Pareillement, à Früebüel nous avons obtenu une amélioration globale des estimations avec une réduction importante d'incertitude. La Figure 27 montre qu'avant calibration Bayésienne les simulations de la teneur en eau du sol à 0.30 m de profondeur avec réduction de précipitation sont loin de ce qui a été observé en réalité et avec une large incertitude. Après calibration, le posterior a nettement amélioré le résultat de la simulation en s'approchant aux observations et en réduisant l'incertitude et cela sous conditions de sécheresse. Toujours avec réduction de précipitation, la Figure 28 présente le résultat de la biomasse récolté où l'on peut constater que sur toutes les dates le posterior était meilleur que le prior tout en réduisant l'incertitude de ~30%.

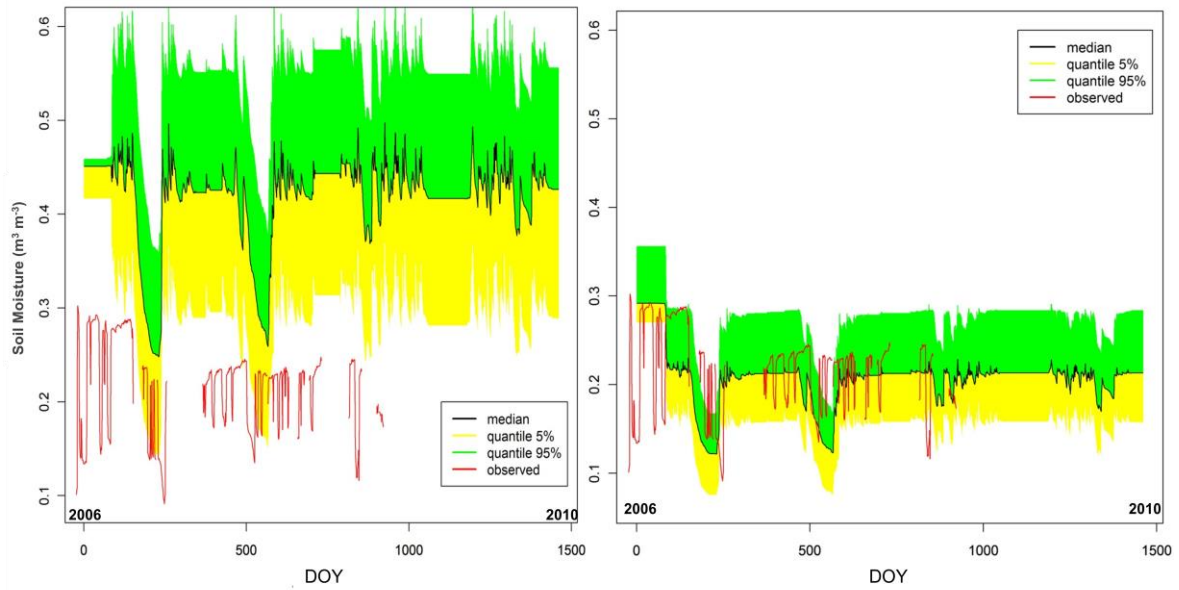


Figure 27. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la teneur en eau du sol à 0.30 m de profondeur avec réduction de précipitation

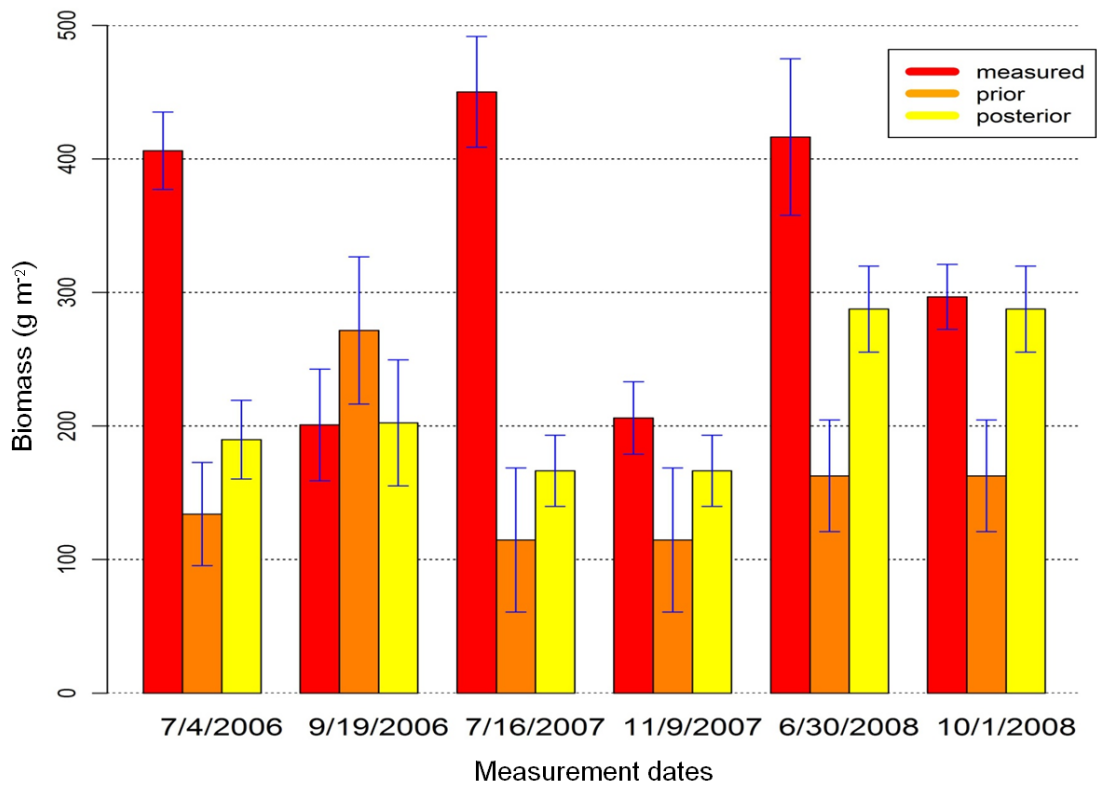


Figure 28. Analyse d'incertitude des estimations du prior et du posterior de la biomasse avec réduction de précipitation

Chamau

Le deuxième site suisse, Chamau n'a pas dérogé de la règle et a présenté lui aussi des améliorations des résultats, comme l'on peut voir par exemple sur la Figure 29, qui montre l'indice de surface foliaire avec et sans réduction de précipitation. Pour les deux traitements, à part deux ou trois dates sur une vingtaine, les résultats du posterior sont meilleurs que ceux du prior. Cela est également vrai pour la teneur en eau du sol et la température du sol à 0.15 m de profondeur en cas de réduction de précipitation (Figure 30). Par contre, les simulations de Chamau ont dévoilé quelques faiblesses du modèle, d'abord avec l'indice de surface foliaire (LAI) que le modèle n'arrive pas à simuler correctement quand les valeurs sont élevées ($>7 \text{ m}^2 \text{ m}^{-2}$, tel étant le cas des deux dates en 2007). Même si le posterior a amélioré le résultat de la simulation, cela reste relativement insuffisant puisque le prior part initialement d'une valeur très faible. Ensuite, les simulations de la teneur en eau du sol ont montré une autre limite du modèle qui n'arrivait pas à dépasser un certain seuil vers le bas (autour de $0.2 \text{ m}^3 \text{ m}^{-3}$, encerclé en vert).

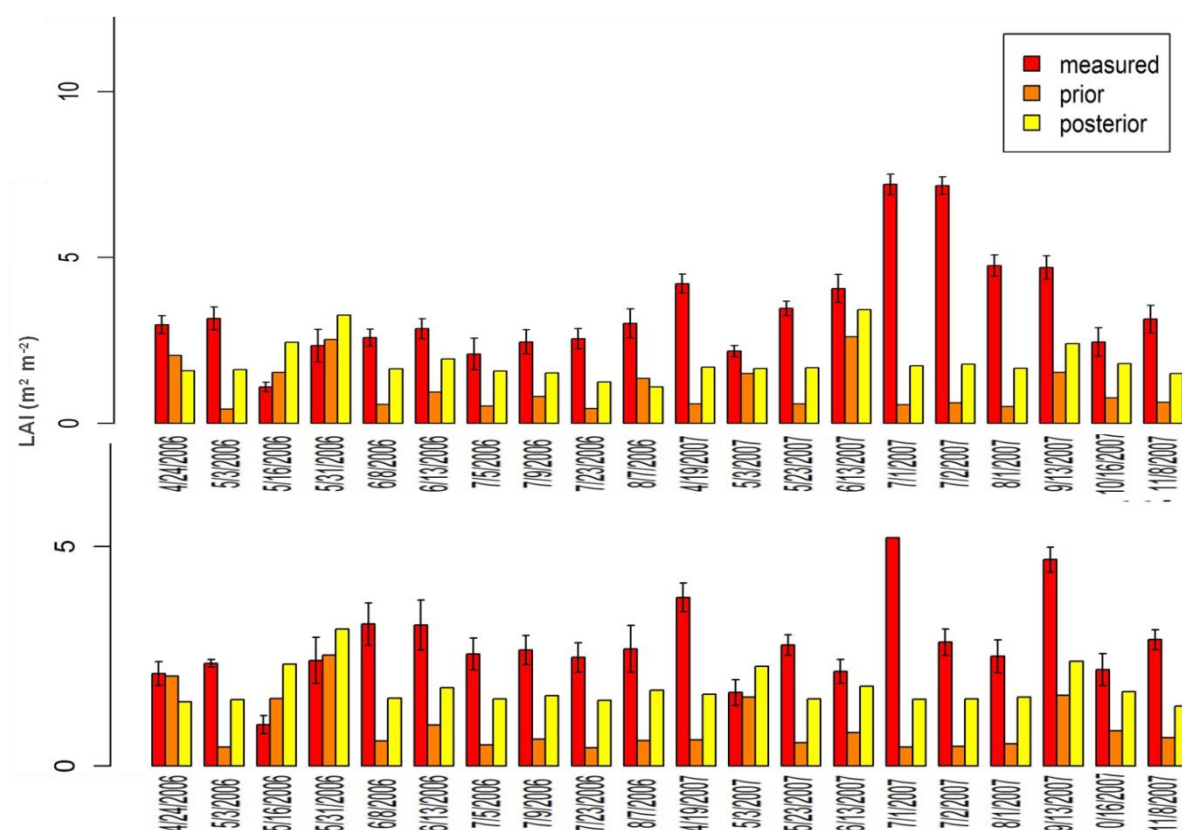


Figure 29. Estimations du prior et du posterior de l'indice des surfaces des feuilles sans (en haut) et avec (en bas) réduction de précipitation

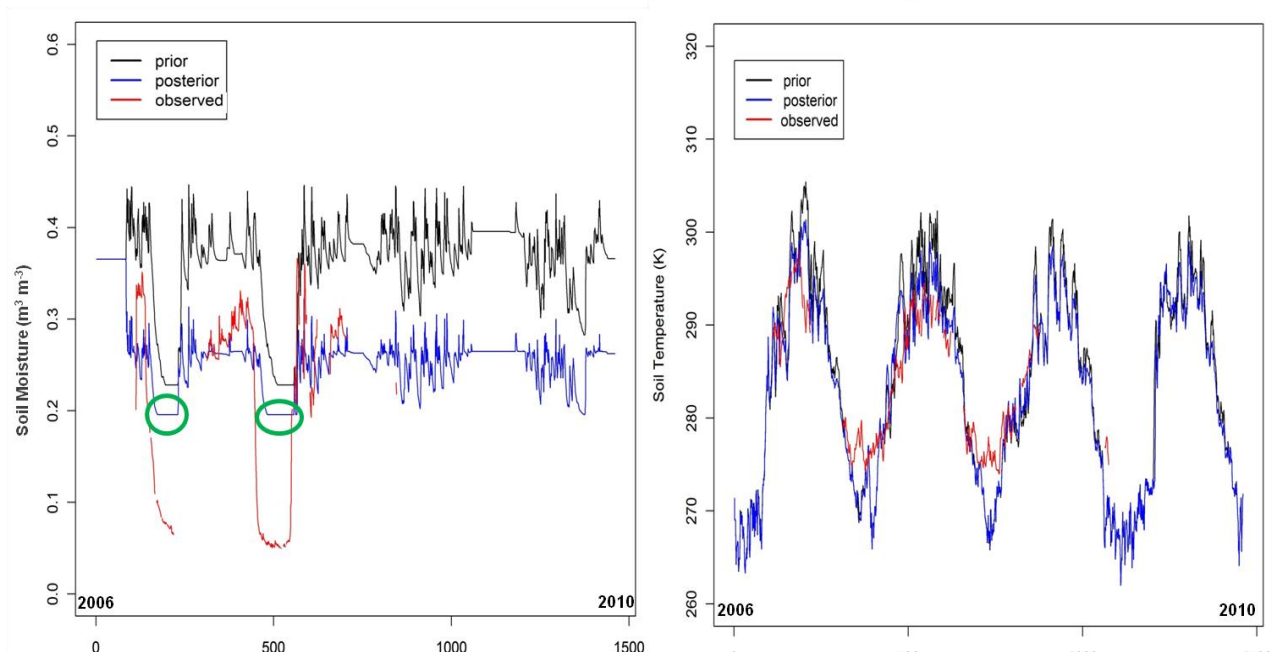


Figure 30. Estimations du prior et du posterior de la teneur en eau du sol (à gauche) et de la température du sol (à droite) à 0.15 m de profondeur avec réduction de précipitation

Calibration Bayésienne de Fruebüel sans LAI

Après avoir appliqué la calibration Bayésienne en utilisant toutes les données disponibles, nous avons testé une calibration Bayésienne de Fruebüel ne tenant pas compte des données de l'indice de surface foliaire (LAI) pour vérifier l'impact d'une telle procédure sur les résultats du modèle.

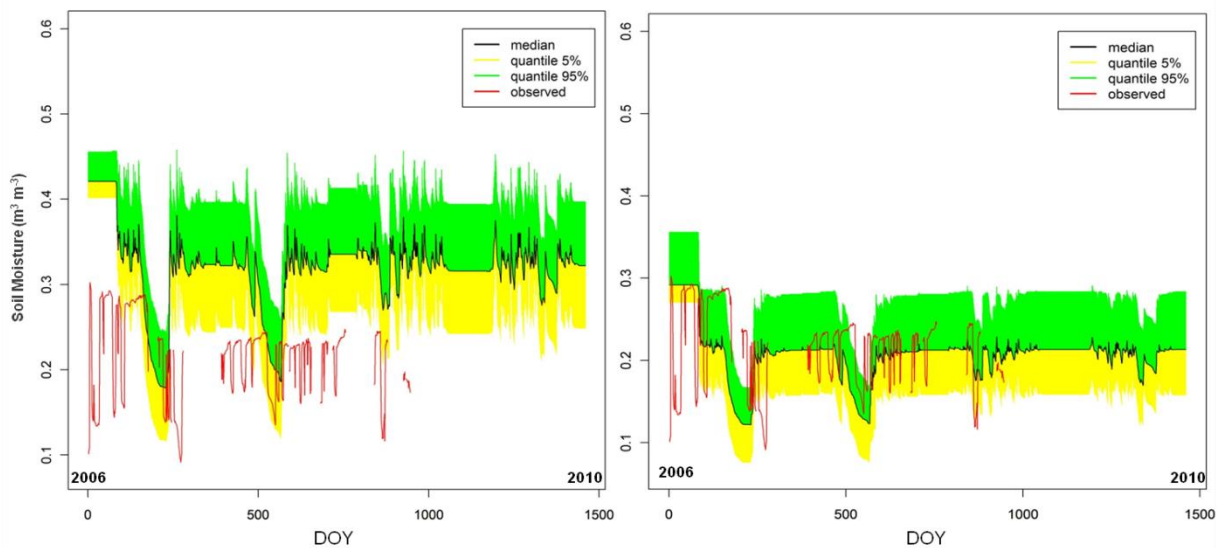


Figure 31. Analyse d'incertitude des résultats de la calibration sans (à gauche) et avec LAI (à droite) de la teneur en eau du sol à 0.30 m de profondeur avec réduction de précipitation

On peut remarquer sur les Figure 31 et Figure 32 que l'incertitude a augmenté, surtout pour la teneur en eau du sol, par rapport à la calibration initiale qui prenait en compte des données du LAI. Cela prouve, d'une part, que le LAI a un impact sur les autres variables de sortie et, d'autre part, que le volume de données utilisé est important pour la mise à jour des connaissances a priori lors d'une calibration Bayésienne (plus on a des données plus le résultat a posteriori de la calibration se rapproche de la valeur observée).

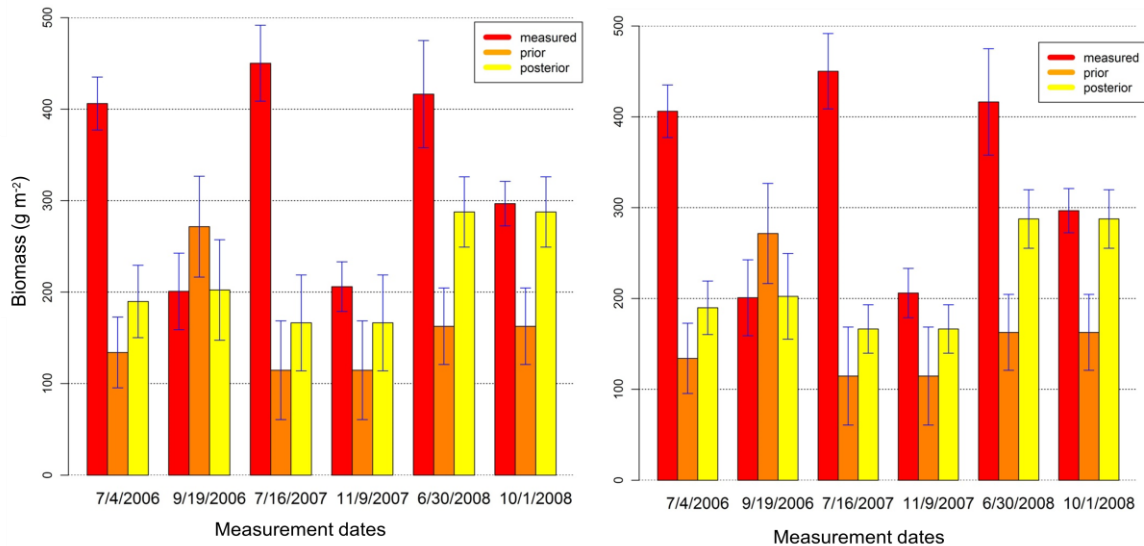


Figure 32. Analyse d'incertitude des résultats de la calibration sans (à gauche) et avec LAI (à droite) de la biomasse avec réduction de précipitation

3.3.3. Evaluation du posteriori des variables de sortie

La capacité des prédictions du modèle résultant des paramètres a priori et a posteriori a été évaluée à l'aide de deux indices statistiques de performance qui sont l'erreur quadratique moyenne (RMSE) et le coefficient de masse résiduelle (CRM), tous les deux basés sur la différence entre les valeurs mesurées (O_i) et les valeurs simulées (P_i), ($n =$ nombre d'échantillons):

$$RMSE = \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{n} \right]^{0.5}$$

$$CRM = \frac{\sum_{i=1}^n (O_i - P_i)}{\sum_{i=1}^n (O_i)}$$

Le RMSE quantifie l'amplitude des différences entre les valeurs simulées et les valeurs mesurées, et varie entre zéro (l'optimum, accord total entre les estimations et les données) et l'infini positif. Le CRM montre la tendance du modèle à sous-estimer (valeurs positives) ou surestimer (valeurs négatives) les observations, zéro étant aussi la valeur optimale.

Le Tableau 8 résume les valeurs du RMSE et CRM obtenues pour quatre variables de sortie à Stubai. Sur les estimations a posteriori, les deux indices sont plus proches de zéro que les estimations a priori pour le LAI et la teneur en eau du sol dans les deux traitements, avec une baisse du RMSE du LAI de plus de 50% et un CRM de la teneur en eau du sol réduit à presque zéro. Pour la surface spécifique des feuilles, des améliorations ont été observées uniquement en cas de réduction des précipitations. Enfin, pas d'amélioration observée sur les estimations du rendement, pour lequel les données disponibles n'étaient pas suffisantes pour tirer des conclusions.

Tableau 8. Evaluation du prior et du posterior des estimations de quatre variables de sortie obtenues avec et sans (Control) réduction de précipitation au site autrichien de Stubai (RMSE, *root mean square error* ; CRM, *coefficient of residual mass*)

Traitement	Information	Indice de surface foliaire (m ² m ⁻²)		Rendement (g MS m ⁻²)		Surface foliaire spécifique (m ² kg ⁻¹)		Moyenne de la teneur en eau du sol (0.05-0.10 m) (m ³ m ⁻³)	
		RMSE	CRM	RMSE	CRM	RMSE	CRM	RMSE	CRM
Control	Prior	1.20	-0.30	15.77	-0.03	1.68	-0.07	0.08	-0.23
	Posterior	0.59	0.15	22.93	0.18	4.34	-0.24	0.03	-0.01
Réduction de précipitation	Prior	1.97	-0.76	5.57	-0.02	11.13	-0.71	0.07	-0.17
	Posterior	0.89	0.35	16.10	0.68	7.73	-0.49	0.03	0.06

Le Tableau 9 récapitule les résultats du RMSE et du CRM obtenus pour deux variables de sortie au deux sites Suisse et pour les deux traitements. La valeur du RMSE de la teneur en eau du sol a été réduite de 50% par rapport au prior à Chamau (pour le traitement) et de 25% par rapport au prior à Frübüel (pour le contrôle et le traitement). Quant au CRM de la teneur en eau du sol, il y a eu une amélioration des résultats pour les deux sites et pour les deux traitements surtout à Frübüel où les valeurs avoisinent le zéro. Pareillement pour le LAI, les simulations a posteriori montrent des résultats meilleurs que celles du prior pour les deux sites.

En général, nous avons constaté que l'amélioration des résultats du modèle obtenue après calibration Bayésienne était plus forte (de ~50%) pour Frübüel que pour Chamau. Cela dépend des différences dans l'hétérogénéité spatiale du site, les caractéristiques de gestion et le climat.

Tableau 9. Evaluation du prior et du posterior des estimations de deux variables de sortie obtenues avec et sans (Control) réduction de précipitation aux deux sites suisses (RMSE, root mean square error ; CRM, coefficient of residual mass)

Site	Traitement	Indice de surface foliaire ($\text{m}^2 \text{m}^{-2}$)				Teneur en eau du sol ($\text{m}^3 \text{m}^{-3}$)			
		CRM		RMSE		CRM		RMSE	
		PRIOR	POSTERIOR	PRIOR	POSTERIOR	PRIOR	POSTERIOR	PRIOR	POSTERIOR
Chamau	Control	0.70	0.45	2.98	2.26	-1.72	-1.40	0.20	0.14
	Réduction de précipitation	0.67	0.36	2.23	1.47	-0.68	-0.18	0.11	0.06
Frübüel	Control	0.55	0.40	2.25	1.79	-1.03	-0.01	0.24	0.06
	Réduction de précipitation	0.46	0.40	1.62	1.44	-1.07	-0.07	0.23	0.06

3.4. Conclusion

La calibration Bayésienne a été appliquée à un certain nombre de paramètres de PaSim en utilisant des données de trois prairies typiques en Europe centrale incluant des traitements induisant des conditions de sécheresse. Malgré les différences entre les sites et les traitements, la calibration Bayésienne site par site a démontré son efficacité dans l'amélioration du paramétrage du modèle même dans des conditions de stress hydrique, ce qui a permis par conséquent d'améliorer les résultats du modèle et de réduire les incertitudes.

L'évaluation des résultats de cette étude a été effectuée en utilisant les mêmes données qui ont servies pour la calibration. En effet, l'accès à une base d'observations plus large aurait permis de tester la robustesse des solutions obtenues à travers une validation indépendante. Puis, les résultats de l'application de la méthode, propres à chaque site, ne permettent pas d'exploiter les nouveaux paramétrages pour réaliser des simulations en

dehors des sites d'étude et, notamment, à l'échelle de grand territoire. Malgré cela, cette étude a permis de mettre en évidence les potentialités de l'approche Bayésienne pour répondre de manière efficace à la réduction des incertitudes touchant aux résultats de la modélisation des prairies conduites en condition de sécheresse.

Dans le chapitre suivant, nous allons aborder la deuxième approche de cette thèse qui consiste à lancer une calibration Bayésienne multi-sites Européens diversifiés climatiquement et géographiquement et avec des données à long termes (≥ 6 ans), tout en passant par une étape d'analyse de sensibilité pour déterminer les paramètres les plus influents à calibrer. L'objectif, dans ce cas, est d'obtenir un paramétrage commun à l'échelle de l'Europe et de le tester les résultats des simulations avec des sites de validation.

Chapitre 4 : Nouveau paramétrage de PaSim à l'échelle de l'Europe (données à long terme)

4.1. Introduction

Dans le chapitre précédent, nous avons démontré les performances de la calibration Bayésienne sur des sites prairiaux conduits en condition de sécheresse induite, notamment sa capacité à améliorer les résultats du modèle et à réduire les incertitudes des paramètres et par conséquent des sorties. Dans ce chapitre, nous menons une deuxième étude plus large et plus globale couvrant divers sites prairiaux européens, dans laquelle nous procédons à une analyse de sensibilité de PaSim afin d'identifier les paramètres de végétation les plus pertinents et qui feront l'objet d'une calibration. Cette fois c'est une calibration Bayésienne multi-sites à l'échelle de l'Europe qui a été mise en œuvre dans le but d'obtenir un nouveau paramétrage de PaSim valable pour les prairies européennes.

La première partie de ce chapitre a été publiée dans *Ecological Modelling* [Ben Touhami et al., 2013a]. Dans cette partie, nous allons présenter l'application de l'analyse de sensibilité à PaSim, notamment la méthode utilisée, les sites de référence, les données et les paramètres à analyser. Ensuite, nous allons présenter et discuter les résultats obtenus de l'analyse de sensibilité et leur stabilité à l'échelle européenne (plasticité de PaSim).

La deuxième partie de ce chapitre a été partiellement illustrée lors de la Greenhouse Gases & Animal Agriculture Conference, 23-26 June, Dublin, Ireland [Ben Touhami et al., 2013b]. Nous y exposons tout le raisonnement de la calibration Bayésienne multi-sites. Le choix des paramètres à calibrer (découlant de l'analyse de sensibilité) et le choix des sites et des données utilisés relèvent d'un protocole que nous avons élaboré. Enfin, nous montrons les résultats issus de cette calibration sur les sites utilisés ainsi que les résultats de l'utilisation du nouveau paramétrage obtenu sur des sites de validation.

Le travail que nous allons voir dans ce chapitre a fait l'objet aussi d'un cours « training session » et une communication orale lors de la réunion annuelle du projet Européen ANIMALCHANGE (Dublin, 26-28 juin 2013). Dans le cadre d'une collaboration entre INRA et Embrapa (<http://www.cnptia.embrapa.br/english.html>) nous avons également assuré deux cours de calibration Bayésienne et d'analyse de sensibilité, et nous avons présenté nos travaux de recherche à Campinas, Brésil, du 12 au 21 novembre 2013.

4.2. Analyse de sensibilité de PaSim

Avant d'entamer la calibration Bayésienne multi-sites, une analyse de sensibilité de PaSim a été effectuée afin de réduire le nombre de paramètres à calibrer pour ne garder que les paramètres les plus influents sur les sorties. Cette étude a nécessité l'élaboration d'un protocole que nous allons détailler dans ce qui suit pour le choix des sites, des données et des paramètres.

4.2.1. Méthodologie

Sites étudiés

Six sites européens couvrant une gamme de latitudes, longitudes et conditions climatiques diverses (Tableau 10), ainsi qu'une variété de pratiques de gestion (Tableau 11), ont été considérés pour cette étude. Ces sites prairiaux, à l'exception d'Oensingen (qui a été mis en place en 2001), sont des prairies semi-naturelles en place depuis longtemps contenant des types de végétation représentatifs de la zone.

Tableau 10. Informations et climats des sites étudiés

Pays	Site	ID	Année	Coordonnées géographiques			Climat			Sources
				Latitude	Longitude	Altitude (m a.s.l.)	Température moyenne de l'air (°C)	Précipitation totale (mm an ⁻¹)	Evapotranspiration de référence (mm an ⁻¹)	
France	Laqueuille	LAQ-1	2002-2009	45°38'N	02°44'E	1040	7.8	1072	769	Klumpff et al. (2011)
		LAQ-2	2002-2009							
Hungary	Bugac-Puszta	BUG	2003-2008	46°41'N	19°36'E	140	10.2	520	890	Pintér et al. (2008)
Italy	Amplero	AMP	2003-2007	41°52'N	13°38'E	884	9.4	781	927	Wohlfahrt et al. (2008)
Switzerland	Friedbüel	FRU	2006-2011	47°06'N	08°32'E	982	7.7	1625	643	Gilgen and Buchmann (2009), Zeeman et al. (2010)
	Oensingen	OEN	2002-2009	47°17'N	07°44'E	450	9.3	1197	736	Ammann et al. (2007)
United Kingdom	Easter Bush	EAB	2002-2008	55°52'N	03°02'W	190	9.0	956	610	Soussana et al. (2007)

Les pratiques de gestion des prairies (Tableau 11) reflètent l'environnement socio-économique dans chaque pays. Les sites étudiés ont été soit fauchés ou pâturés, soit soumis

à une combinaison des deux. La classification des sites entre « extensifs » et « intensifs » était essentiellement basée sur l'intensité de la fertilisation, le nombre de fauches et la densité des animaux et /ou la durée de pâturage.

Tableau 11. La gestion des sites étudiés

ID	Utilisation	Gestion				
		Fauches (an ⁻¹)	Fertilisation		Pâturage	
			Événements (an ⁻¹)	Quantité totale apportée (kg N ha ⁻¹ an ⁻¹)	Chargement (UGB ⁸ ha ⁻¹ an ⁻¹)	Durée (jours an ⁻¹)
LAQ-1	Intensive	-	3	210	1.1	160
LAQ-2	Extensive	-	-	-	0.6	160
BUG	Extensive	-	-	-	0.3	180
AMP	Semi-intensive	1	-	-	0.5	200
FRU	Extensive	1-2	-	-	0.03	10
OEN	Intensive	4-6	4	214	-	-
EAB	Intensive	0-2	1-4	160	1.4	160

La classification des sites s'est basée principalement sur des critères climatiques et géographiques. Pour cela, la température moyenne et les précipitations totales ont été obtenus à partir des données météorologiques horaires disponibles pour chaque site ainsi que la quantité annuelle d'évapotranspiration de référence (ET_0 , mm), cette dernière étant calculée en utilisant la méthode Penman-Monteith avec un pas de temps horaire [Walter et al., 2005]. Ce qui nous a permis de calculer deux indicateurs bioclimatiques (Figure 33). L'indice d'aridité de Martonne (b) a été élaboré à partir de la formule de Gottmann [De Martonne, 1942]. Cet indice s'est avéré approprié dans les études de l'impact du changement climatique sur les systèmes herbagers, car il prend en compte à la fois la variabilité annuelle et intra-annuelle de la température et la précipitation. La possibilité d'établir une distinction entre les conditions climatiques associées aux gradients d'aridité est donnée par les limites publiées par [Diodato et Ceccarelli, 2004] : $b < 5$: aridité extrême ; $5 \leq b \leq 14$: aridité ; $15 \leq b \leq 19$: semi-aridité ; $20 \leq b \leq 29$: sous-humidité ; $30 \leq b \leq 59$: humidité ; $b > 59$: forte humidité. Le deuxième indicateur est le bilan hydrique climatique ($P-ET_0$, mm) qui a été également calculé sur la base des valeurs annuelles de précipitations

⁸ UGB : Unité de Gros Bétail (équivalent français du LSU) est une unité de référence permettant d'agrèger le bétail de différentes espèces et de différents âges en utilisant des coefficients spécifiques établis initialement sur la base des besoins nutritionnels de chaque type d'animal.

(P) et évapotranspiration (ET_0) afin d'évaluer le déficit (valeurs négatives) ou l'excédent (valeurs positives) d'eau de chaque site. Le déficit climatique d'eau, en particulier, est très important pour la végétation car il touche directement la physiologie des plantes. Il représente la demande d'évaporation non satisfaite par l'eau disponible et est considéré comme une mesure de la sécheresse absolue (indépendamment de la végétation actuelle du site). Pour chaque site sélectionné, les simulations ont été effectuées pour toutes les années des données de météo et de gestion disponibles. Cela a permis à l'analyse de sensibilité de prendre en compte de la variabilité interannuelle au niveau de chaque site.

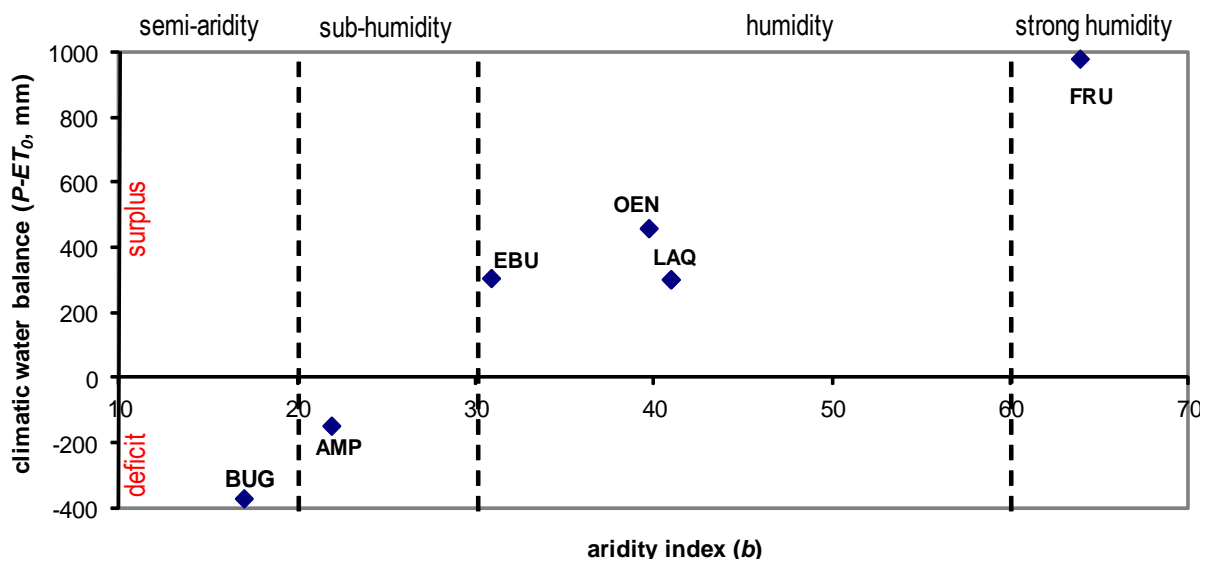


Figure 33. Classification des sites selon deux indices bioclimatiques (l'indice d'aridité et le bilan hydrique climatique) liés à la disponibilité de l'eau

Le critère géographique du choix des six sites de cette étude se manifeste par leur répartition le long d'un gradient allant du nord-ouest jusqu'au sud-est (Figure 34), notamment des îles Britanniques jusqu'en Italie et Hongrie, y compris les hauts plateaux volcaniques du centre de la France, couvrant ainsi la plupart des zones climatiques européennes. L'indice d'aridité (Figure 33) varie du semi-aride (Bugac-Puszta, Hongrie) jusqu'au fort-humide (Früebüel, Suisse), la valeur moyenne étant comprise entre 17 à Bugac et 64 à Früebüel (région de montagne en Suisse). Bugac-Puszta représente un climat continental, avec des étés chauds, averses de pluie et des hivers modérément froids, en plus il est caractérisé par des déficits annuels d'eau. Le site suisse de Früebüel est représentatif des montagnes alpines, dominées par des conditions humides. Les autres sites sont pour la plupart humides, avec des valeurs de l'indice d'aridité entre 20 et 40.

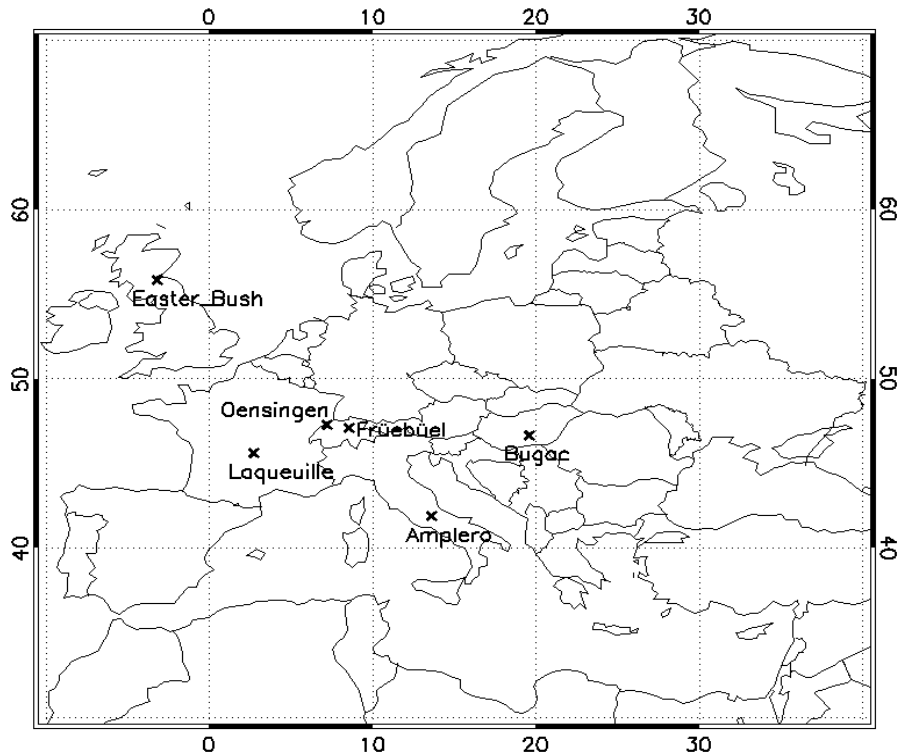


Figure 34. Localisation géographique des sites

Variables de sortie et paramètres de PaSim

La sensibilité de neuf variables de sortie liées à la végétation, au sol et à l'atmosphère a été examinée. Ces variables, qui sont les suivantes, semblaient suffisantes pour apprécier la relation structurelle entre les réalisations du modèle et ses paramètres : la production primaire brute (GPP, $\text{kg C m}^{-2} \text{ yr}^{-1}$), la production primaire nette (NPP, $\text{kg C m}^{-2} \text{ yr}^{-1}$), l'indice de surface des feuilles (LAI, $\text{m}^2 \text{ m}^{-2} \text{ d}^{-1}$), le rendement (Y, $\text{kg MS m}^{-2} \text{ yr}^{-1}$), la teneur en eau du sol (SWC, $\text{m}^3 \text{ m}^{-3} \text{ d}^{-1}$), la température du sol (T_{SOIL} , K d^{-1}), la concentration du substrat de carbone en plante (C, $\text{kg C kg}^{-1} \text{ MS d}^{-1}$), la production nette du biome (NBP, $\text{kg C m}^{-2} \text{ yr}^{-1}$) et le méthane émis par les herbivores au cours de la période de pâturage (CH_4 , $\text{kg CH}_4\text{-C eq. m}^{-2} \text{ yr}^{-1}$). Pour les variables ayant une production annuelle (GPP, NPP, le rendement et la NBP), l'impact de la variation des paramètres a été examiné sur la moyenne des valeurs annuelles cumulées. Pour les autres variables, la moyenne des valeurs quotidiennes simulées sur plusieurs années est la sortie de référence pour l'analyse de sensibilité.

La production primaire est une estimation de la production de la plante de composés organiques à partir de dioxyde de carbone (CO_2) atmosphérique, elle est soit brute (GPP) soit nette (NPP), cette dernière prenant en compte les pertes dues à la respiration cellulaire.

La production nette du biome (NBP) est le gain ou la perte nette de carbone au niveau de l'écosystème. La concentration en carbone de la plante (C) est une variable clé dans la description de l'état et de la dynamique de la plante, jouant un rôle important dans la croissance et la répartition des assimilats ; son taux de variation est déterminé par l'afflux de carbone acquis dans la photosynthèse et les pertes dues à la respiration. Le rendement de l'herbe (Y) est la quantité de biomasse de matière sèche qui est retirée du terrain à chaque événement de fauche (coupe). L'indice de surface foliaire (LAI) est une mesure de la capacité d'utilisation de la lumière par la plante, qui est à la base de la prévision de la croissance de la production primaire, l'évapotranspiration et la croissance de la végétation. Le méthane (CH₄) dû à la fermentation entérique des ruminants au pâturage est important car il contribue aux émissions mondiales de gaz à effet de serre. Les connaissances biophysiques de l'écosystème des prairies ont été complétées par deux variables du sol : teneur en eau et température (SWC et T_{SOIL}), dont le profil provient de la modélisation de la physique du sol véhiculée par le bilan énergétique de surface. Elles ont été estimées pour chacune des six couches qui composent le profil du sol (exigence de base de PaSim).

La variation des sorties en réponse au changement des paramètres a été étudiée en utilisant la méthode de Morris, telle que fournie par le langage R (<http://rss.acs.unt.edu/Rdoc/library/sensitivity/html/sensitivity-package.html>). Cette implémentation comprend des améliorations par rapport à la méthode originale en ce qui concerne la stratégie d'échantillonnage (optimisation du remplissage de l'espace et design basé sur le simplexe). La librairie sensibilité du langage R a été exécutée pour calculer la moyenne (μ) et l'écart type (σ) des variations des sorties par rapport aux paramètres dans six trajectoires et six niveaux de l'espace des paramètres. En principe, afin d'obtenir un échantillon significatif, un nombre de niveaux à explorer doit être pondéré par un nombre de trajectoires. Notre choix correspond à une indication de la robustesse des résultats d'analyse de sensibilité que le nombre de trajectoires est au moins cinq [Confalonieri et al., 2010a] et un nombre égal de trajectoires et de niveaux (six, six) est pris en compte (comme conseillé par [Saltelli et al., 2004] et confirmé par [Richter et al., 2010]).

Étant donné que l'analyse de sensibilité nécessite de connaître les distributions de probabilité des valeurs des paramètres du modèle afin de calculer les mesures de sensibilité, un ensemble de valeurs a été associé à chaque paramètre. Vingt-huit paramètres d'entrée, liés à la végétation, sont considérés dans cette étude dont les valeurs de référence sont données dans le Tableau 12, tels que fournis par l'examen de la littérature disponible

sur PaSim [Johnson et Thornley, 1985 ; Thornley et Verberne, 1989 ; Riedo et al, 1998, 1999 ; Graux, 2011]. Pour chaque paramètre, les valeurs moyennes ont été établies entre les bornes supérieure et inférieure, et les écarts-types ont été calculés sous forme de fraction (1/6, ex. [Foscarini et al., 2010]) du rang de variabilité.

La librairie sensibilité de R a été introduite à l'aide de l'ensemble des valeurs du Tableau 12 et PaSim a été exécuté de façon répétée avec une combinaison de valeurs des paramètres, échantillonnées à partir de deux distributions (en l'absence de connaissance d'une distribution appropriée). D'après d'autres études d'analyse de sensibilité (ex. [Foscarini et al., 2010]), les deux distributions normale (Gaussienne) et uniforme ont été employées pour évaluer l'influence de la forme de distribution sur les résultats de l'analyse. L'hypothèse que les valeurs des paramètres sont uniformément réparties sur une gamme appropriée de variation est fréquente, surtout lorsque l'objectif principal est de comprendre le comportement du modèle [Monod et al., 2006]. Généralement, si une sortie est insensible aux paramètres avec une distribution uniforme, elle le sera encore plus pour une autre distribution car la distribution uniforme attribue l'incertitude la plus raisonnable pour les paramètres [Vose Software, 2007]. La distribution Gaussienne repose plutôt sur l'hypothèse que, pour chaque paramètre, certaines valeurs (près de la moyenne) sont plus probables que d'autres. La distribution normale considérée varie selon les limites inférieures et supérieures du Tableau 12.

Bien que l'analyse de sensibilité nécessite beaucoup moins de temps de calcul que la calibration Bayésienne, nous avons eu recours au cluster du LIMOS pour lancer l'analyse de sensibilité comme pour la calibration, et pour le traitement et l'analyse des résultats.

Tableau 12. Liste des paramètres considérés pour l'analyse de sensibilité

Nom	Paramètres		Valeurs		Description
	Symbole	Unité	Min	Max	
Paramètre de la distribution de la matière sèche des racines par rapport aux différentes couches de sol	b_root_profile	-	5	15	Il calcule la proportion de racines à une profondeur donnée.
Capacité photosynthétique foliaire pendant la période reproductrice	pmco2rep	$\mu\text{mol m}^{-2} \text{s}^{-1}$	15.0	22.5	Ils représentent l'influence de la phase de développement sur le taux de lumière de la photosynthèse des feuilles à saturation (défini dans les conditions normales de température, de la concentration de
Capacité photosynthétique foliaire pendant la période végétative	pmco2veg	$\mu\text{mol m}^{-2} \text{s}^{-1}$	10	15	

					CO ₂ dans l'atmosphère et de la concentration en N de la plante), qui est une composante du taux de couvert photosynthétique.
Proportions de fibres digestibles de l'épis - age 1	DNDFear1	kg digestible fibres kg ⁻¹ fibres	0.71	0.77	
Proportions de fibres digestibles de l'épis - age 2	DNDFear2	kg digestible fibres kg ⁻¹ fibres	0.62	0.65	
Proportions de fibres digestibles de l'épis - age 3	DNDFear3	kg digestible fibres kg ⁻¹ fibres	0.54	0.57	Ces paramètres influent sur la quantité
Proportions de fibres digestibles de l'épis - age 4	DNDFear4	kg digestible fibres kg ⁻¹ fibres	0.41	0.45	quotidienne de détergent neutre
Proportions de fibres digestibles de limbe - age 1	DNDFlam1	kg digestible fibres kg ⁻¹ fibres	0.90	0.97	des fibres digestibles (<i>DNDF</i>)
Proportions de fibres digestibles de limbe - age 2	DNDFlam2	kg digestible fibres kg ⁻¹ fibres	0.83	0.85	à partir de différents
Proportions de fibres digestibles de limbe - age 3	DNDFlam3	kg digestible fibres kg ⁻¹ fibres	0.77	0.80	composants de la plante (épis, limbe, gaine / tige) dans la consommation de
Proportions de fibres digestibles de limbe - age 4	DNDFlam4	kg digestible fibres kg ⁻¹ fibres	0.75	0.77	l'animal, en fonction de l'âge des tissus (quatre classes).
Proportions de fibres digestibles de la tige - age 1	DNDFstem1	kg digestible fibres kg ⁻¹ fibres	0.81	0.88	
Proportions de fibres digestibles de la tige - age 2	DNDFstem2	kg digestible fibres kg ⁻¹ fibres	0.74	0.76	
Proportions de fibres digestibles de la tige - age 3	DNDFstem3	kg digestible fibres kg ⁻¹ fibres	0.63	0.65	
Proportions de fibres digestibles de la tige - age 4	DNDFstem4	kg digestible fibres kg ⁻¹ fibres	0.58	0.60	
Surface foliaire spécifique maximale	slam	m ² kg ⁻¹	28	35	Il s'agit de la valeur maximale de la surface foliaire spécifique, définie comme étant le rapport de la surface de la feuille et le poids de la matière sèche, utilisée pour dériver la zone de surface foliaire de la biomasse foliaire.
paramètre contrôlant la teneur en N dans la matière sèche structurelle	fnref	-	0.017	0.021	Ce paramètre est utilisé pour calculer la concentration en azote de la matière sèche structurelle produite récemment.
Teneur en fibres dans les épis ingérés par les herbivores	NDFear	-	0.64	0.68	Ces paramètres représentent la proportion de fibres (<i>NDF</i>) dans les tissus ingérés (épis, limbe, gaine / tige).
Teneur en fibres dans les limbes ingérés par les herbivores	NDFlam	-	0.52	0.57	
Teneur en fibres dans les tiges et gaines ingérées par les herbivores	NDFstem	-	0.60	0.64	
Teneur en C dans la matière sèche structurelle racinaire	fcr	kg C kg ⁻¹ DM	0.45	0.50	Ces paramètres se multiplient les taux de croissance des racines et pousses, respectivement, pour obtenir la variation du C de substrat.
Teneur en C dans la matière sèche structurelle aérienne	fcsh	kg C kg ⁻¹ DM	0.43	0.47	

Hauteur du couvert à la floraison	hcanmax	m	0.3	0.7	Il détermine la hauteur du couvert pour calculer les flux de chaleur sensible et latente du couvert et de la surface du sol.
Taux de rotation de la partie structurelle aérienne à 20 °C	ageing_1gamr20	d ⁻¹	0.036	0.054	Ils représentent la durée de vie des feuilles et de la racine à une température constante, utilisés pour calculer le flux de résidus végétaux
Taux de rotation de la racine à 20 °C	ageing_1gamsh20	d ⁻¹	0.0096	0.0144	
Facteur de normalisation de l'indice de développement	tasumrep	K-d	604	814	Ce paramètre normalise l'indice de l'étape de développement (<i>devstage</i>) de telle manière que <i>devstage</i> = 1 marque le passage du stade reproductif au stade végétatif.
Valeur de l'indice de développement	devear	-	0.47	0.62	Ce paramètre représente une étape intermédiaire entre le début et la fin de l'émergence des épis, lorsque la végétation atteint le stade de "épis à 10 cm".
Concentration maximale en N de la plante	un_1ntotmax	kg N kg ⁻¹ DM	0.028	0.042	Ce paramètre limite l'absorption de N par la plante.

La méthode de Morris utilise un plan expérimental de type OAT ("*One-factor-At-a-Time*"), où chaque facteur bouge aléatoirement par rapport à la combinaison précédente. Grâce à cela, une trajectoire est décrite dans l'espace formé de tous les paramètres. La sensibilité de chaque paramètre est estimée à travers l'effet de la variation élémentaire du paramètre sur la variable d'intérêt, et cela à partir de plusieurs points de l'espace. Peu de trajectoires étant nécessaires pour obtenir un résultat robuste, six trajectoires de 29 points (28 paramètres + 1 point initial) ont été utilisées dans notre cas. Pour chaque paramètre, six niveaux ont été définis, chaque niveau correspondant aux valeurs suivantes de la fonction de répartition : 0 ; 0,2 ; 0,4 ; 0,6 ; 0,8 ; 1 (Figure 35). Cette conception a été répétée en se basant sur deux hypothèses de distribution des paramètres du modèle, une distribution uniforme et une distribution gaussienne.

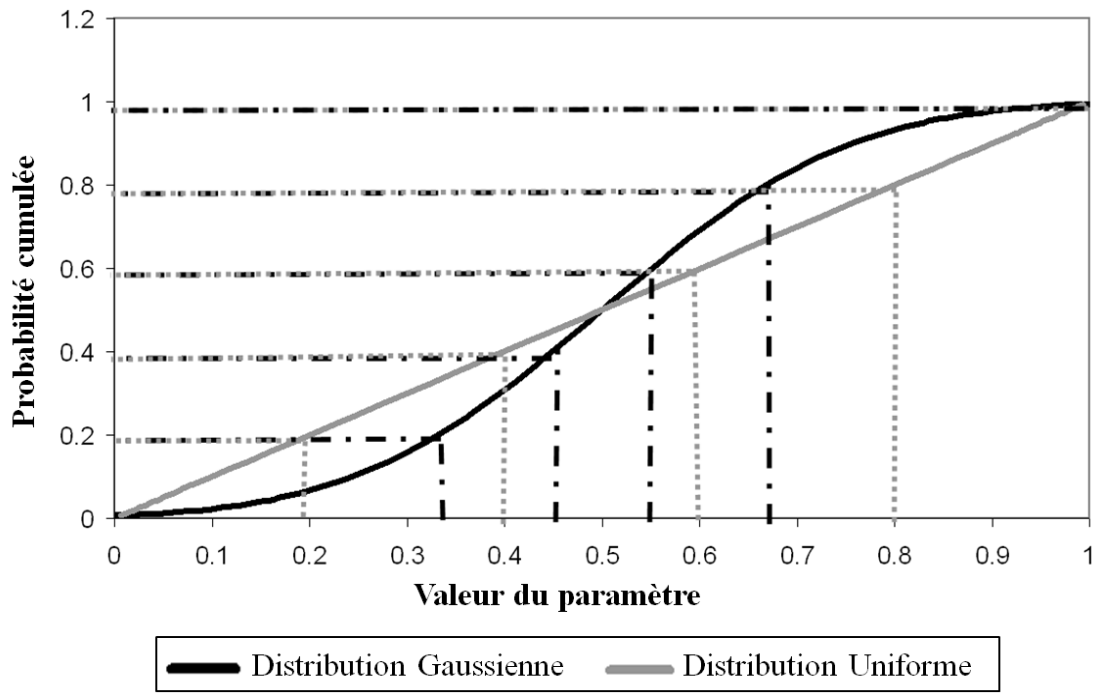


Figure 35. Niveaux statistiques définis à partir de deux fonctions de distribution cumulée : uniforme (gris) et Gaussienne (noire)

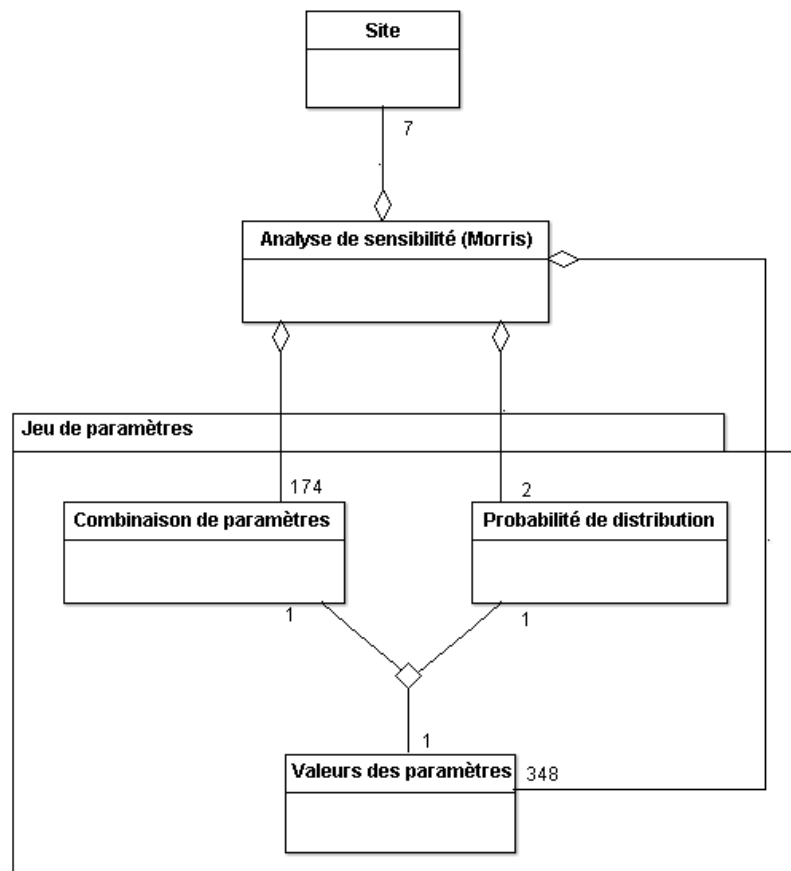


Figure 36. Diagramme du plan expérimental pour l'analyse de sensibilité

La Figure 36 montre le plan expérimental utilisé pour l'analyse de sensibilité. La classe Valeurs des paramètres résulte de l'association entre les Combinaisons de paramètres et la Probabilité de distribution. La cardinalité de la classe Combinaison de paramètres est de 174 pour six trajectoires de 29 points (dans un design de Morris). La cardinalité de la classe Valeurs des paramètres, quant à elle est de 348, puisqu'il y a deux types de distributions de probabilité (2×174).

4.2.2. Plasticité du modèle

Lors de cette étude, nous avons aussi visé à étudier les propriétés du modèle à travers sa plasticité qui est la capacité du modèle à modifier son comportement (par exemple la pertinence de ses paramètres) dans des conditions différentes (ex. par le climat et la gestion). Basé sur [Confalonieri et al., 2012], un indicateur de plasticité du modèle (L) a été calculé en fonction de l'équation suivante :

$$L = TDCC \cdot e^{\sigma_{SAM}^{-1}}$$

où $TDCC$ est le coefficient de concordance [Iman et Conover, 1987], compris entre 0 (absence de concordance) et 1 (concordance totale), et σ_{SAM} est l'écart-type de l'indicateur agro-météorologique synthétique (SAM) de [Confalonieri et al., 2010b], variant entre -1 et 1 dont voici l'équation :

$$SAM = \frac{P - ET_0}{P + ET_0}$$

P (mm) et ET_0 (mm) sont, respectivement, les précipitations annuelles moyennes et le taux d'évapotranspiration de référence du site. Cet indicateur est facilement calculé en utilisant des données climatiques disponibles et il est pertinent pour les systèmes de production agricoles parce que les principaux effets du climat sur les plantes sont déterminés par les interactions de l'énergie et de l'eau [Stephenson, 1990].

$TDCC$ est couramment utilisé pour mesurer l'accord entre les séries de classement, avec une accentuation sur les premiers rangs, qui est réalisée en utilisant les inverses des rangs comme scores [Savage, 1956]. Dans les études d'analyse de sensibilité, le $TDCC$ est appliqué aux classements des paramètres obtenus avec chaque répétition afin de déterminer s'il existe un accord entre les répétitions [Helton et al., 2005]. Ici, les sites sont considérés comme des répétitions. L'indicateur de plasticité L est évalué dans une fourchette allant de

0 (plasticité la plus élevée) à environ 1,5 [Confalonieri et al., 2012]. La relation exponentielle entre σ_{SAM} et $TDCC$ s'est avérée capable de réduire l'asymétrie dans la répartition de L .

4.2.3. Résultats et discussion

L'exécution du processus d'analyse de sensibilité a pris quatre jours pour les deux distributions de référence Gaussienne et uniforme, toutes les deux lancées en parallèle sur le cluster du LIMOS.

Les résultats des analyses ont été évalués graphiquement par l'intermédiaire de matrices codées et colorées de μ et σ pour aider à interpréter visuellement les variations des sorties du modèle par rapport aux changements dans les valeurs des paramètres. Le degré de sensibilité varie du vert (peu ou pas de pertinence) au rouge (plus pertinent), sur une échelle allant de zéro à max (μ , σ). Pour chaque sortie cible, les paramètres influents étaient ceux dont $\mu_{moy} < \mu \leq \mu_{max}$ (moyenne et max de μ à travers les paramètres et les sites). Cela correspond approximativement à un critère (quelque peu subjectif) du premier point de discontinuité de droite à gauche sur l'échelle de μ , conseillé pour distinguer les paramètres les plus influents (situés à la droite d'une discontinuité) de ceux moins- ou non-influents (ex. [Saltelli et al, 2004 ; Confalonieri et al, 2006 ; Foscarini et al, 2010]). L'interprétation des résultats a été faite en combinant des matrices colorées et des schémas de diagnostic (σ contre μ) pour chaque site et chaque sortie.

Ainsi, deux listes de paramètres influents pour chaque variable de sortie et chaque site ont été obtenues, en fonction de l'hypothèse de distribution des paramètres : Gaussienne et uniforme. Les valeurs de concordance et de plasticité ont été également calculées pour les deux distributions. Seuls quelques exemples de résultats (variables de sortie et sites sélectionnés) sont présentés graphiquement et les principaux résultats (classements, concordance, plasticité) discutés ci-après, tandis que le reste des résultats graphiques est joint dans l'annexe II. L'ensemble des résultats de cette étude est accessible en ligne (https://www1.clermont.inra.fr/urep/modeles/PASIM/matrix_gaussian/matrix_gaussian.html).

Distribution Gaussienne

Les matrices colorées de la Figure 37 montrent le degré de sensibilité (μ , panneau de gauche ; σ , panneau de droite) des estimations de la GPP utilisant les paramètres étudiés dans chaque site. Pour la GPP, prise comme exemple, la Figure 38 trace σ par rapport à μ pour chaque paramètre à un site sélectionné (Bugac-Puszta, Hongrie). Elle montre que chaque paramètre avec une valeur élevée de μ a aussi une grande valeur de σ . De gauche vers la droite de l'échelle de μ , les valeurs de μ n'augmentent pas toujours graduellement, mais avec des discontinuités permettant de distinguer les paramètres les plus influents des autres paramètres moins ou non influents.

Trois groupes de paramètres avec des frontières bien distinctes sont évidentes. Les paramètres les moins pertinents abordent l'origine dans la zone inférieure gauche, et sont très loin de la zone en haut à droite. La deuxième série des paramètres haut-classés incluent slam, pmco2rep, fcsh, fcr et tasumrep, tous servant à modéliser le développement et la croissance de la végétation dans PaSim. Les paramètres de ce groupe ne sont pas à proximité de la zone en haut à droite. Un groupe de deux paramètres classés dans la région en haut à droite est clairement séparé des autres. Le paramètre de la fraction de N dans la matière sèche structurale (fnref) est clairement le paramètre le plus influent. Le taux de photosynthèse pour la phase végétative (pmco2veg) est également un paramètre pertinent. Avec une moyenne μ d'environ 0,5 (contre 0,06 en moyenne générale), ce groupe indique une grande influence sur l'estimation de la GPP. Ces résultats reflètent bien les résultats qualitatifs observés dans la matrice des couleurs (Figure 37) où les deux paramètres les plus pertinents ont des cellules allant de l'orange au rouge. En effet, la Figure 38 est une sorte de zoom sur la colonne du site Bugac-Puszta de la Figure 37. Les résultats globaux de l'analyse de sensibilité utilisant une distribution Gaussienne pour tous les paramètres, les sorties et les sites sont dressés dans le Tableau 13.

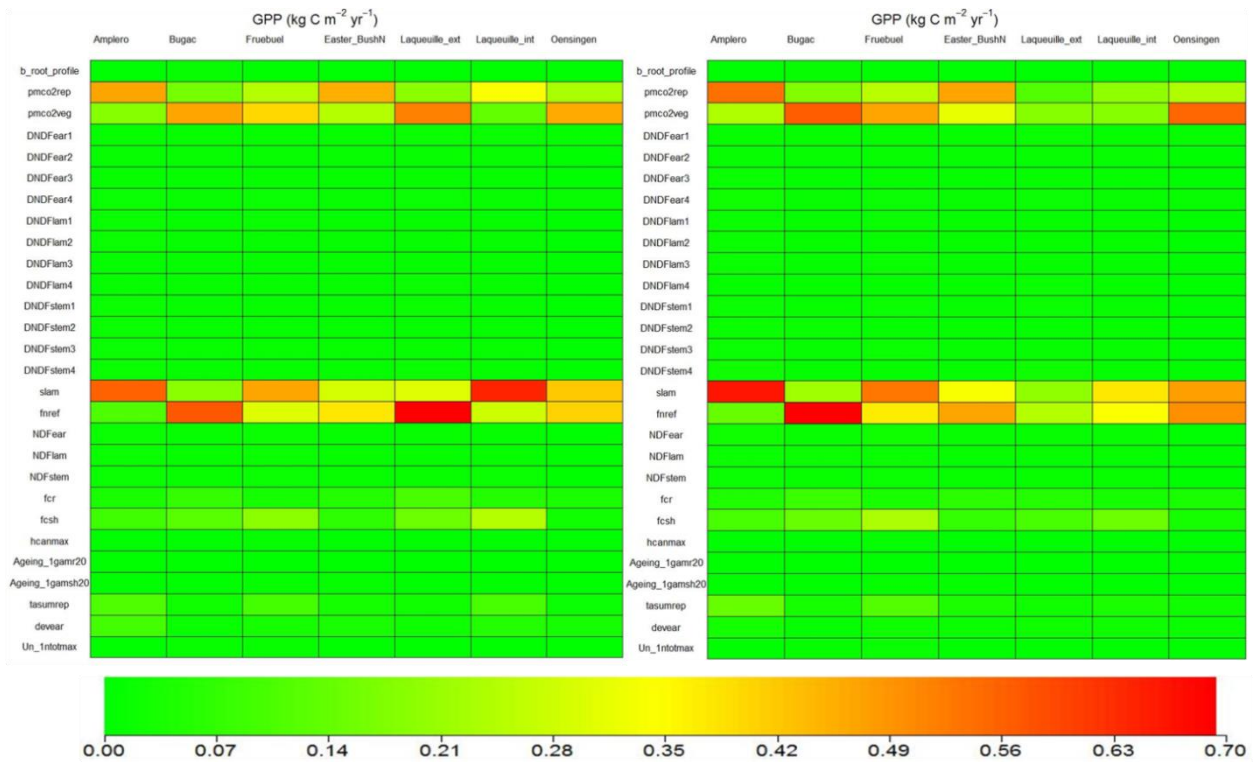


Figure 37. Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire brute (GPP) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres

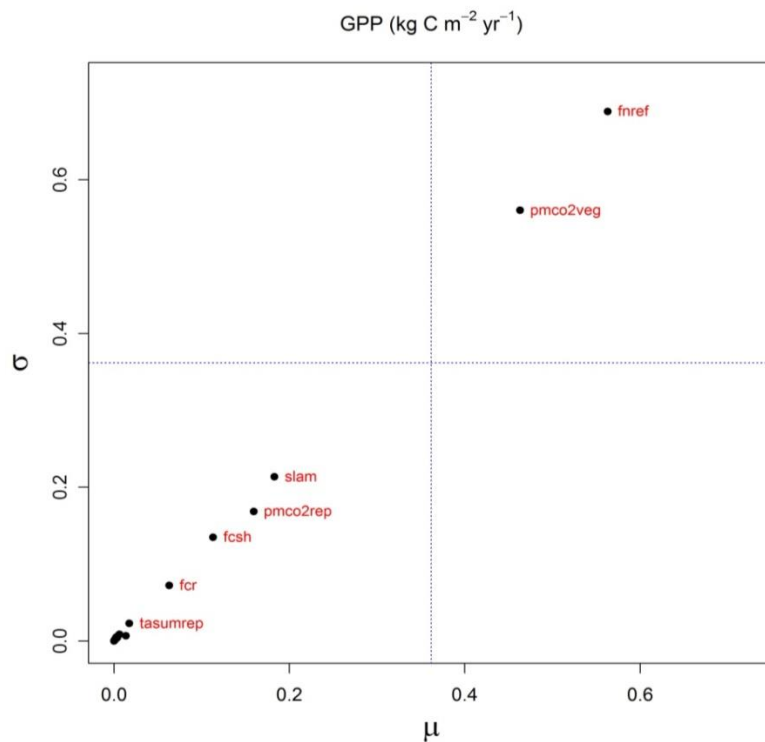


Figure 38. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Puszt en utilisant une distribution Gaussienne pour les valeurs des paramètres

Distribution uniforme

Comme dans l'exemple ci-dessus, les matrices de couleurs sont données (Figure 39) pour dresser le degré de sensibilité des estimations du GPP prenant Bugac-Pusztza (Hongrie) comme exemple. Les résultats de l'analyse de Morris montrent que le paramètre de la fraction de N de matière sèche structurale (fnref) donne la valeur la plus élevée que se soit pour μ ou σ . Ce paramètre, le seul placé dans la partie supérieure droite du schéma (Figure 40), est clairement séparé des paramètres restants.

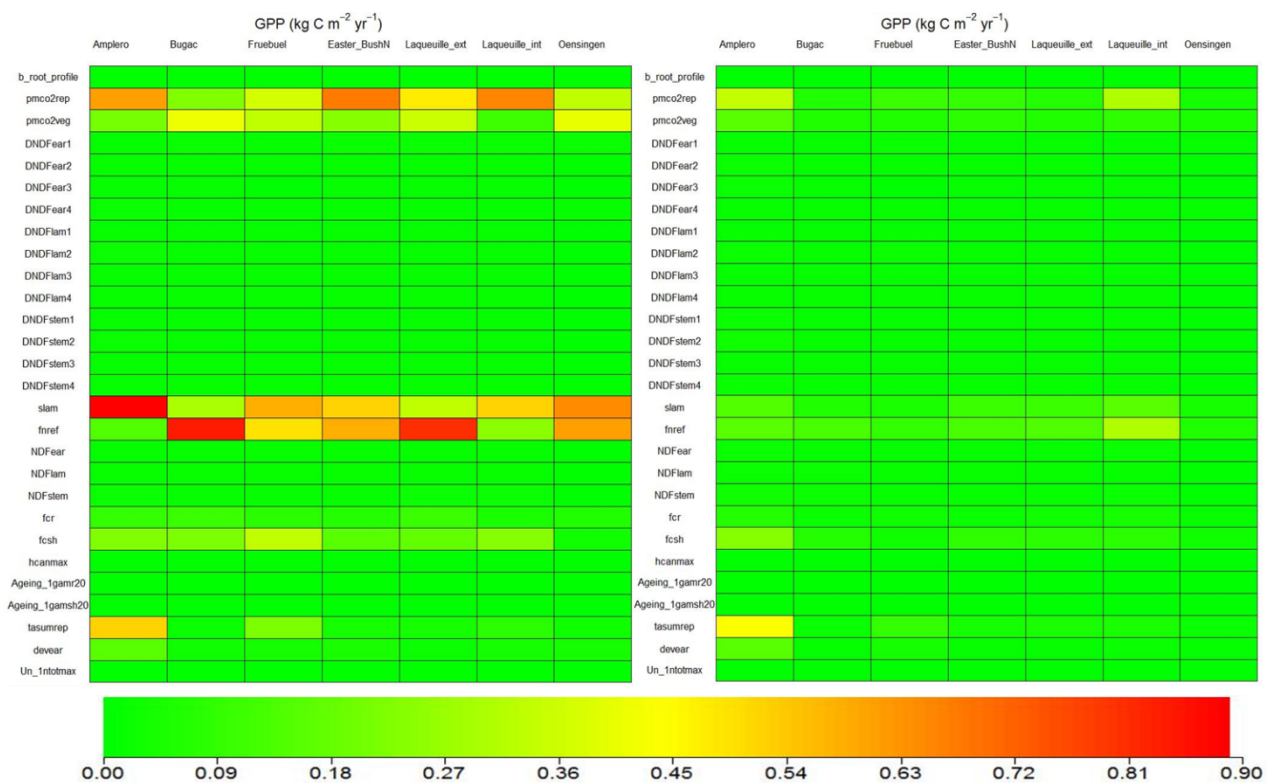


Figure 39. Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire brute (GPP) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres

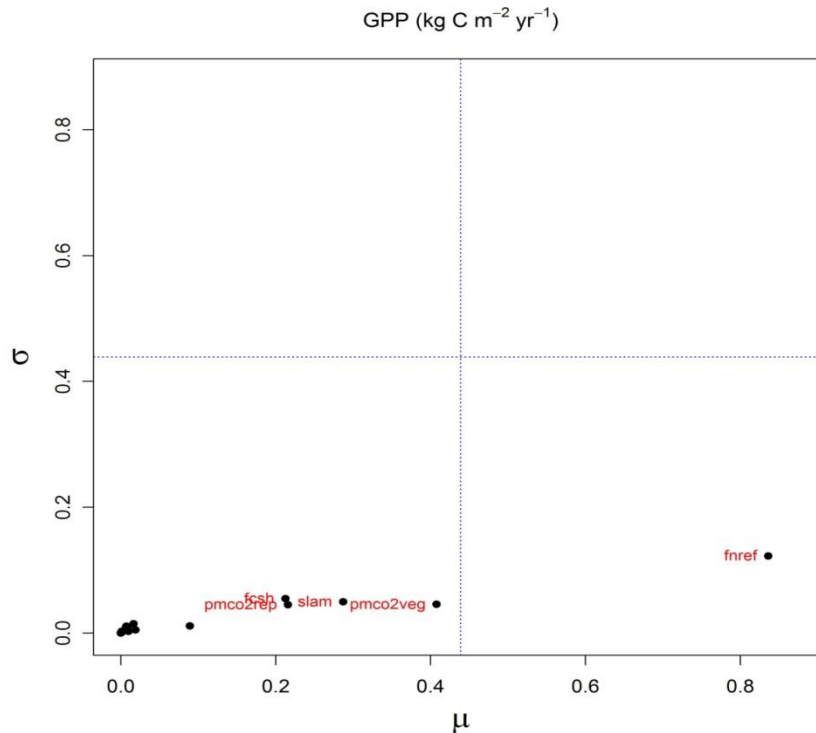


Figure 40. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Pusztas en utilisant une distribution uniforme pour les valeurs des paramètres

La sensibilité de PaSim s'est avérée assez stable à travers les sites européens, avec seulement quelques différences en termes de paramètres influents. Dans l'ensemble (Tableau 13 et Tableau 14), presque toutes les sorties cibles se sont montrées plus sensibles à sept paramètres du modèle : la surface spécifique foliaire maximale (slam), les taux de lumière saturée de la photosynthèse des feuilles pour les deux stades reproductif et végétatif (pmco2rep et pmco2veg, respectivement), la teneur en carbone dans la matière sèche structurale aérienne (fsh), le paramètre contrôlant la teneur en azote dans la matière sèche structurale (fnref), la phase de développement quand l'épiaison commence (devear), le facteur de normalisation pour le développement (tasumrep). Deux autres paramètres se sont révélés influents : les fractions de fibres dans les limbes (NDFlam) et dans les tiges/gaines (NDFstem) ingérées par les herbivores. Ces deux paramètres apparaissent pertinents seulement pour les émissions de CH₄ des animaux. Donc ces neuf paramètres représentent les paramètres les plus influents résultants de l'union des résultats de l'analyse de sensibilité des deux approches Gaussienne et uniforme. Le principe de l'union a été considéré plutôt que l'intersection afin d'élargir la base des paramètres et n'en exclure aucun.

Cependant, le classement des paramètres varie selon le site et c'est cette différence de « ranking » qui montre la plasticité du modèle, comme en témoigne la pauvre concordance entre les classements (en moyenne, autour de $TDCC = 0,3-0,4$) et une grande plasticité ($L < 0,5$) (Tableau 15). En moyenne, c'est avec la distribution uniforme que le modèle présente une plasticité légèrement plus élevée (0,155 contre 0,177 pour la distribution Gaussienne). Cela montre donc que PaSim est un modèle plastique capable de modifier son comportement suivant différentes conditions, ce qui est un atout d'un point de vue utilisateur.

Au-delà de cette conclusion générale, toutefois, les résultats montrent un schéma complexe de différences entre les variables de sortie que les différences dans les hypothèses de répartition ne peuvent pas seules les expliquer. Pour une meilleure compréhension de la sensibilité du modèle à ces paramètres, les résultats de l'analyse de sensibilité sont discutés ci-dessous en relation avec le rôle des paramètres influents (groupées en trois classes) dans le modèle.

Tableau 13. Classement des paramètres les plus influents obtenu avec la méthode Morris (μ) pour un ensemble de variables de sortie dans chaque site en utilisant une distribution Gaussienne des valeurs des paramètres

Sites	Variables de sortie								
	GPP	NPP	LAI	Yield	SWC	T _{soil}	C	CH ₄	NBP
Oensingen	slam pmco2rep fnref	slam pmco2veg fnref	slam	pmco2veg fesh pmco2rep	slam fnref	slam	pmco2veg fnref	-	slam pmco2veg
Friedbüel	slam pmco2veg fnref	slam pmco2rep fesh pmco2veg	slam	pmco2rep	slam fnref	slam	pmco2veg slam fnref	pmco2veg NDFlam NDFstem fnref	fnref slam
Laqueuille									
extensive	pmco2veg fnref	pmco2rep pmco2veg slam	slam fesh	-	fnref slam	slam	slam pmco2veg fnref	NDFstem NDFlam	tasumrep
intensive	slam pmco2rep fnref	slam pmco2rep fnref	slam	-	fnref slam	slam	fnref pmco2veg slam	NDFlam slam fnref NDFstem	tasumrep pmco2veg
Amplero	slam pmco2rep	slam pmco2rep	slam	pmco2rep	pmco2veg pmco2rep slam devar	slam pmco2rep	tasumrep fnref pmco2veg slam	×	slam pmco2veg pmco2rep
Easter Bush	pmco2rep fnref slam	pmco2rep slam fnref	slam	pmco2rep fesh	pmco2rep slam fnref fesh	slam pmco2rep	fnref pmco2veg slam	NDFlam NDFstem	pmco2veg pmco2rep
Bugac- Puszta	pmco2veg fnref	fnref pmco2veg	slam fesh	-	fnref pmco2veg slam	slam pmco2rep	fnref pmco2veg slam	NDFlam NDFstem	slam pmco2veg fnref

Tableau 14. Classement des paramètres les plus influents obtenu avec la méthode Morris (μ) pour un ensemble de variables de sortie dans chaque site en utilisant une distribution uniforme des valeurs des paramètres

Sites	Variables de sortie								
	GPP	NPP	LAI	Yield	SWC	T _{soil}	C	CH ₄	NBP
Oensingen	slam fnref pmco2veg	slam fnref pmco2rep	slam	fsh pmco2veg pmco2rep	slam fnref	slam	pmco2veg fnref slam	-	slam
Früebüel	slam fnref	slam fsh pmco2rep	slam	pmco2rep slam	slam fnref	slam fnref pmco2rep	slam fnref pmco2veg	NDFlam NDFstem	fnref pmco2rep
Laqueuille									
extensive	fnref pmco2rep	pmco2rep fnref slam	slam fsh	-	fnref slam pmco2rep	fnref	slam pmco2veg fnref	NDFstem NDFlam	pmco2rep tasumrep
intensive	pmco2rep slam	pmco2rep slam	slam pmco2rep	-	pmco2rep fnref slam	slam pmco2rep	fnref pmco2veg slam	fsh pmco2rep NDFstem	pmco2rep tasumrep
Amplero	slam pmco2rep tasumrep	slam pmco2rep tasumrep	slam tasumrep	pmco2rep slam tasumrep fnref	pmco2rep slam pmco2veg tasumrep	slam tasumrep	Slam fnref tasumrep pmco2veg	×	slam fsh pmco2rep tasumrep
Easter Bush	pmco2rep fnref slam	pmco2rep slam fnref	slam fsh	fsh pmco2rep	pmco2rep slam fsh	slam pmco2rep	pmco2veg fnref slam	NDFstem NDFlam fsh	pmco2rep
Bugac-Pusza	fnref pmco2veg	fnref pmco2veg	slam fsh	-	fnref pmco2veg slam	slam pmco2rep fsh	slam fnref pmco2veg	NDFstem NDFlam	slam fsh

Tableau 15. Coefficient de concordance (TDCC) et plasticité du modèle (L) pour chaque variable de sortie en utilisant une distribution Gaussienne et une distribution uniforme

Variables de sortie	Gaussian		Uniform	
	TDCC	L	TDCC	L
GPP	0.004	0.002	0.191	0.088
NPP	0.272	0.126	0.377	0.174
LAI	~1.000	0.463	0.885	0.409
Yield	0.119	0.055	0.268	0.124
SWC	0.352	0.163	0.335	0.155
T _{soil}	~1.000	0.463	0.449	0.208
C	0.159	0.074	0.299	0.138
CH ₄	0.379	0.175	0.136	0.063
NBP	0.155	0.072	0.079	0.036
Moyenne	0.382	0.177	0.335	0.155

slam, fnref, fcsh

PaSim s'est révélé très sensible à la valeur maximale de la surface foliaire spécifique (*slam*), utilisée pour calculer la surface foliaire du couvert. Le *slam* est un facteur éco-physiologique clé qui influe sur la physiologie de la feuille, la photosynthèse, et le gain de carbone de la plante entière. Grâce à la relation avec le substrat de carbone dans la matière sèche structurelle (Annexe III.C), le modèle tient compte des variations saisonnières de la surface foliaire spécifique qui pourrait résulter de la sécheresse saisonnière ou de la saisonnalité dans la production de feuilles, ou du rayonnement solaire. Toutefois, le site le plus aride de cette étude (Bugac-Puszt, Hongrie) indique que le *slam* n'est pas parmi les paramètres les plus influents sur les sorties de GPP et NPP. Après la disponibilité de l'eau, c'est l'azote qui est en effet le principal facteur qui peut entraver la productivité dans les systèmes arides. Le stress hydrique auquel les prairies situées dans des zones arides et semi-arides doivent faire face poussera donc à allouer plus d'azote pour soutenir la croissance, ce qui peut expliquer la sensibilité élevée à un paramètre lié à l'azote (*fnref*, Annexe III.F) dans de telles conditions.

Sous l'hypothèse de distribution uniforme (Tableau 14), le paramètre *slam* n'est pas pertinent pour les sites humides de Easter Bush (Royaume-Uni) et Oensingen (Suisse). Il semble que la teneur en carbone dans la matière sèche structurelle aérienne (*fcsh*, Annexe III.E) peut devenir un paramètre de conduite dans ces conditions. Les variables de sortie du sol (teneur en eau et température) sont généralement, sensibles à *slam* et *fnref*, en résultat d'échange entre la végétation et la masse du sol et les budgets énergétique. Le microclimat de la surface du sol, caractérisé par la température du sol et les conditions de l'eau, peut en effet être modifié par le développement de la plante et les conditions de sa croissance. Par exemple, la transpiration de la plante entraîne une extraction d'eau de chaque couche du sol, ce qui influence la conductivité thermique car celle-ci dépend de la teneur en eau du sol.

pmco2rep, pmco2veg, devear, tasumrep

Le taux de la photosynthèse du couvert végétal est contrôlé dans PaSim par divers processus physiologiques, y compris l'effet d'interaction de la température, de l'azote et du CO₂ (Annexe III.D). Le développement des plantes (Annexe III.A) et les paramètres liés à la phénologie (Annexe III.B) modulent les effets de la phénologie de la végétation sur la

photosynthèse, avec des conséquences importantes pour le bilan de carbone des prairies. Cela explique pourquoi $pmco2rep$ et/ou $pmco2veg$, qui dépendent du stade de développement de la plante, ont une influence sur pratiquement toutes les variables de sortie, à l'exception de l'indice de surface foliaire LAI (si l'on fait l'exception de la prairie intensive de Laqueuille avec une distribution uniforme, Tableau 14).

Par contre, les variations du paramètre relatif au stade de développement auquel l'épiaison commence ($devear$) ont peu ou pas d'influence sur les résultats du modèle. L'influence de $devear$ n'a été observée que sur la teneur en eau du sol à Amplero avec distribution Gaussienne (Tableau 13). C'est en effet un paramètre clé pour assimiler la répartition entre les pousses et les racines ainsi que les sous-compartiments de pousses (Annexe III.B). La variation du facteur de normalisation utilisé par le modèle pour calculer l'indice de stade de développement ($tasumrep$) est influente principalement sur deux sorties, la production nette du biome (NBP) et la concentration de substrat de carbone dans la plante (C), qui sont directement liés au budget de carbone. À Amplero, $tasumrep$ présentait une influence (dont l'intensité varie) sur toutes les sorties de ce site (ex. Figure 41).

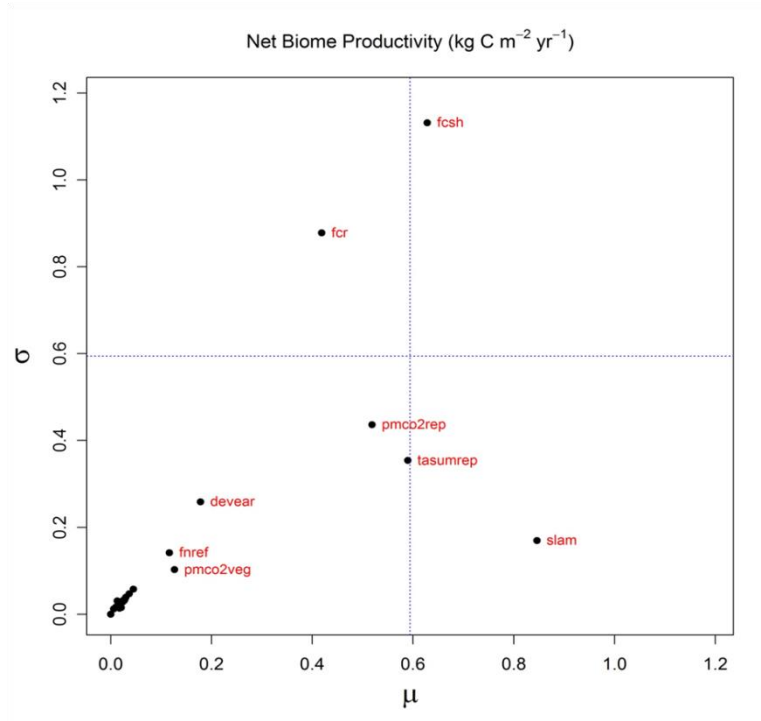


Figure 41. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production nette de biome (NBP) pour le site Amplero en utilisant une distribution uniforme pour les valeurs des paramètres

Il est à noter que pmco2veg uniquement (et non pmco2rep) est pertinent dans tous les cas pour la concentration du substrat de carbone dans la plante (C). Ceci peut s'expliquer par la réduction de la capacité photosynthétique des nouvelles feuilles qui se produisent dans la phase végétative ($pmco2veg < pmco2rep$). Dans le stade de la reproduction, en revanche, l'extension de la tige de la fleur conserve les feuilles en développement soumis à une forte irradiation près du sommet de la canopée ce qui rendrait le processus moins sensible au paramètre pertinent.

NDFlam, NDFstem

Concernant les émissions de CH_4 de la part des animaux, les paramètres liés à la teneur en fibres dans les feuilles et les tiges (Annexe III.G) logiquement ont tendance à avoir plus d'influence que d'autres. Pour NBP, l'analyse de sensibilité a montré une valeur élevée de σ pour la fraction de fibres dans les tiges et les gaines d'ingestion (NDFstem) à Laqueuille (gestion intensive) avec la distribution Gaussienne (Figure 42).

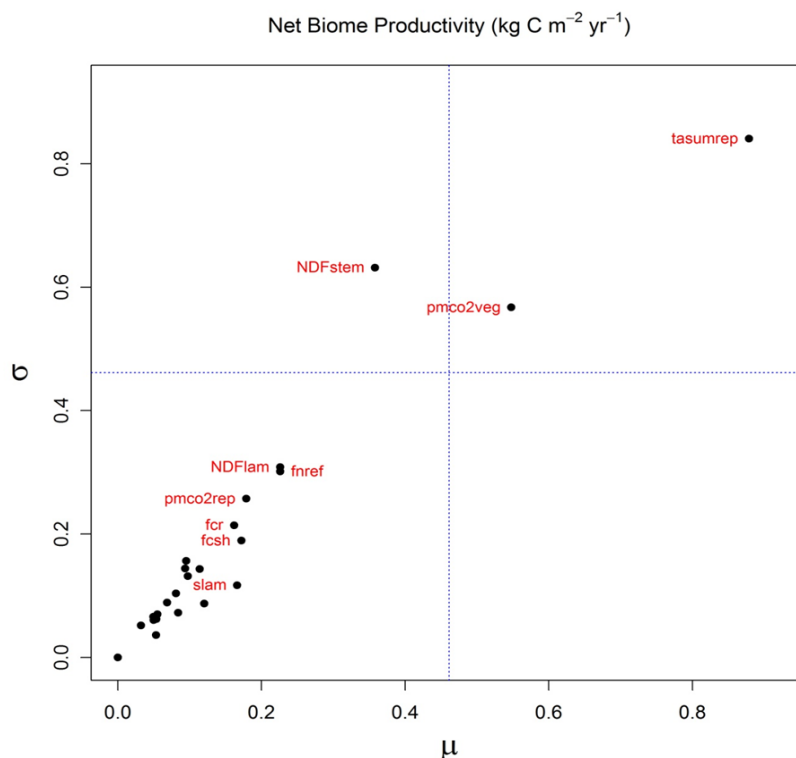


Figure 42. La moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production nette de biome (NBP) pour le site Laqueuille (gestion intensive) en utilisant une distribution Gaussienne pour les valeurs des paramètres

Enfin, nous avons constaté une proportionnalité entre la moyenne (μ) et l'écart-type (σ) dans le cas d'une distribution Gaussienne des paramètres (ex. Figure 43), ce qui montre une certaine stabilité des résultats du modèle. Contrairement, pour la distribution uniforme les valeurs de l'écart-type (σ) étaient faibles pratiquement dans tous les cas, ce qui signifie qu'il y a des faibles interactions et non-linéarités entre les différents paramètres.

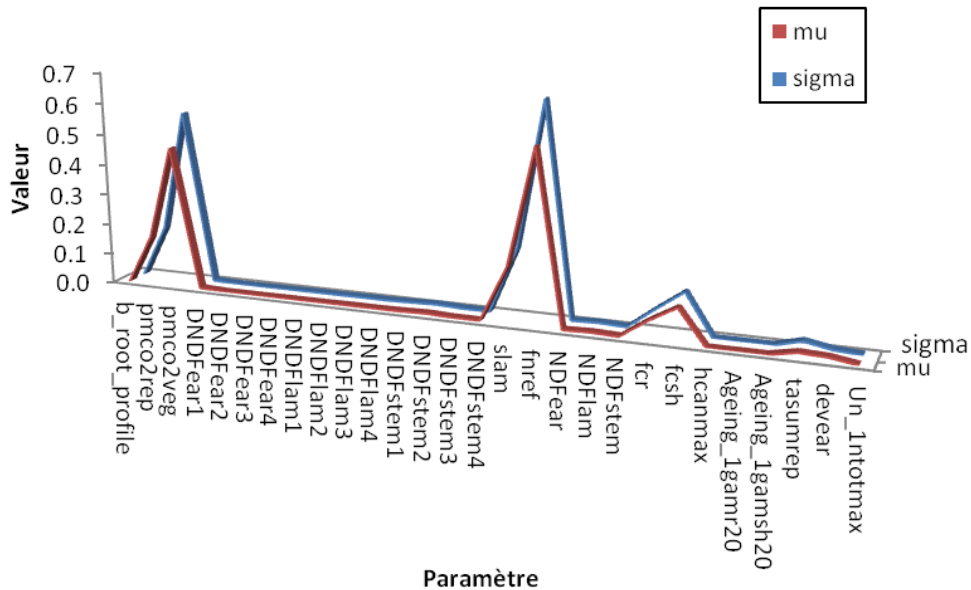


Figure 43. Proportionnalité entre les deux indices de sensibilité la moyenne (μ) et l'écart-type (σ) de l'effet élémentaire associé à la production primaire brute (GPP) pour le site Bugac-Pusza en utilisant une distribution Gaussienne pour les valeurs des paramètres

4.3. Calibration Bayésienne multi-sites à l'échelle de l'Europe

4.3.1. Matériels et méthodes

Après une étape d'analyse de sensibilité qui s'est avérée indispensable pour l'identification des paramètres auxquels le modèle est plus sensible, les neuf paramètres clés (Tableau 16) ayant un impact majeur sur les sorties ont fait l'objet d'une calibration Bayésienne. Par le biais de cette calibration, on a visé à obtenir un paramétrage commun du modèle à l'échelle de l'Europe sur des sites non manipulés et avec des données à long terme. Pour cela nous avons utilisé les six sites considérés dans l'analyse de sensibilité (Tableau 10).

La différence de cette étude avec celle vue dans le chapitre précédent est que cette calibration a été réalisée sur plusieurs sites en même temps, couvrant divers zones climatiques et géographiques, utilisant un nombre plus important d'années d'observation. Nous avons également utilisé deux types de distribution a priori des paramètres (Gaussienne et uniforme) afin d'évaluer la stabilité du modèle aux hypothèses relatives à la distribution des valeurs des paramètres.

Dans notre protocole, élaboré spécialement pour cette étude et relayé aux collègues en Europe, nous avons décidé de garder des sites (Figure 44) pour la validation indépendante du nouveau paramétrage. Ces sites que nous allons présenter dans la section suivante n'ont pas été inclus dans la calibration.

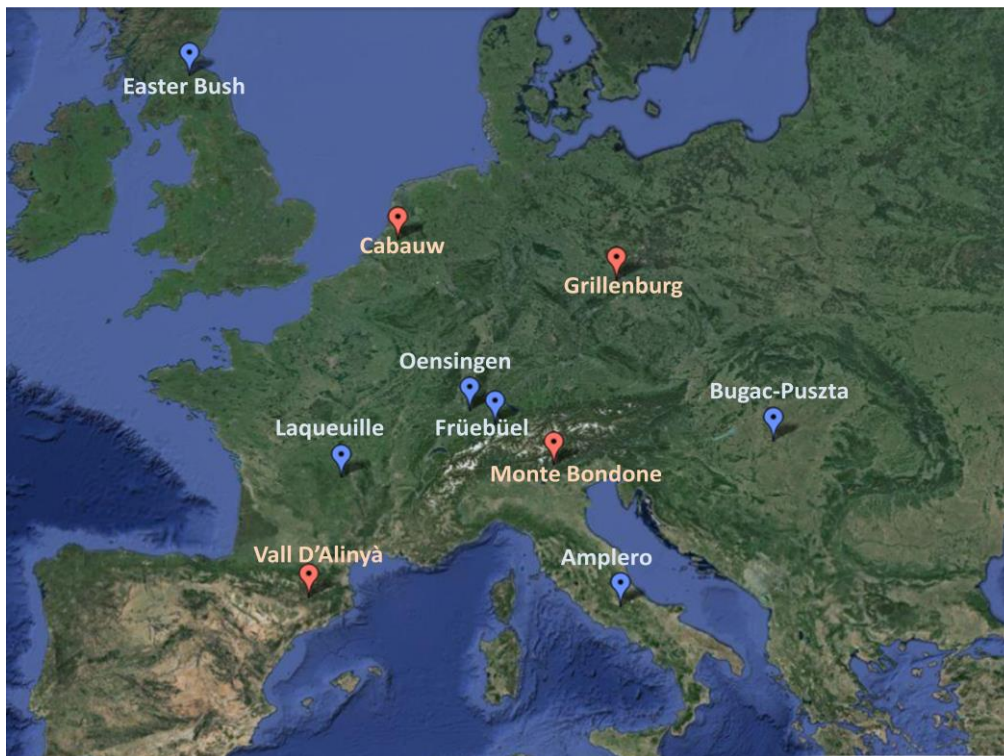


Figure 44. Localisation des sites prairiaux utilisés pour la calibration (en bleu) et la validation (en rouge)

Tableau 16. Paramètres utilisés pour la calibration Bayésienne multi-sites

Paramètre	Définition	Unité	Valeur par défaut
slam	Surface foliaire spécifique maximale	m ² kg ⁻¹	28
pmco2rep	Capacité photosynthétique foliaire pendant la période reproductrice	μmol m ⁻² s ⁻¹	18.75
pmco2veg	Capacité photosynthétique foliaire pendant la période végétative	μmol m ⁻² s ⁻¹	12.50
fcsh	Teneur en C dans la matière sèche structurelle aérienne	kg C kg ⁻¹ DM	0.43
fnref	Paramètre contrôlant la teneur en N dans la matière sèche structurelle	-	0.017
devear	Valeur de l'indice de développement	-	0.545
Tasumrep	Facteur de normalisation de l'indice de développement	K-d	708.8
NDFlam	Teneur en fibres dans les limbes ingérés par les herbivores	-	0.52
NDFstem	Teneur en fibres dans les tiges et gaines ingérées par les herbivores	-	0.60

Le Tableau 17 présente les variables de sortie utilisées lors de cette étude afin de mettre à jour la distribution a priori des paramètres. Selon leur disponibilité par site et par année, cinq variables de sortie ont été considérées pour la calibration Bayésienne multi-sites, dont deux liées au sol (teneur en eau et température), deux liées à la végétation (biomasse et indice de surface foliaire) et une variable liée au flux de gaz à effet de serre (échange net de l'écosystème en carbone). Pour le choix des données à mobiliser dans la calibration, on s'est basé d'abord sur leur disponibilité, leur rapport qualité/quantité et leur diversité (sol, végétation, atmosphère). Ces données viennent en grande partie de la base de données de projets européens, à partir du recensement auprès des partenaires de FLUXNET (<http://www.europe-fluxdata.eu>), qui ont également fournis directement une partie des données utilisées. Certains traitements sur les données ont été nécessaires pour les adapter au bon format requis par PaSim et par le code de la calibration Bayésienne.

Tableau 17. Variables de sortie utilisées dans la calibration Bayésienne multi-sites

Variable de sortie	Définition	Unité
NEE	Échange net de l'écosystème en carbone	$\text{g C m}^{-2} \text{d}^{-1}$
SWC	Teneur en eau du sol	$\text{m}^3 \text{m}^{-3}$
T_{soil}	Température du sol	K
Biomasse aérienne	Rendement de la matière sèche aérienne (avant ou après fauche selon le site)	g m^{-2}
LAI	Indice de surface foliaire	$\text{m}^2 \text{m}^{-2}$

Pendant cette calibration avec des données de long terme, les simulations ont tournées en même temps sur l'ensemble des sites afin d'aboutir sur une seule distribution a posteriori des paramètres pour chaque hypothèse relative à la distribution a priori de leur valeurs (Gaussienne et uniforme).

4.3.2. Résultats et discussion

Dans cette étude, le temps de calcul de la calibration Bayésienne a pris à peu près 35 jours sur le cluster du LIMOS. Par rapport à l'étude précédente, nous avons certes plus de sites et plus d'années, mais moins de paramètres à calibrer donc moins de tirages et moins de tests d'acceptation. En plus, les neuf paramètres calibrés ne sont pas spécifiques aux sites, ce qui a permis d'alléger le processus de calibration puisque nous n'avons pas eu à réécrire à chaque fois les fichiers des conditions initiales et spécifiques aux sites nécessaires en entrée pour faire tourner le modèle.

L'a posteriori des paramètres

Le taux d'acceptation des paramètres candidats dans le processus MCMC a été de ~35% dans le cas de la distribution uniforme des paramètres et de 30% avec la distribution Gaussienne. Cela montre une certaine stabilité de la méthode de la calibration Bayésienne indépendamment du type de distribution a priori utilisée. Cette légère différence entre les deux taux peut s'expliquer par la variation des valeurs des paramètres tirées d'une distribution Gaussienne, qui peut générer plus de rejet de valeurs candidats. Le Tableau 18

dresse la liste des nouvelles valeurs des paramètres obtenues après calibration Bayésienne pour les deux distributions utilisées.

Tableau 18. Nouveau paramétrage européen de PaSim obtenu après calibration Bayésienne multi-sites en utilisant deux distributions des paramètres (uniforme et Gaussienne)

Paramètre	Unité	Valeur (distribution uniforme)	Valeur (distribution Gaussienne)
pmco2rep	$\mu\text{mol m}^{-2} \text{s}^{-1}$	22.086	19.075
pmco2veg	$\mu\text{mol m}^{-2} \text{s}^{-1}$	14.550	12.449
slam	$\text{m}^2 \text{kg}^{-1}$	31.732	27.512
fnref	-	0.020	0.016
NDFlam	-	0.545	0.518
NDFstem	-	0.631	0.600
fcsh	$\text{kg C kg}^{-1} \text{DM}$	0.463	0.430
tasumrep	K-d	741.655	682.084
Devear	-	0.609	0.549

Tableau 19. Exemple de distributions du prior et du posterior des paramètres à travers la moyenne, le CV et les valeurs du changement du CV, coefficient de variation obtenue après calibration Bayésienne multi-sites utilisant le nouveau paramétrage (distribution uniforme)

Paramètre du modèle	Prior		Posterior		% changement CV
	Moyenne	CV	Moyenne	CV	
pmco2rep	18.53	0.25	18.66	0.07	-70.28
pmco2veg	12.04	0.10	12.49	0.07	-28.41
slam	26.84	0.12	28.04	0.09	-29.86
fnref	0.02	0.12	0.02	0.09	-28.50
NDFlam	0.51	0.10	0.52	0.04	-57.34
NDFstem	0.58	0.10	0.60	0.04	-59.68
fcsh	0.42	0.10	0.43	0.04	-56.39
tasumrep	683.75	0.11	708.82	0.06	-46.65
devear	0.53	0.12	0.55	0.06	-52.20

Le Tableau 6 montre les valeurs de la moyenne et du CV des paramètres calculées pour le prior et le posterior, obtenues en utilisant une distribution uniforme pour le prior. On peut noter que l'incertitude du slam, qui est l'un des paramètres les plus influents selon l'analyse de sensibilité, a été réduite de ~30%. L'incertitude du paramètre pmco2rep lié à la photosynthèse pendant le stade de reproduction de la plante a été réduite de 70% (c'est un paramètre qui apparaît influent sur une majorité des sites et des variables de sorties, d'après l'analyse de sensibilité). Avec la distribution uniforme, la réduction globale d'incertitude apportée par la distribution a posteriori a été de l'ordre de ~48% par rapport à la distribution a priori pour l'ensemble des paramètres.

Analyse du posteriori des variables de sortie

Nous avons utilisé le nouveau paramétrage obtenu à travers la calibration Bayésienne pour simuler les variables de sortie étudiées. Dans ce qui suit nous allons présenter quelques résultats à titre d'exemple pour chaque variable de sortie (Tableau 17) dans différents sites plus une variable de validation, le méthane (CH₄, g CH₄ C m⁻² d⁻¹) entérique émis par les animaux au pâturage, mesuré sur le site de Laqueuille géré intensivement et non utilisé pendant la calibration.

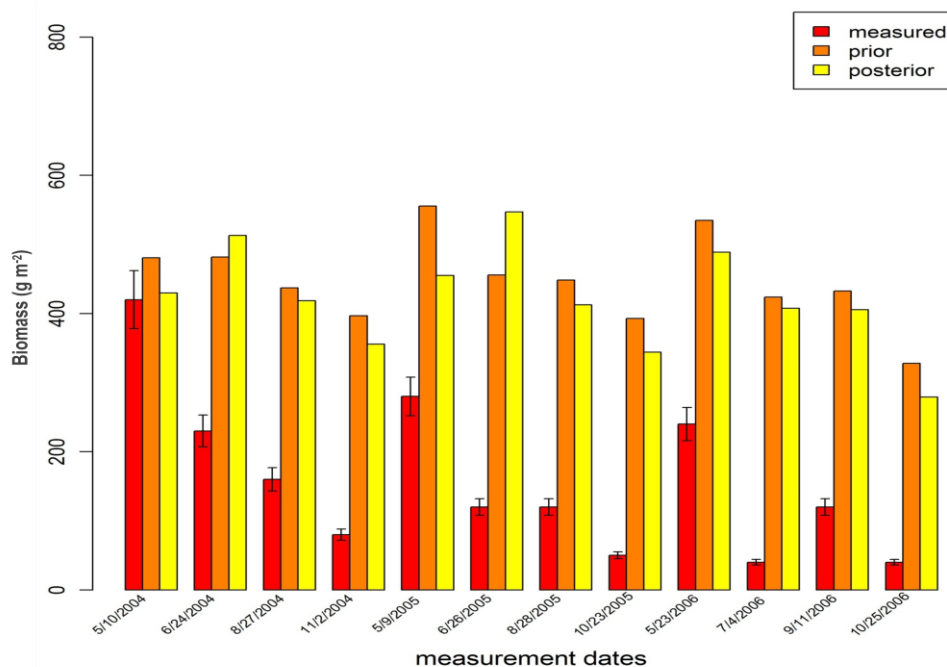


Figure 45. Comparaison des valeurs mesurées (avec les barres d'écart-type), du prior et du posterior de la biomasse aérienne au site suisse d'Oensingen en utilisant une distribution uniforme

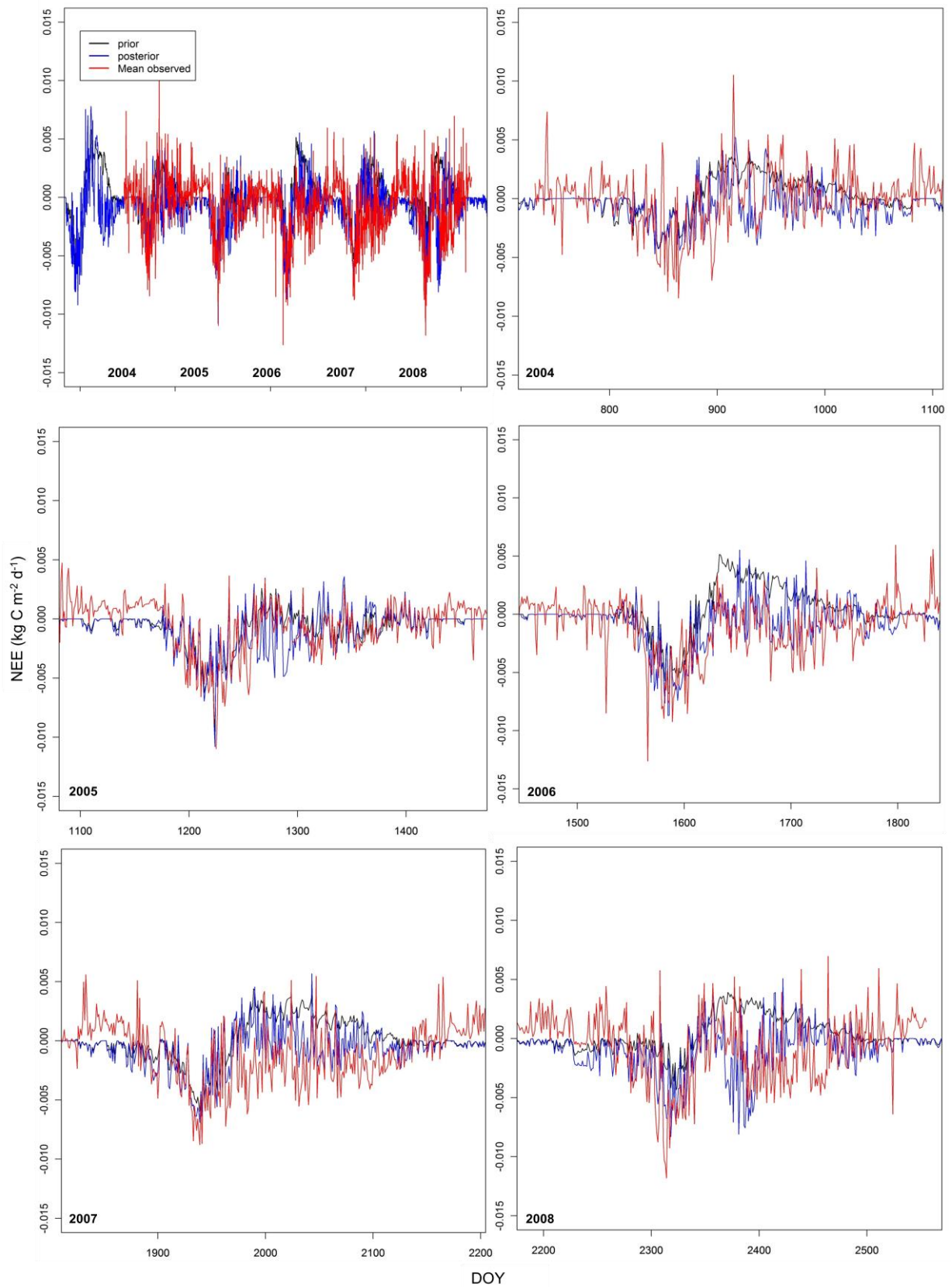


Figure 46. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes pour le site français de Laqueuille géré intensivement en utilisant une distribution uniforme

L'étude de la biomasse aérienne sur le site suisse d'Oensingen, par exemple, a montré (Figure 45), à l'exception de deux dates sur douze, une amélioration du résultat du posterior par rapport au prior. Malgré des valeurs a priori assez élevées par rapport aux observations, le posterior a permis de réduire cette surestimation du prior.

Pour l'échange net de l'écosystème en carbone (NEE, Figure 46) relatif au site intensif de Laqueuille, comme nous avons des données journalières sur plusieurs années, il est difficile d'analyser graphiquement les résultats sur toute la période. C'est pourquoi nous avons montré le résultat sur toute la période d'abord, puis nous l'avons découpé la période année par année, ce qui fait une sorte de zoom sur le premier graphique en haut à gauche. Cela nous permet de constater l'amélioration des simulations apportée par le posteriori surtout sur les trois dernières années 2006, 2007 et 2008. Cette amélioration du résultat de la NEE, qui est une sortie liée aux émissions de gaz à effet de serre et dont le fait de la simuler correctement est important, se confirme à travers l'analyse des régressions linéaires de la NEE à Laqueuille (Figure 47) pour les deux types de gestion extensive (à gauche) et intensive (à droite). Nous remarquons que dans les deux cas de gestion, la valeur du coefficient de détermination (R^2) de la régression linéaire du posterior a doublé par rapport au prior, sachant que $0 \leq R^2 \leq 1$ ($R^2 = 1$ représentant l'optimum de manière que les points sont exactement alignés sur la droite de régression), et les valeurs de la pente et de l'ordonnée sont respectivement plus proches de 1 et 0 (valeurs optimales).

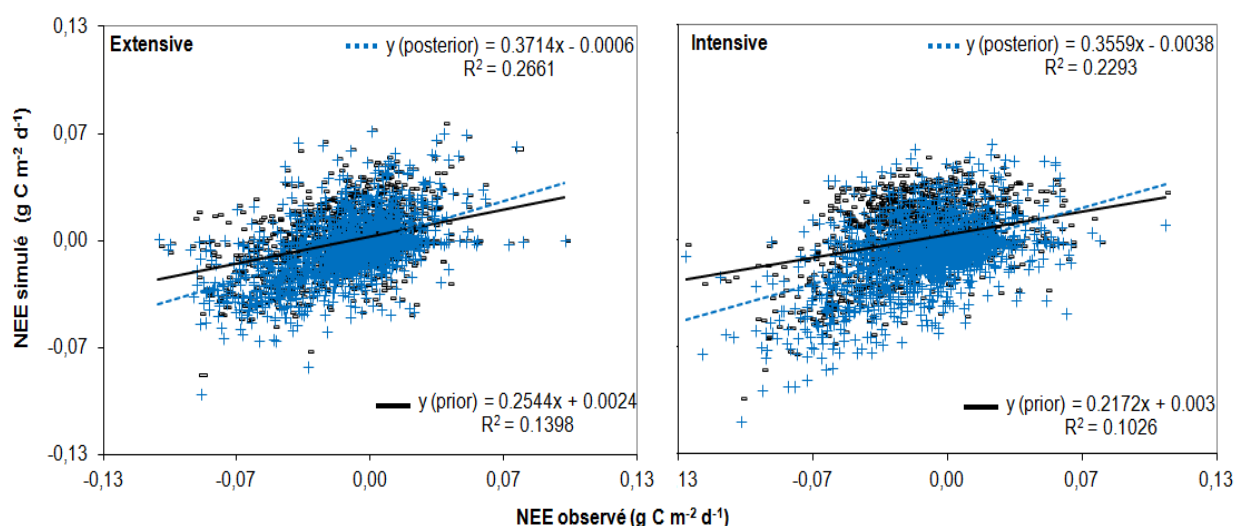


Figure 47. Régression linéaire des valeurs simulées de la NEE ($\text{g C m}^{-2} \text{d}^{-1}$) (prior [-] et posterior [+]) versus les valeurs observées à Laqueuille (2004-2008) avec une gestion extensive (gauche) et intensive (droite), en utilisant une distribution uniforme

En ce qui concerne les deux variables de sortie du sol (la teneur en eau et la température), l'exemple de Laqueuille avec gestion extensive (Figure 48) montre que le posterior est très légèrement mieux que le prior, surtout pour la température mais la différence n'est pas considérable, tellement les simulations du posterior et du prior sont très proches qu'on ne peut pas en tirer une conclusion définitive.

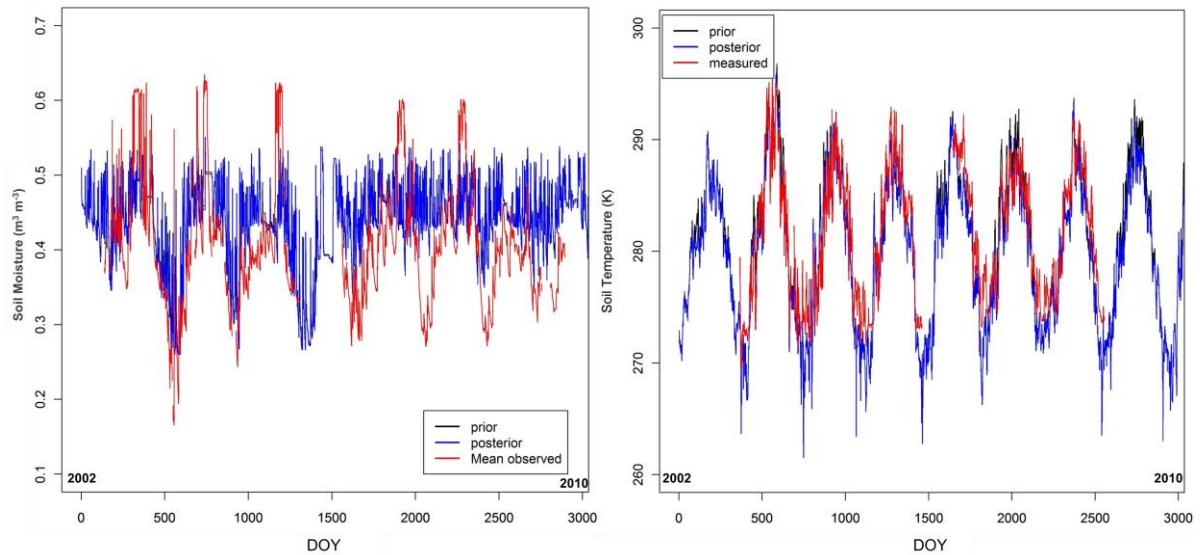


Figure 48. Comparaison des valeurs mesurées, du prior et du posterior de la teneur en eau du sol (à gauche) et de la température du sol (à droite) à Laqueuille avec une gestion extensive (2003-2009), en utilisant une distribution uniforme

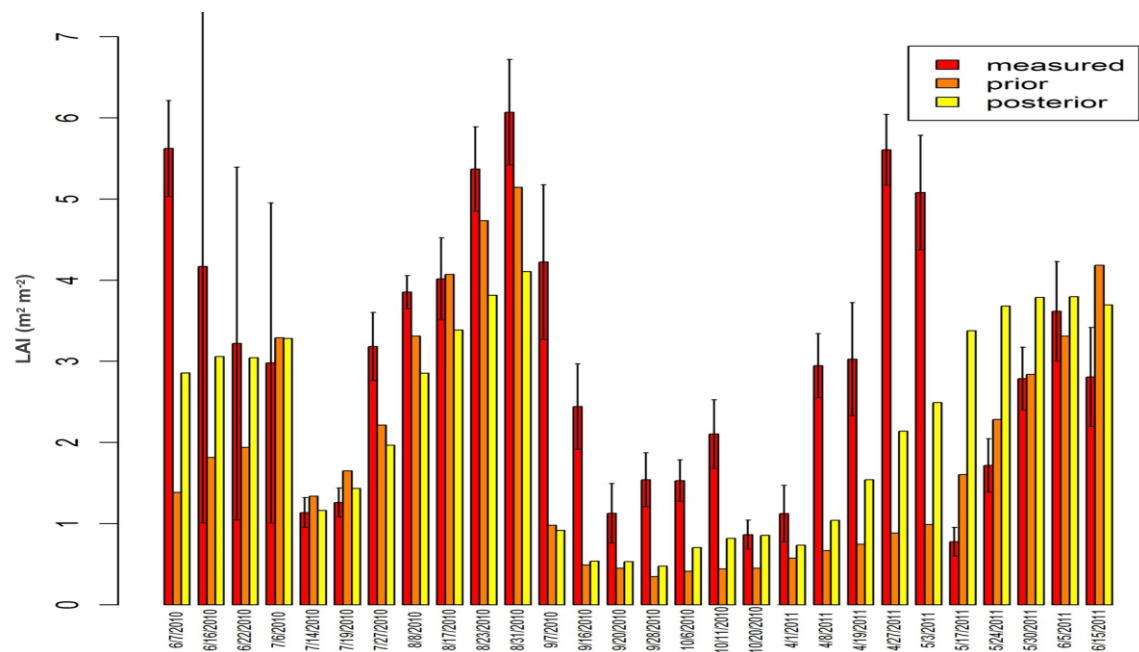


Figure 49. Estimations du prior et du posterior de l'indice de surface foliaire au site suisse de Frübüel en utilisant une distribution uniforme

La Figure 49 montre les estimations du prior et du posterior de l'indice de surface foliaire (LAI) au site suisse de Frübüel. Nous pouvons remarquer que les simulations du posterior illustrent globalement une amélioration des résultats par rapport au prior sur la majorité des dates observées malgré les grandes incertitudes liées aux données mesurées (visibles à travers les valeurs élevées des écart-types).

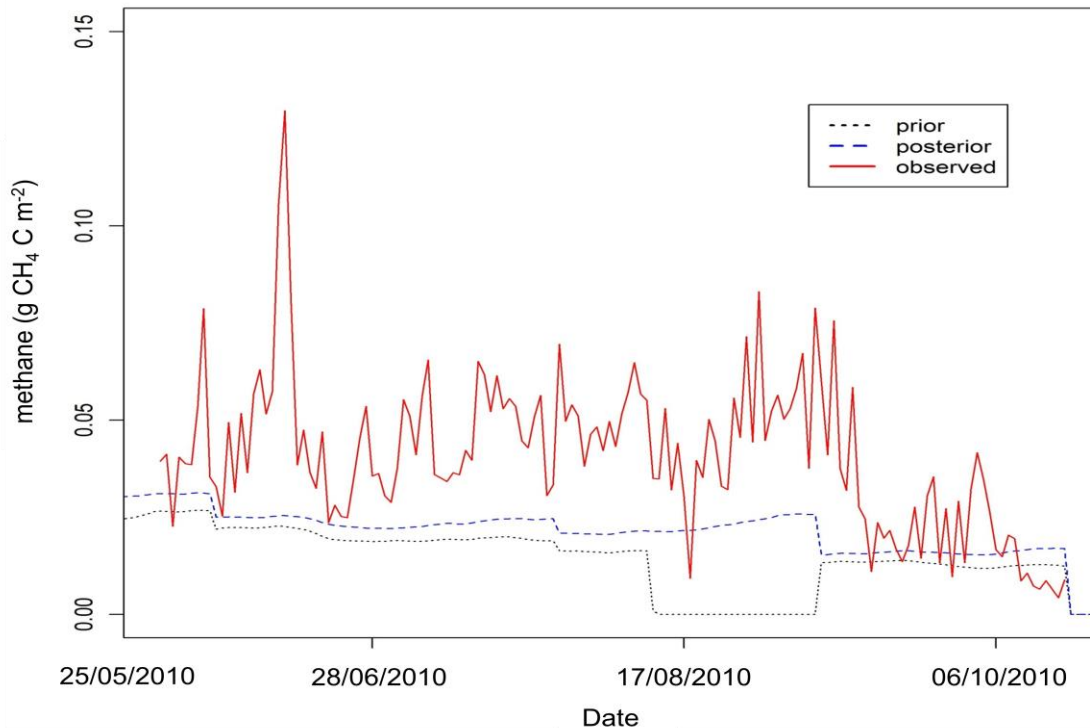


Figure 50. CH_4 ($\text{g CH}_4 \text{ C m}^{-2} \text{ d}^{-1}$) entérique simulé (prior et posterior) et observé à Laqueuille avec gestion intensive pour la période entre mai et octobre 2010 en utilisant une distribution uniforme

Enfin, voici les simulations du méthane entérique (CH_4) à Laqueuille avec une gestion intensive pour la période entre mai et octobre 2010 (Figure 50). C'est une variable de validation, puisque le méthane n'a pas été utilisé dans la calibration Bayésienne multi-sites, ni l'année 2010 d'ailleurs puisque l'accès à ces données n'a pas été possible qu'à la fin de notre étude. La Figure 50 montre l'amélioration obtenue avec le nouveau paramétrage a posteriori, mais aussi que le modèle ne simule pas correctement les fluctuations journalières des valeurs de CH_4 . Cela est également dû au dispositif de mesure qui est fixe alors que les vaches pâturent bougent et changent de positions régulièrement dans la parcelle. Il est à noter que, avec la simulation du posterior, le système émet du CH_4 entérique en été parce que la biomasse de l'herbe est disponible et le pâturage peut se produire. Cela se rapproche de ce qui se passe dans la réalité, ce qui n'est pas le cas avec la simulation du prior qui affiche zéro émission au mois d'août 2010.

4.4. Validation du nouveau paramétrage de PaSim

Nous avons montré précédemment que la calibration Bayésienne est capable de réduire les incertitudes sur les sorties du modèle liées aux paramètres et d'améliorer ainsi les résultats des simulations sur les sites déjà utilisés pour la calibration. Afin d'évaluer le nouveau paramétrage européen, il est également important de garder certains sites pour la validation qui n'ont pas fait partie de l'étude de calibration.

La validation du modèle est réalisée en aval de la calibration. Les prédictions sont confrontées aux mesures indépendantes observées afin de juger la performance du modèle et sa qualité de prédiction. Les données pour la validation doivent donc être indépendantes des données utilisées pour calibrer le modèle.

Pour cela quatre sites ont été choisis (Tableau 20), diversifiés géographiquement (Figure 44) et climatiquement (Figure 51), tel que montré par deux indices bioclimatiques (les mêmes utilisés pour classer les sites de calibration). Les sites de validations varient du sous humide (deux sites) jusqu'au humide (deux sites) avec le site sous humide de Vall D'Alinyà (Espagne) ayant un bilan hydrique climatique à l'équilibre.

Tableau 20. Informations et climats des sites de validation

Pays	Site	ID	Années	Coordonnées géographiques			Climat		
				Latitude	Longitude	Altitude (m a.s.l.)	Température moyenne de l'air (°C)	Précipitation totale (mm an ⁻¹)	Evapotranspiration de référence (mm an ⁻¹)
Espagne	Vall D'Alinyà	VDA	2004-2008	42°15'N	01°44'E	1770	6	908	889
Allemagne	Grillenburg	Gri	2004-2008	50°94'N	13°51'E	375	8	946	682
Italie	Monte Bondone	MBo	2003-2007	41°01'N	11°04'E	1550	5	1003	694
Pays-Bas	Cabauw	Ca	2004-2007	51°97'N	04°92'W	0.01	11	823	718

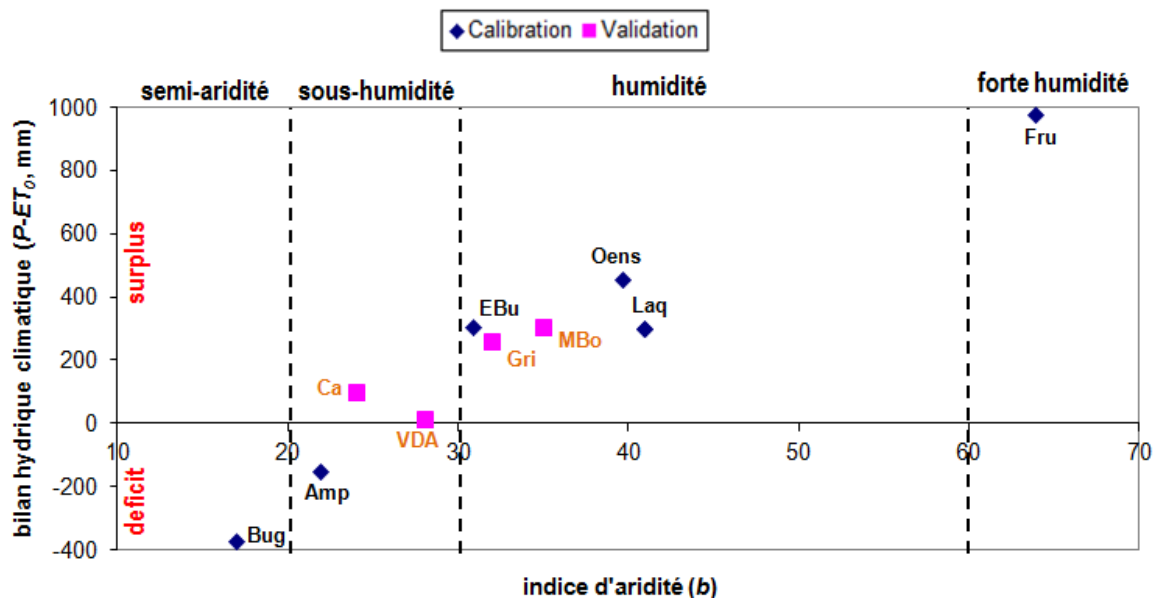


Figure 51. Classification des sites de calibration (en bleu) et des sites de validation (en rouge) selon deux indices bioclimatiques (l'indice d'aridité et le bilan hydrique climatique)

Nous allons montrer, dans ce qui suit, quelques exemples de résultats obtenus sur différents sites et variables de sortie en utilisant le nouveau paramétrage européen obtenu avec les deux distributions uniforme et Gaussienne. La Figure 52 présente le résultat de l'analyse d'incertitude des estimations du prior et du posterior de la biomasse aérienne au site allemand de Grillenburg en utilisant une distribution uniforme, où nous pouvons clairement observer qu'il y a eu une amélioration du résultat des simulations en utilisant le nouveau paramétrage par rapport à l'ancien, mais aussi une réduction de l'incertitude qui se traduit par des barres d'écart-type plus petites.

Pour la production primaire brute (GPP) au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme (Figure 53), nous pouvons constater une légère amélioration des simulations entre le posterior et le prior où le posterior est légèrement plus proche des observations.

La Figure 54 montre les estimations du prior et du posterior de l'échange net de l'écosystème en carbone (NEE) de 2004 à 2008 au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme. Il y a une similitude entre les deux courbes du posterior et du prior avec une légère amélioration du posterior, mais malgré cela les valeurs simulées par le modèle restent relativement élevées par rapport aux observations.

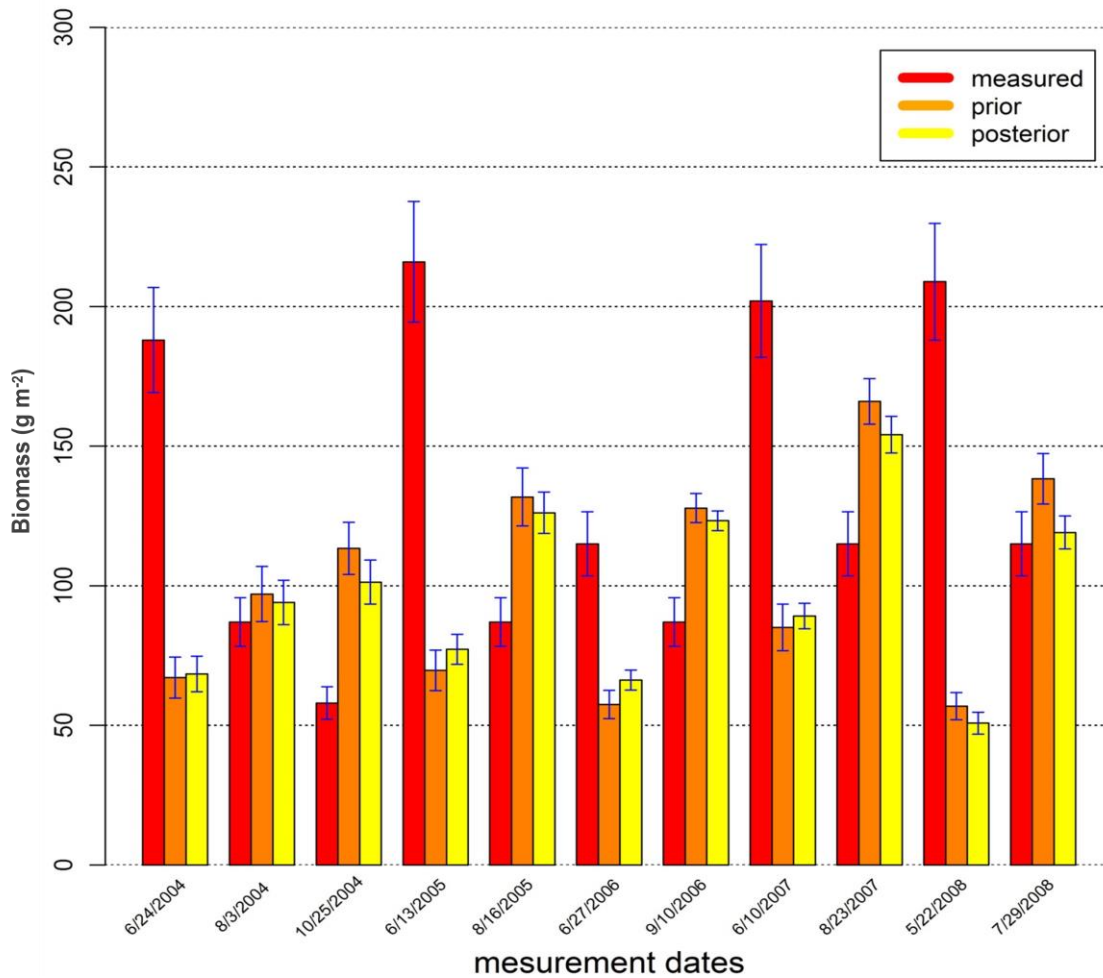


Figure 52. Analyse d'incertitude des estimations du prior et du posterior de la biomasse aérienne au site Allemand de Grillenburg en utilisant une distribution uniforme

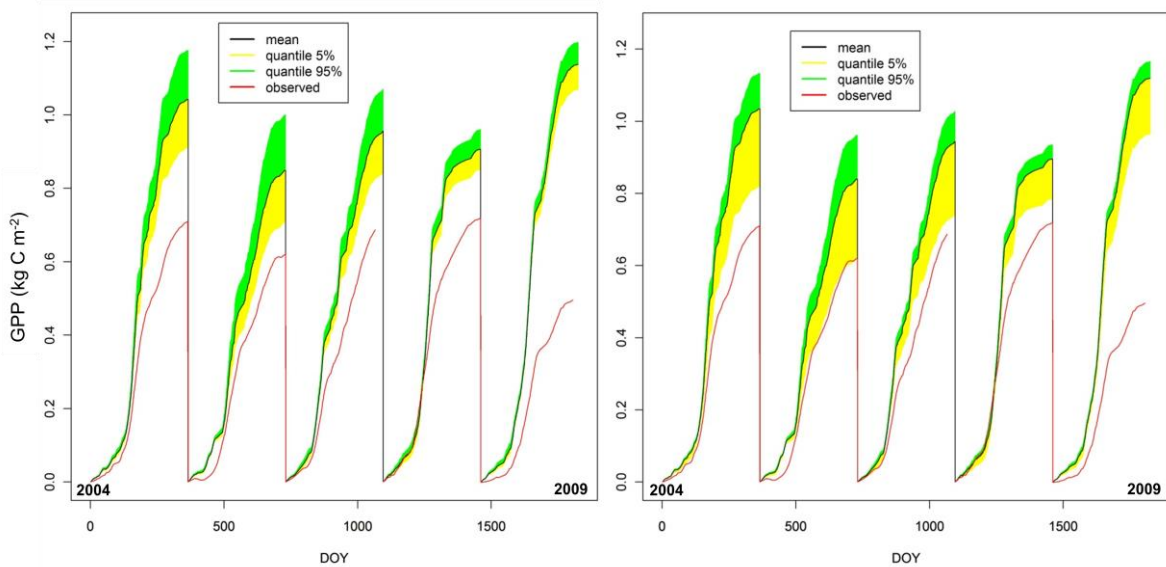


Figure 53. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la production primaire brute cumulée (GPP) au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme

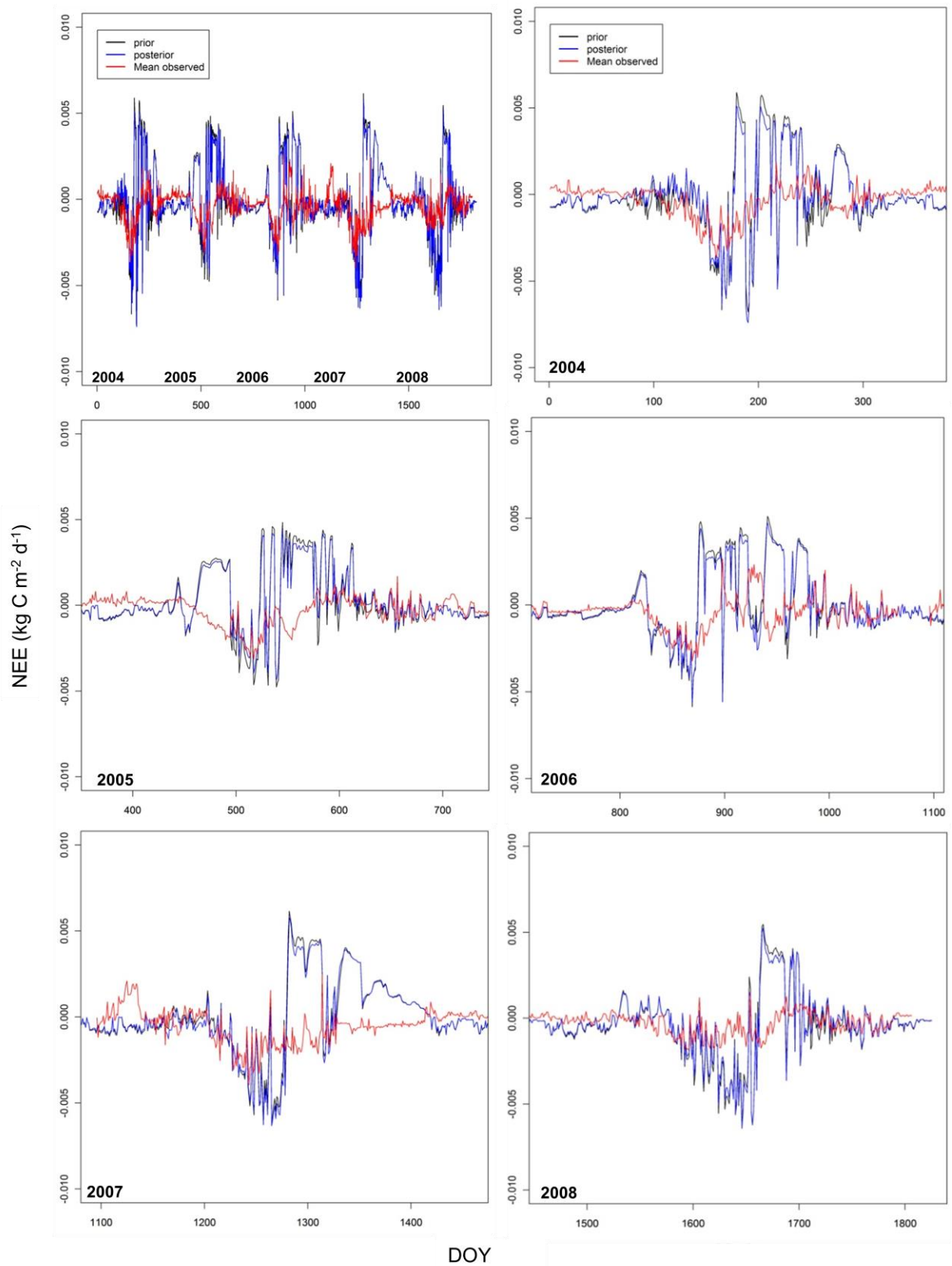


Figure 54. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes) pour le site espagnol de Vall d'Alinyà en utilisant une distribution uniforme

L'analyse d'incertitude des estimations du prior et du posterior de la température du sol sur cinq ans au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme (Figure 55) montre que malgré le bon résultat obtenu tant avec le posterior qu'avec le prior, le posterior arrive à réduire l'incertitude de cette sortie par rapport au prior et pratiquement sur toutes les années observées. On peut aussi observer que le modèle arrive à simuler correctement la température du sol pendant les périodes hivernales contrairement à ce qu'on a pu relever comme limite dans l'étude précédente. Cela peut s'expliquer par les caractéristiques du site étudié, situé au sud et dont la température du sol moyenne est de 9 °C et où il n'y a pas des grandes étapes neigeuses qui, elles, ne sont pas prises en compte dans les équations de PaSim.

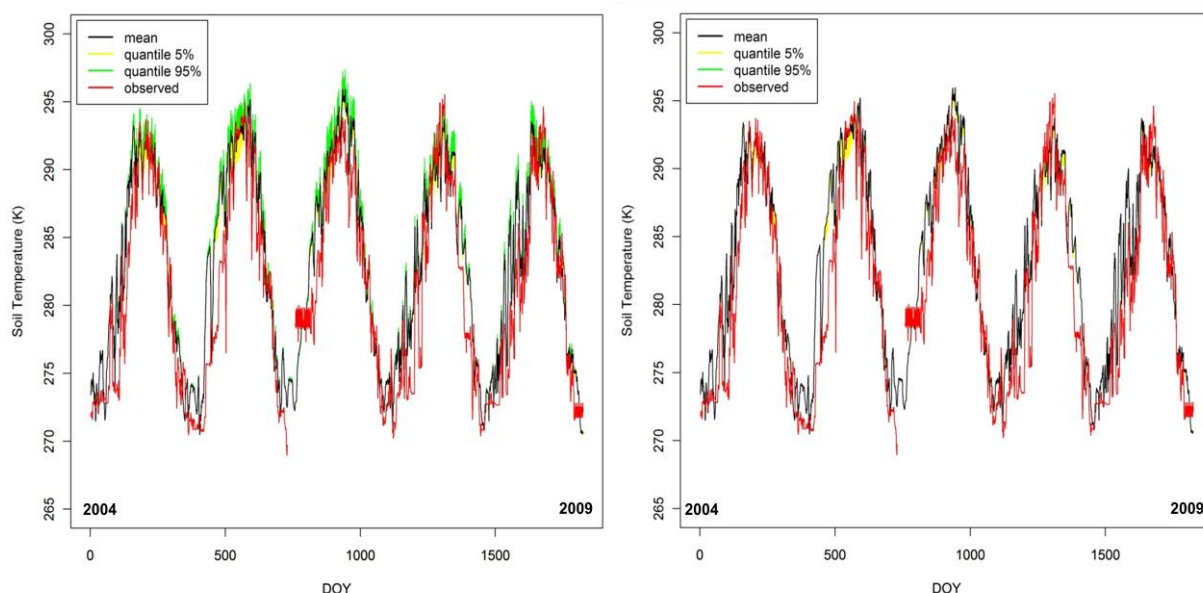


Figure 55. Analyse d'incertitude des estimations du prior (à gauche) et du posterior (à droite) de la température du sol au site espagnol de Vall d'Alinyà en utilisant une distribution uniforme

L'analyse d'incertitude des estimations du prior et du posterior de la biomasse au site hollandais de Cabauw en utilisant une distribution uniforme (Figure 56) vient confirmer l'amélioration des résultats obtenus avec le nouveau paramétrage. On peut observer que, sur la majorité des dates de mesure, le posterior affiche une performance meilleure avec une réduction importante d'incertitude.

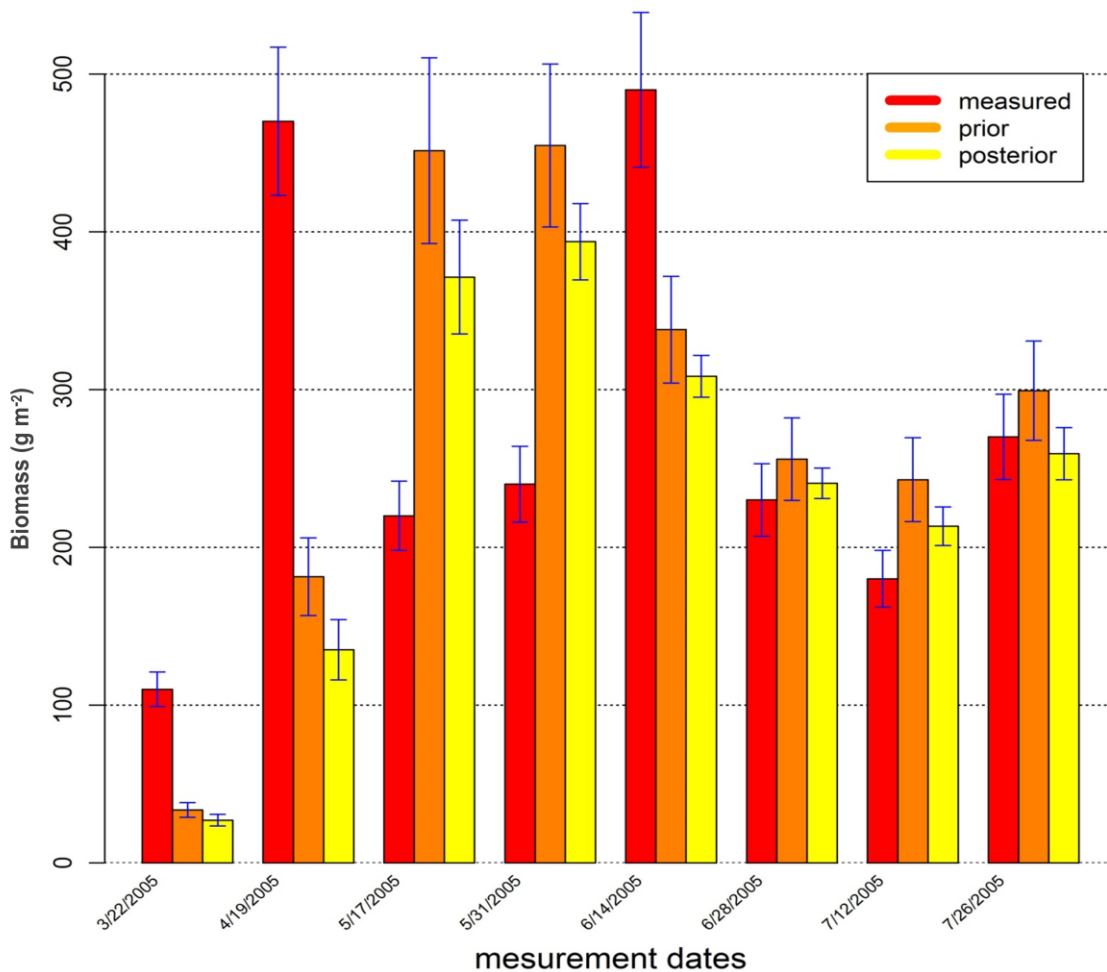


Figure 56. Analyse d'incertitude des estimations du prior et du posterior de la biomasse au site hollandais de Cabauw en utilisant une distribution uniforme

Nous avons aussi testé le nouveau paramétrage obtenu avec une distribution Gaussienne des valeurs a priori des paramètres du modèle. La Figure 57 illustre les estimations du prior et du posterior de la production primaire brute cumulée (GPP) au site espagnol de Vall d'Alinyà, où nous pouvons constater que les simulations du posterior sont nettement meilleures que celles du prior.

La Figure 58 montre les estimations du prior et du posterior de la NEE de 2004 à 2008 au site espagnol Vall d'Alinyà. Le résultat est similaire à celui obtenu avec la distribution uniforme (Figure 54) où l'on peut observer une similitude entre le prior et le posterior avec une légère avance pour ce dernier.

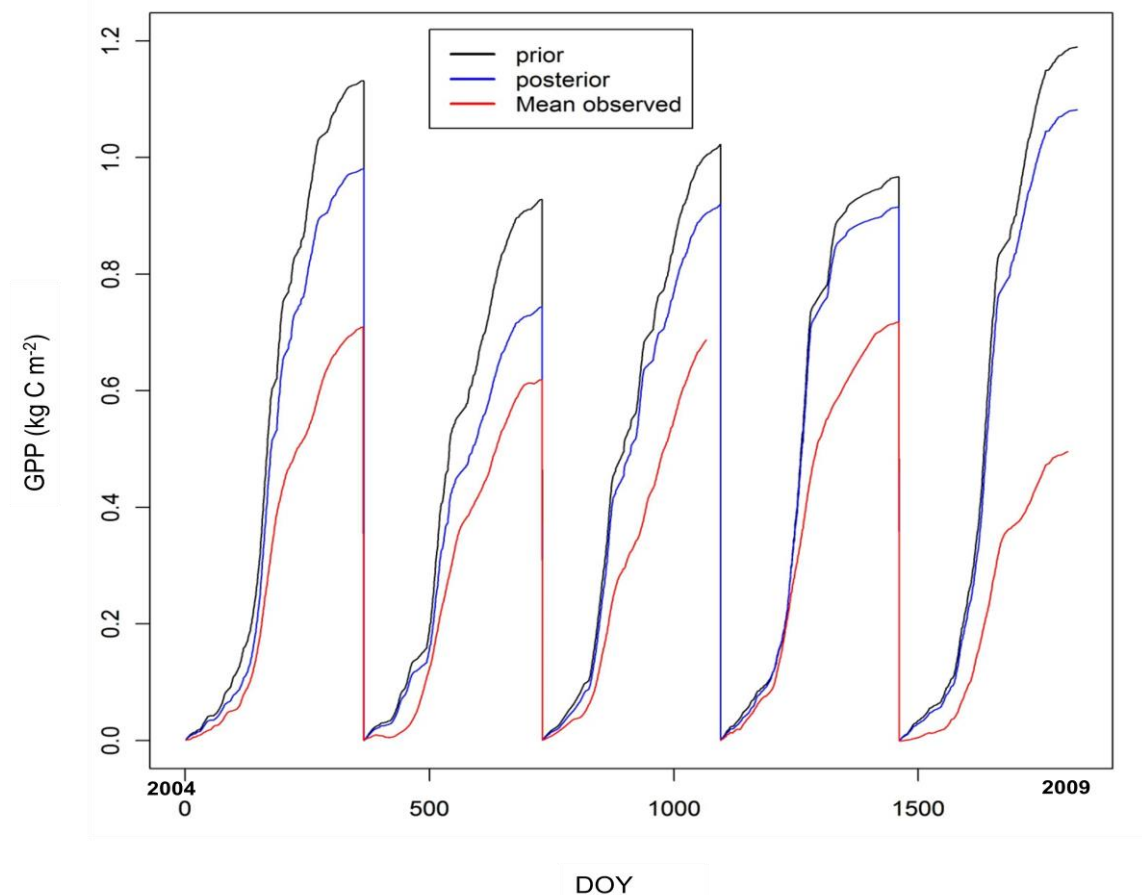


Figure 57. Comparaison des valeurs mesurées, du prior et du posterior de la production primaire brute cumulée (GPP) au site espagnol de Vall d'Alinyà en utilisant une distribution Gaussienne

Enfin, nous avons dressé un tableau récapitulatif des résultats de l'évaluation des estimations du prior et du posterior des quatre variables de sortie obtenues sur l'ensemble des sites de validation avec les deux approches uniforme et Gaussienne (Tableau 21), en utilisant les deux indices statistiques de performance qui sont l'erreur quadratique moyenne (RMSE) et le coefficient de masse résiduelle (CRM). La teneur en eau du sol montre certaine stabilité entre le prior et le posterior mais le résultat est légèrement meilleur en uniforme, et pareil pour la NEE. La GPP a été simulée plus correctement avec la distribution Gaussienne. Par contre, la biomasse aérienne estimée a posteriori a été meilleure avec la distribution uniforme. Ce qui est important à retenir, c'est que nous avons obtenu une amélioration globale des résultats sur les variables de sortie étudiées à l'aide du nouveau paramétrage dont la distribution a priori est soit uniforme soit Gaussienne, ce qui montre que la calibration Bayésienne multi-sites a permis d'aboutir sur un paramétrage commun fiable à l'échelle de l'Europe permettant d'obtenir des résultats plus précis sur des sites prairiaux représentant des conditions diverses et variées de ce continent.

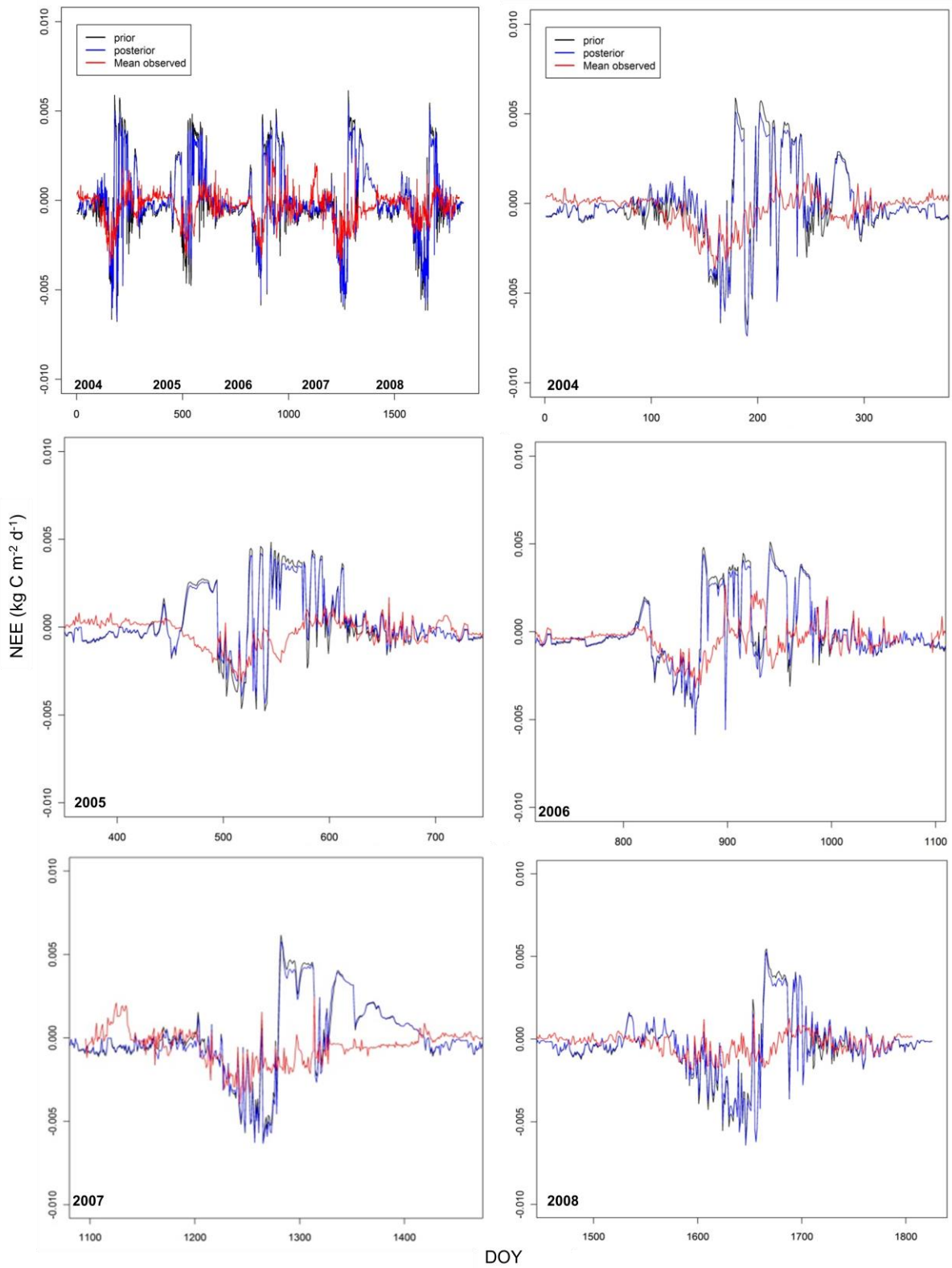


Figure 58. Comparaison des valeurs mesurées, du prior et du posterior de la NEE sur la période 2004-2008 (en haut à gauche), et pour chaque année (les autres graphes) pour le site espagnol de Vall d'Alinyà en utilisant une distribution Gaussienne

Tableau 21. Evaluation des estimations a priori et a posteriori de quatre variables de sortie sur l'ensemble des sites de validation avec les deux distributions uniforme et Gaussienne (RMSE, *root mean square error* ; CRM, *coefficient of residual mass*)

Distribution	Information	Teneur en eau du sol ($\text{m}^3 \text{m}^{-3}$)		NEE (g C m^{-2})		GPP (g C m^{-2})		Biomasse aérienne (g m^{-2})	
		RMSE	CRM	RMSE	CRM	RMSE	CRM	RMSE	CRM
Uniforme	Prior	0,12	0,09	0,02	-1,21	0,35	0,04	15,76	0,39
	Posterior	0,12	0,08	0,02	0,76	0,34	0,02	15,13	0,33
Gaussienne	Prior	0,13	0,10	0,02	-1,21	0,34	0,05	17,28	0,44
	Posterior	0,13	0,09	0,02	0,76	0,30	0,01	16,24	0,40

4.5. Conclusion

Dans ce chapitre, nous avons appliqué une démarche complète pour l'obtention d'un nouveau paramétrage de PaSim valable à l'échelle de l'Europe. Cette étude a commencé par une analyse de sensibilité multi-sites pour déterminer les paramètres les plus influents sur les sorties du modèle. Après avoir dressé un classement des paramètres, neuf paramètres pertinents liés à la végétation sont ressortis et une certaine plasticité du modèle a été observée. La deuxième étape a consisté à passer ces paramètres à une calibration Bayésienne multi-sites couvrant le continent avec des données à long terme où nous avons posé deux hypothèses relativement à la distribution a priori des valeurs des paramètres (uniforme et Gaussienne). Enfin, nous avons testé le paramétrage obtenu à l'issue de cette calibration sur des sites de validation en comparant les résultats du modèle aux observations.

Les résultats graphiques et analytiques obtenus sur les sites de calibration, mais également sur les sites de validation, montrent : 1) une stabilité de la méthode quant au choix de la distribution a priori, 2) une amélioration globale des résultats des simulations sur la majorité des variables de sortie étudiées en utilisant le nouveau paramétrage, et 3) une réduction d'incertitudes sur les paramètres et, par conséquent, sur les variables de sorties. À la différence de la calibration précédente, ce paramétrage n'est pas spécifique à chaque site. Il peut donc être utilisé pour réaliser des simulations de prairies à l'échelle de l'Europe.

Conclusion générale

En guise de conclusion, nous allons revenir sur les objectifs de cette thèse et la démarche appliquée dans ce cadre. Ensuite nous reviendrons sur les principales conclusions, observations et propositions qui émergent de cette étude. Enfin, nous terminerons par la présentation de quelques pistes de réflexion en perspective, concernant l'application de la calibration Bayésienne dans des contextes différents.

Les buts de cette étude doctorale effectuée dans un cadre applicatif était : 1) de fournir des réponses scientifiques sur le paramétrage des modèles complexes de simulations d'un écosystème prairial dans le domaine des impacts des changements climatiques, 2) de quantifier l'incertitude liée au paramètres des modèles et d'utiliser des techniques tel que la calibration Bayésienne pour pouvoir les réduire même en cas de conditions extrêmes, et 3) d'arriver à obtenir un paramétrage fiable et unique couvrant des zones géographiques et climatiques tel que le continent européen tout en améliorant les résultats des simulations.

Dans ce cadre, l'originalité de notre travail de thèse a été d'adapter la méthode Bayésienne à un modèle éco-systémique complexe (PaSim) et de montrer ses avantages potentiels dans la réduction d'incertitudes et l'amélioration des résultats, en combinant notamment méthodes statistiques (technique Bayésienne et analyse de sensibilité par la méthode de Morris) et outils informatiques (couplage code R-PaSim et utilisation du calcul intensif). Cette démarche nous a conduits à produire un paramétrage commun pour modéliser les prairies européennes, et à fournir un outil informatique générique donnant la possibilité de réaliser des calibrations Bayésienne avec d'autres modèles et sur d'autres sites. Or, à notre connaissance, aucune étude de cette ampleur n'a été effectuée pour paramétrer un modèle complexe de simulation des prairies à l'échelle de grand territoire, en passant par la calibration Bayésienne et précédée par une analyse de sensibilité des paramètres du modèle.

Après avoir défini les concepts de base relatifs aux écosystèmes prairiaux, leurs rôles et leurs réponses aux changements climatiques ainsi que leurs contributions aux émissions de gaz à effet de serre, nous avons dressé un état de l'art des techniques de calibration dont la calibration Bayésienne et des méthodes d'analyse de sensibilité dont celle de Morris et

leurs utilisations avec des modèles d'écosystèmes. Nous avons présenté aussi la notion de calcul intensif, ses outils et son utilité pour les simulations des modèles complexes.

Les modèles éco-systémiques sont développés pour fournir des simulations étalées sur le temps et l'espace pour lesquels des observations peuvent être peu ou pas disponibles. L'application du modèle à un nouveau site ou à une échelle spatiale plus large a donc besoin de la mise au point d'une stratégie de calibration permettant d'estimer des valeurs communes et globales des paramètres, tout cela en tenant compte des incertitudes liées au paramétrage du modèle qui, à leur tour, se propagent jusqu'aux sorties et dont le degré d'incertitude s'amplifie en cas d'événements climatiques extrêmes sur sites étudiés.

Le travail de cette thèse se compose de deux approches d'application de la calibration Bayésienne. Dans la première étude, nous avons développé une méthode de calibration des paramètres du modèle PaSim basée sur une approche Bayésienne, appliquée indépendamment sur des sites d'expérimentation à court et moyen termes incluant des manipulations de conditions de sécheresse. L'analyse des résultats issus de cette calibration a montré l'efficacité de la technique Bayésienne dans l'amélioration du paramétrage du modèle même dans des conditions de sécheresse, ce qui a permis par conséquent d'améliorer les résultats du modèle et de réduire les incertitudes. La limite de cette étude est que les calibrations étaient spécifiques au site étudié et les paramètres calibrés provenaient d'études antérieures sans analyse effectuée sur leur degré d'influence sur les sorties du modèle. Aussi, la validation indépendante du modèle calibré n'a pas été possible du fait de manque de données.

Dans la deuxième étude, nous avons mené une démarche plus large et plus globale couvrant divers sites prairiaux européens, où une calibration Bayésienne multi-sites a été mise en œuvre dans le but d'obtenir un nouveau paramétrage de PaSim valable pour les prairies européennes. Cette étude a commencé par une analyse de sensibilité multi-sites visant à déterminer les paramètres de végétation les plus influents sur les sorties du modèle. Dans l'absence d'une connaissance à priori de la distribution des valeurs des paramètres de PaSim, cette étude s'est appuyée sur deux hypothèses contrastées : distribution uniforme et distribution Gaussienne. Après avoir dressé un classement des paramètres, neuf paramètres pertinents liés à la végétation sont ressortis ainsi que l'aspect « plastique » du modèle. La deuxième étape a consisté à soumettre ces paramètres à une calibration Bayésienne multi-sites couvrant le continent avec des données à long terme où

nous nous sommes également appuyés sur deux hypothèses pour la distribution a priori des valeurs des paramètres (uniforme et Gaussienne). Afin de valider les résultats du modèle, nous avons testé ensuite le nouveau paramétrage obtenu à l'issue de cette calibration sur des sites de validation non inclus dans la calibration. Après analyse des résultats à la fois sur les sites de calibration et les sites de validation, nous avons constaté d'abord une stabilité de la méthode quant au choix de la distribution à priori, une amélioration globale des résultats des simulations sur la majorité des variables de sortie étudiées en utilisant le nouveau paramétrage et une réduction des incertitudes sur les paramètres et, par conséquent, sur les variables de sortie. Certes, cette calibration multi-sites donne des résultats moins bons qu'une calibration spécifique au site mais la différence est que ce paramétrage est générique et peut donc être utilisé sur d'autres prairies européennes, ce qui rend le modèle opérationnel dans la perspective d'études à l'échelle de grand territoire.

Perspectives

En termes de perspective, plusieurs pistes de réflexion se dégagent. Par exemple, une procédure peut être pensée où une deuxième calibration Bayésienne est tournée en utilisant la FDP a posteriori d'une première calibration comme FDP a priori de la deuxième, et tester ainsi le résultat obtenu, avec le risque que le paramétrage a priori soit dépendant dans ce cas des données utilisées pour la première calibration. On peut également réfléchir à l'implémentation de la calibration Bayésienne à l'aide d'un langage informatique plus évolué (tel que JAVA ou C++) en profitant des nouvelles bibliothèques mathématiques et statistiques dédiées à ces langages de programmation (ex. Bibliothèque Bayes++). Cela conduirait à améliorer le temps de calcul de la calibration, mais offrirait aussi la possibilité d'intégrer la calibration dans une interface homme-machine et produire ainsi un outil de calibration Bayésienne simplifié où l'on peut spécifier les sites, les sorties et les paramètres à calibrer et lancer la calibration à travers une interface graphique.

Appliquer la calibration Bayésienne (qui est générique) à d'autres modèles de prairie en utilisant les mêmes données et comparer ensuite les résultats, ce serait une sorte d'évaluation comparative de modèles. On peut aussi appliquer la calibration Bayésienne en utilisant des sites non européens, tempérés ainsi que tropicaux, afin de couvrir davantage de zones climatiques et géographiques. Dans ce contexte, des collègues travaillant déjà

avec PaSim [Ma et al., 2014a,b] ont commencé à utiliser notre nouveau paramétrage de PaSim sur plusieurs sites prairiaux européens dans le but de comparer les résultats des simulations avec d'autres modèles dans le cadre du projet MACSUR (<http://www.macsur.eu>) de la *Joint Research Programming Initiative on Agriculture, Food Security and Climate Change* (FACCE-JPI, <http://www.faccejpi.com>). Et une collaboration a déjà commencée avec des chercheurs du Norwegian Agricultural Economics Research Institute à Oslo en Norvège avec lesquels nous avons partagé nos codes sources de calibration Bayésienne, nos données et notre nouveau paramétrage pour les intégrer dans des études plus larges de calibration de modèles complexes et d'extrapolation des résultats à d'autres sites.

Tout cela témoigne de la possibilité d'ouvrir un vaste champ d'actions scientifiques dans la perspective d'améliorer les performances des modèles de simulation des écosystèmes afin de répondre aux questions soulevées dans les études sur les impacts des changements globaux (dont les changements climatiques ne sont qu'un aspect).

Bibliographie

- Acutis, M., Confalonieri, R., 2006. Optimization algorithms for calibrating cropping systems simulation models. A case study with simplex-derived methods integrated in the WARM simulation environment. *Italian Journal of Agrometeorology* 3, 26–34.
- Agreste (La statistique, l'évaluation et la prospective agricole), 2000. La conduite des prairies en France: pratiques intensives et rendements élevés dans le Nord-ouest. *Cahiers*, 4, 15-26. <http://www.agreste.agriculture.gouv.fr>.
- Agreste (La statistique, l'évaluation et la prospective agricole), 2008. Statistique agricole annuelle. <http://www.agreste.agriculture.gouv.fr>.
- Agreste (La statistique, l'évaluation et la prospective agricole), 2010. Agreste Infos rapides Animaux de boucherie - décembre 2010. <http://www.agreste.agriculture.gouv.fr>.
- Agreste (La statistique, l'évaluation et la prospective agricole), 2012. Statistique agricole annuelle. <http://www.agreste.agriculture.gouv.fr>.
- Albert, J., 2009. *Bayesian Computation with R*. 2nd edition, Springer. ISBN 978-0-387-92298-0.
- Amigues J.P., P. Debaeke, B. Itier, G. Lemaire, B. Seguin, F. Tardieu, A. Thomas (éditeurs), 2006. Sécheresse et agriculture. Réduire la vulnérabilité de l'agriculture à un risque accru de manque d'eau. Expertise scientifique collective, synthèse du rapport, INRA (France), 72 p.
- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. Lond.*, 53, 370-418.
- Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. *Water Resources Research* 23, 1393-1442.
- Ben Touhami, H., Lardy, R., Bahn, M., Bellocchi, G., Barra, V., 2012a. Bayesian calibration of the Pasture Simulation Model (PaSim) to simulate European grasslands under climate extremes: case study at Stubai (Austria). *Proceedings of the 11th PSAM & ESREL Conference, 25-29 June, Helsinki, Finland, Vol. 1*, pp. 567-577.
- Ben Touhami, H., Lardy, R., Barra, V., Bellocchi, G., 2013a. Screening parameters in the Pasture Simulation Model using the Morris method. *Ecological Modelling*, 266, 42-57.
- Ben Touhami, H., Lardy, R., Gilgen, A. K., Buchmann, N., Barra V., Bellocchi G., 2012b. Bayesian calibration of the Pasture Simulation Model (PaSim) to simulate Swiss grasslands under climate extremes. *Proceedings of the 12th Congress of the European Society for Agronomy, 20-24 August, Helsinki, Finland*, 441-5.
- Ben Touhami, H., Lardy, R., Klumpp, K., Barra, V., Bellocchi, G., 2013b. Bayesian calibration of the Pasture Simulation model (PaSim) to simulate emissions from long-term European grassland sites: a case study at Laqueuille (France). *Greenhouse Gases & Animal Agriculture Conference, 23-26 June, Dublin, Ireland*, p. 580.

- Box, G.E.P., Draper, N.R., 1987. Empirical Model-Building and Response Surfaces, p. 424, Wiley. ISBN 0471810339.
- Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M., H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J., M., Meynard, J., M., Delecolle, R., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18, 311–346.
- Brisson, N., Ruget, F., Gate, P., Lorgeau, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M., H., Bouthier, A., Ripoche, D., Mary, B., Justes, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie*, 22, 69-92.
- Brochot, C., 2004. Paramétrisation des modèles physiologiques toxico/pharmacocinétiques. Université Pierre et Marie Curie - Paris 6, thèse de doctorat, 156 p.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22, 1509-1518.
- Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2007. The role of sensitivity analysis in ecological modelling. *Ecological Modelling* 203, 167–182.
- Carlin, B.P., Louis, T.A., 2009. Data sets in Bayesian Methods for Data Analysis, 3rd ed., Chapman and Hall/CRC Press, ISBN: 978-1-58488-6976.
- Carranza, P., Kuzniak, A., 2007. Approche bayésienne « cachée » et approche fréquentiste « ambiguë ». 4èmes Rencontres Internationales, Analyse Statistique Implicative, 8p.
- Chevassus-au-Louis, B., Salles, J.-M., Pujol, J.-L., 2009. Approche économique de la biodiversité et des services liés aux écosystèmes. Contribution à la décision publique. Centre d'analyse stratégique. 376 p.
- Clark, J. S., 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2-14.
- Confalonieri, R., Acutis, M., Bellocchi, G., Cerrani, I., Tarantola, S., Donatelli, M., Genovese, G., 2006. Exploratory sensitivity analysis of CropSyst, WARM and WOFOST: a case-study with rice biomass simulations. *Italian Journal of Agrometeorology* 3, 17-25.
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., Acutis, M., 2010a. Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling*, 221, 1897-1906.
- Confalonieri, R., Bregaglio, S., Acutis, M., 2010b. A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. *Ecological Modelling* 221, 960-964.
- Confalonieri, R., Bregaglio, S., Acutis, M., 2012. Quantifying plasticity in simulation models. *Ecological Modelling* 225, 159-166.

- Coquillard, P., Hill, D., 1997. Modélisation et simulation d'écosystèmes. Des modèles déterministes aux simulations à événements discrets. Masson publ., Paris, France, 273 pp.
- Cotter S.C., 1979. A screening design for factorial experiments with interactions. *Biometrika* 66, 317-320.
- Cruz, P., Duru, M., Therond, O., Theau, J.P., Ducourtieux, C., Jouany, C., Al Haj Khaled, R., Ansquer, P., 2002. Une nouvelle approche pour caractériser les prairies et leur valeur d'usage. *Fourrages*, 172, 335–354.
- Cryer, S.A., Havens, P.L., 1999. Regional sensitivity analysis using a factorial method for the USDA model GLEAMS. *Environmental Modelling & Software* 14, 613–624.
- Cukier, H., Levine, R.I., Shuler, K., 1978. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26,1-42.
- De Martonne, E., 1942. Nouvelle carte mondiale de l'indice d'aridité. *Annales de Géographie* 51, 242-250.
- Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 138.
- Diodato, N., Ceccarelli, M., 2004. Multivariate indicator Kriging approach using a GIS to classify soil degradation for Mediterranean agricultural lands. *Ecological Indicators* 4, 177-187.
- Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M., Badran, F., Thiria S., Hérault, L., 2002. Réseaux de neurones - Méthodologie et applications. Édition Eyrolles, Paris, France.
- Field C.B., Barros V., Stocker, T.F., Qin D., Dokken D.J., Ebi K.L., Mastrandrea M.D., Mach K.J., Plattner G.-K., Allen S.K., Tignor M., Midgley P.M., 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. A special report on working groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York, p. 582.
- Fiore, S., Aloisio, G., 2011. Grid and cloud database management. Springer-Verlag, Berlin.
- Foscarini, F., Bellocchi, G., Confalonieri, R., Savini, C., van den Eede, G., 2010. Sensitivity analysis in fuzzy systems: integration of SimLab and DANA. *Environmental Modelling & Software* 25, 1256-1260.
- Foster, I., Kesselman, C., 1999. The Grid: Blueprint for a new computing infrastructure, Morgan Kaufmann, 675p.
- Gabrielle, B., Laville, P., Duval, O., Nicoulaud, B., Germon, J.C., Hénault, C., 2006. Process-based modeling of nitrous oxide emissions from wheat-cropped soils at the subregional scale. *Global Biogeochem. Cycles*, 20, GB4018.
- Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software*, 22, 1000-1020.

- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis. 2nd edition. CRC Press, 668 pp.
- Gelman, A., Roberts, G., Gilks, W., 1996. Efficient metropolis jumping hules. Bayesian statistics, 5, 599-608.
- Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., Sitch, S., 2004. Terrestrial vegetation and water balance–Hydrological evaluation of a dynamic global vegetation model, J. Hydrol., 286, 249–270.
- Gotelli, N.J., 1998. A primer of ecology, Sunderland, Sinauer, 206 p. 2nd Eds.
- Graux, A.-I., 2011. Modélisation des impacts du changement climatique sur les écosystèmes prairiaux. Voies d’adaptation des systèmes fourragers, Université Blaise Pascal, Thèse de doctorat, 535 p.
- Graux, A.-I., Bellocchi, G., Lardy, R., Soussana, J.-F., 2012a. Ensemble modelling of climate change risks and opportunities for managed grasslands in France. Agricultural and Forest Meteorology, 170, 114-131.
- Helton, J.C., Davis, F.J., Johnson, J.D., 2005. A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. Reliability Engineering & System Safety 89, 305-330.
- Hill, D., 2009. Simulation Informatique au Service des Sciences de la Vie, Annals. Université Blaise Pascal, 210 p.
- Hoffman, A.R., Traub, J.F., 1989. Supercomputers: directions in technology and applications. National Academies Press, 102p.
- Iman, R.L., Conover, W.J., 1987. A measure of top-down correlation. Technometrics 293, 351–357.
- Institut de l’Elevage, 2007. La prairie, un enjeu économique et sociétal. Dossier économie de l’élevage, Hors série, spécial prairie, 40 pp.
- IPCC (Intergovernmental Panel on Climate Change), 2007. Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, 976 p.
- IPCC (Intergovernmental Panel on Climate Change), 2013. Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, 1535 p.
- Johnson, I.R., Thornley, J.H.M., 1985. Dynamic model of the response of a vegetative grass crop to light, temperature and nitrogen. Plant, Cell and Environment 8, 485-499.
- Jorgensen S. E., 1994. Fundamentals of Ecological Modelling, Elsevier, 632 p. 2nd Eds.
- Kaufman, A., 1973. Introduction à la logique floue. Ed. Techniques Ingénieur.
- Kleijnen, J., van Groenendaal, W., 1992. Simulation a statistical perspective, Wiley, Chichester, UK, 241p.

- Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., Cioppa, T.M., 2005. A User's Guide to the Brave New World of Designing Simulation Experiments. *INFORMS Journal on Computing*, 17, 263-289.
- Kobilinsky, A., 1990. Analyse factorielle discriminante. In *Analyse discriminante sur variables continues*. Collection Didactique. Ed. INRIA. CEDIS-Diffusion INRIA, Rocquencourt France, 65-78.
- Kung, S.Y., Mak, M.W., Lin, H.S., 2005. Biometric authentication: a machine learning approach. Ed. Prentice Hall Professional Technical Reference.
- Lardy, R., 2013. Calcul intensif pour l'évaluation de la vulnérabilité en utilisant une approche d'Ingénierie Dirigée par les Modèles. Application à la vulnérabilité des prairies au changement climatique sous contraintes de plans d'expériences. Université Blaise Pascal, Thèse de doctorat, 254 p.
- Legendre, A.M., 1805. Nouvelles méthodes pour la détermination des orbites des comètes avec un supplément contenant divers perfectionnements de ces méthodes et leur application aux deux comètes de 1805. Courcier, Paris.
- Lehuger, S., Gabrielle, B., van Oijen, M., Makowski, D., Germon, J.-C., Morvan, T., Hénault, C., 2009. Bayesian calibration of the nitrous oxide emission module of an agroecosystem model. *Agriculture, Ecosystems and Environment*.
- Levenberg, K., 1944. A Method for the Solution of Certain Problems in Least Squares. *The Quarterly of Applied Mathematics*. 2,164-168.
- Lindley, D.V., 1971. *A Mathematical Theory of Evidence*. Princeton University Press.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337.
- Ma S., Acutis M., Barcza Z., Ben Touhami H., Doro L., Hidy D., Köchy M., Minet J., Lellei-Kovács E., Perego A., Rolinski S., Ruget F., Seddaiu G., Wu L., Bellocchi G., 2014a. The grassland model intercomparison of the MACSUR (Modelling European Agriculture with Climate Change for Food Security) European knowledge hub. International Environmental Modelling and Software Society (iEMSs) International Congress "Bold Visions for Environmental Modelling", June 15-19, San Diego, CA, USA (soumis).
- Ma, S., Lardy, R., Graux, A.I., Ben Touhami, H., Klumpp, K., Martin, R., Bellocchi, G., 2014b. On approaches and applications of Pasture Simulation model to simulate carbon and water exchanges in grassland systems. *Environmental Modelling & Software* (accepté).
- Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M.H., 2006a. Parameter estimation for crop models. In: Wallach, D., Makowski, D., Jones, J.W. Eds, *Working with dynamic crop models. Evaluation, analysis, parameterization and applications*. Amsterdam: Elsevier, 101-149.
- Makowski, D., Jeuffroy, M. H., 2002. Estimation des paramètres des modèles de culture. In *Ecole-Chercheurs - Pour une bonne utilisation des modèles de culture* chapter 4. INRA.

- Makowski, D., Naud, C., Jeuffroy, M.H., Barbottin, A., Monod, H., 2006b. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. *Reliability Engineering & System Safety*, 91, 1142-1147.
- Marquardt, D., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*. 11, 431-441.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087–1092.
- Monod, H., Naud, C., Makowski, D., 2006. Uncertainty and sensitivity analysis for crop models. In: Wallach, D., Makowski, D., Jones, J. (Eds.), *Working with dynamic crop models*. Elsevier, Amsterdam, The Netherlands, pp. 55-100.
- Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161-174.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Computer Journal*, 7, 308-313.
- Olesen, J.E., Carter, T.R., Díaz-Ambrona, C.H. Fronzek, S., Heidmann, T., Hickler, T., Holt, T., Minguez, M.I., Morales, P., Palutikof, J., Quemada, M., Ruiz-Ramos, M., Rubæk, G., Sau, F., Smith, B., Sykes, M., 2007. Uncertainties in projected impacts of climate change on European agriculture and terrestrial ecosystems based on scenarios from regional climate models. *Climatic Change*, 81, 123–143.
- Parent, E., Bernier, J., 2007. *Le raisonnement bayésien: modélisation et inférence*. Springer.
- Parent, É., Bernier, J., 2007. *Le raisonnement bayésien: Modélisation et inférence, Statistique et probabilités appliquées*. Springer, 364 pp.
- Peyraud, J.-L., 2012. Réforme de la PAC et prairies permanentes. Dans : *La PAC a 50 ans : le bel âge ? Colloque organisé par Institut National de la Recherche Agronomique dans le cadre du Salon International de l'Agriculture*, 26 février 2013, Paris.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing Series. New-York.
- Pujol, G., 2009. Simplex-based screening designs for estimating metamodels. *Reliability Engineering & System Safety* 94, 1156-1160.
- Rabitz, H., 1989. System analysis at molecular scale. *Science* 246, 221–226.
- Richter, G.M., Acutis, M., Trevisiol, P., Latiri, K., Confalonieri, R., 2010. Sensitivity analysis for a complex crop model applied to Durum wheat in the Mediterranean. *European Journal of Agronomy* 32, 127-136.
- Riedo, M., Grub, A., Fuhrer, J., 1998. A pasture simulation model for dry matter production, and fluxes of carbon, nitrogen, water and energy. *Ecological Modelling* 105, 141–183.

- Riedo, M., Grub, A., Rosset, M., Fuhrer, J., 1998. A pasture simulation model for dry matter production and fluxes of carbon, nitrogen, water and energy. *Ecological Modelling*, 105, 41-183.
- Riedo, M., Gyalistras, D., Fischlin, A., Fuhrer, J., 1999. Using an ecosystem model linked to GCM-derived local weather scenarios to analyse effects of climate change and elevated CO₂ on dry matter production and partitioning, and water use in temperate managed grasslands. *Global Change Biology* 5, 213-223.
- Robert. C.P., 2001. L'analyse statistique bayésienne. *Courrier des statistiques* n° 100.
- Ruget, F., Durand J.L., Moreau J.C., Lacroix B., Ripoche D., Graux A.I., Bernard F., 2013. Impacts des changements climatiques sur les productions de fourrages (prairies, luzerne, maïs) : variabilité selon les régions et les saisons. *Fourrages*, 214, 99-110.
- Ruget, F., Novak, S., Granger, S., 2006. Use of the ISOP system, based on STICS model for the assessment of forage production. *Adaptation to grassland and spatialized application*. *Fourrages*, 186, 241–256.
- Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15, 377-395.
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. *Sensitivity Analysis in Practice*. Wiley, New York, NY, USA, pp. 219.
- Saltelli, A., Tarantola, S., Chan, K., 1999. A quantitative, model independent method for global sensitivity analysis of model output. *Technometrics* 41, 39-56.
- Savage, I.R., 1956. Contributions to the theory of rank order statistics - the two-sample case. *Annals of Mathematical Statistics* 27, 590-615.
- Schmitt, M., Bahn, M., Wohlfahrt, G., Tappeiner, U., Cernusca, A., 2010. Land use affects the net ecosystem CO₂ exchange and its components in mountain grasslands. *Biogeosciences*, 7, 2297-2309.
- Seguin, B., 2007. Le réchauffement climatique et ses conséquences pour la viticulture. Proceedings of the “Global warming, which potential impacts on the vineyards?” Congress, Dijon, France, 1-9.
- Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., Venevsky, S. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Global Vegetation Model. *Global Change Biology* 9, 161-185.
- Sivia, D.S., 1996. *Data analysis: a Bayesian tutorial*. Oxford University Press, Oxford.
- Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., O'Mara, F., Rice, C., Scholes, B., Sirotenko, O., 2007. Agriculture. In *Climate Change 2007 : Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Metz, B. and Davidson, O. R. and Bosch, P. R. and Dave, R. and Meyer, L. A.[eds.]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

- Sobol, I.M., 1993. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1, 407-414.
- Soussana, J.-F., Graux, A.-I., Tubiello, F.N., 2010. Improving the use of modelling for projections of climate change impacts on crops and pastures, *J. Exp. Bot.*, 61, 2217–2228.
- Stephenson, N.L., 1990. Climatic control of vegetation distribution: the role of the water balance. *American Naturalist* 135, 649-670.
- Sutton, M. A., Nemitz, E., Erisman, J.W., Beier, C., Bahl, K. B., Cellier, P., de Vries, W., Cotrufo, F., Skiba, U., Di Marco, C., Jones, S., Laville, P., Soussana, J. F., Loubet, B., Twigg, M., Famulari, D., Whitehead, J., Gallagher, M. W., Neftel, A., Flechard, C. R., Herrmann, B., Calanca, P. L., Schjoerring, J. K., Daemmgen, U., Horvath, L., Tang, Y. S., Emmett, B. A., Tietema, A., Penuelas, J., Kesik, M., Brueggemann, N., Pilegaard, K., Vesala, T., Campbell, C. L., Olesen, J. E., Dragosits, U., Theobald, M. R., Levy, P., Mobbs, D. C., Milne, R., Viovy, N., Vuichard, N., Smith, J. U., Smith, P., Bergamaschi, P., Fowler, D., Reis, S., 2007. Challenges in quantifying biosphere-atmosphere exchange of nitrogen species. *Environmental Pollution* 150, 125–139.
- Thornley J., H., M., 1998. *Grassland dynamics: an ecosystem simulation model*. CAB International, Wallingford, United Kingdom, p 241.
- Thornley, J., H., M., 2001. Modelling grassland ecosystems. *Proceedings of the XIX International Grassland Congress*, in: Mattos W.R.S., da Silva S.C. (Eds.), Sao Paulo, Brazil, 1029–1035.
- Thornley, J.H.M., Verberne, E.L.J., 1989. A model of nitrogen flows in grassland. *Plant, Cell & Environment* 12, 863-886.
- Transtrum, M.K., Sethna, J.P., 2012. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. Preprint submitted to *Journal of Computational Physics*, January 30.
- Turton, I., Openshaw, S., 1998. High-performance computing and geography: developments, issues, and case studies. *Environment Planning A* 30, 1839-1856.
- Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models : bridging the gap between models and data. *Tree Physiology* 25, 915-927.
- Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. *Environmental Modelling & Software*, 25, 310-319.
- Varma, A., Morbidelli, M., Wu, H., 1999. *Parametric Sensitivity in Chemical Systems*. Cambridge University Press.
- Vose Software, 2007. Help file for ModelRisk version 5 (available at <http://www.vosesoftware.com/vosesoftware/ModelRiskHelp>).
- Vuichard, N., Ciais, P., Viovy, N., Calanca, P., Soussana, J.-F., 2007b. Estimating the greenhouse gas fluxes of European grasslands with a process-based model: 2. Simulations at the continental level. *Global biogeochemical cycles* 21, GB1005.

- Vuichard, N., Soussana, J.-F., Ciais, P., Viovy, N., Ammann, C., Calanca, P., Clifton-Brown, J., Fuhrer, J., Jones, M., Martin, C., 2007a. Estimating the greenhouse gas fluxes of European grasslands with a process-based model: 1. Model evaluation from in situ measurements, *Global Biogeochem. Cycles*, 21, GB1004.
- Wallach, D., Buis, S., Lecharpentier, P., Bourges, J., Clastre, P., Launay, M., Bergez, J.-E., Guérif, M., Soudais, J., Justes, E., 2011. A package of parameter estimation methods and implementation for the STICS crop-soil model. *Environmental Modelling & Software*, 26, 1-9.
- Walter, I.A., Allen, R.G., Elliott, R., Itenfisu, D., Brown, P., Jensen, M.E., Mecham, B., Howell, T.A., Snyder, R.L., Eching, S., Spofford, T., Hattendorf, M., Martin, D., Cuenca, R.H., Wright, J.L., 2005. The ASCE standardized reference evapotranspiration equation. Final report of the Task Committee on Standardization of Reference Evapotranspiration. Environmental and Water Resources Institute of the American Society of Civil Engineers, Reston, VA, USA, p. 59.
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C., Wang, Y.-P., 2009. Improving land surface models with FLUXNET data. *Biogeosciences Discussions* 6, 2785-2835.
- Zeeman, M.J., Hiller, R., Gilgen, A.K., Michna, P., Pluss, P., Buchmann, N., Eugster, W., 2010. Management and climate impacts on net CO₂ fluxes and carbon budgets of three grasslands along an elevational gradient in Switzerland. *Agricultural and Forest Meteorology*. 150, 519-530.

Annexes

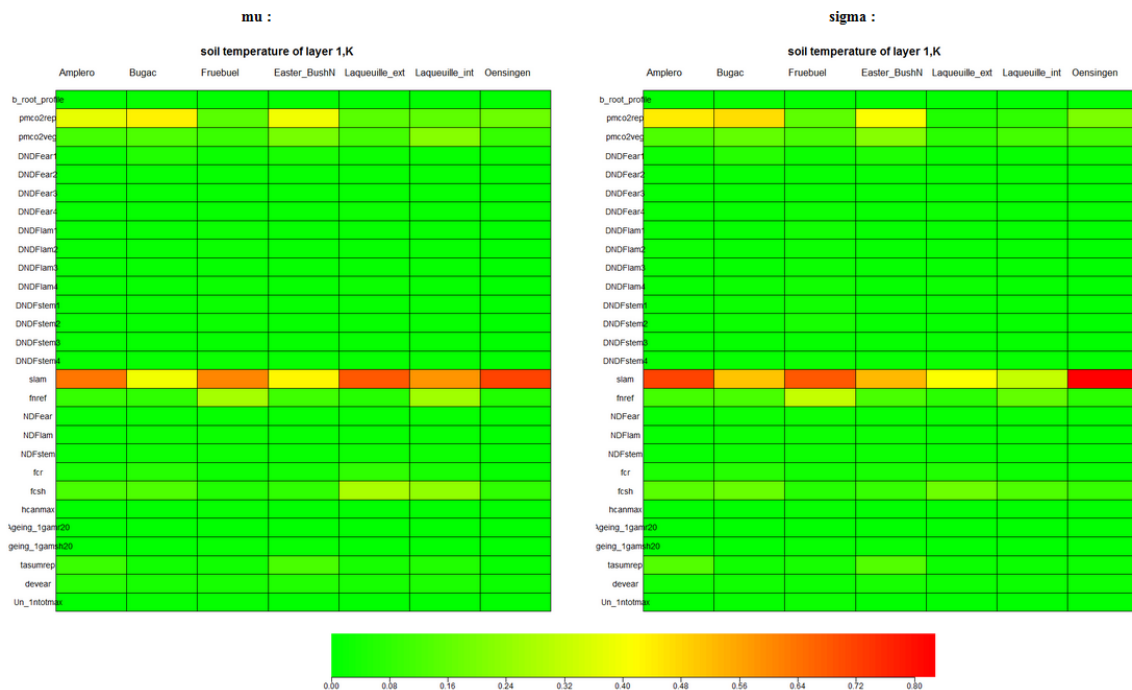
I. Implémentation du code principal de la calibration Bayésienne sous R

```
Bayesian_Calibration.R
1 source("Preliminaries_rev2.R") # contain necessary data (lai, biomass..) and path for other daily data (tsoil, SWC, NEE..)
2 load("init.Rdata") # initial conditions
3
4 nChain <- as.integer(100000) # set the length of the Markov Chain to be generated (nChain)
5 nBI <- as.integer(2000) # burn-in (nBI : number of iterations to stabilize the model)
6 pChain <- matrix(0, nrow=nChain, ncol=npars) # - matrix used for storing the candidates (pChain);
7 pChAcc <- matrix(0, nrow=nChain, ncol=npars) # - the accepted states (pChAcc) with their index in the chain (jChAcc)
8 jChAcc <- matrix(0, nrow=nChain, ncol=npars) #--npars : number of marameters
9 covPar <- matrix(0, nrow=npars, ncol=npars) # - the mean vector (avePar) and covariance (covPar)
10
11 pValues <- runif(npars, min=parmin, max=parmax) # Start the chain using uniform random values sampled in [parmin, parmax]
12
13 Param(1) = pValues[1]
14 Param(2) = pValues[2]
15 ...
16
17 pasim <- lance_pasim() # First call to pasim (to get outputs).
18
19 # Calculate corresponding prior probability and likelihood
20 logPrior0 <- sum( dunif(pValues, min=parmin, max=parmax, log=TRUE) )
21 logL0 <- calcul_logL(pasim)
22 #calcul_logL a fonction giving Likelihood: <- sum( -0.5*((simulated - observed)/standard deviation (obs))^2
23 #- 0.5*log(2.*pi) - log(standard deviation (obs)))
24
25 psetMAP <- pValues
26 logMAP <- logPrior0 + logL0
27 pChain[1,] <- pValues
28
29 # proposal variance matrix needed to define the multivariate normal step away from the current vector (vcovProp)
30 vcovProp <- diag( (0.01*(parmax - parmin)/parmod)^2 ) #construct a diagonal matrix
31 scalProp <- 2.4^2/npars # This is the scaling factor proposed by Gelman et al. (1996) , it will be multiplied by the
32 matrix
33
34 # Build up the chain. Candidates for the parameter values (candidatepValues) are assumed to stem from a multivariate
35 #normal distribution (mvnrm) with mean at the current state and covariance given by scalProp*covPar.
36 #If the candidates are outside [parmin, parmax], then candidates within this range are sampled using reflection at parmin or parmax.
37
38 covPar <- vcovProp
39 kc <- 0 # number of accepted parameters candidate
40
41 for (j in 2:nChain)
42 {
43   candidatepValues <- mvnrm(n=1, pValues, scalProp*covPar, tol=1e-6, empirical=FALSE)
44
45   #-- Call the model with the candidate values, and check acceptance criterium.
46
47   Param(1) = candidatepValues [1]
48   Param(2) = candidatepValues [2]
49   ...
50
51   pasim <- lance_pasim()
52
53   #-- Update prior probabilities and likelihood
54   logPrior1 <- sum( dunif(candidatepValues, min=parmin/parmod, max=parmax/parmod, log=T) )
55
56   #-- Check whether the candidates are accepted. If yes and if burn-in (2000 first iterations to stabilize the model)
57   #has been completed, then store corresponding values and positions within the chain in pChAcc and jChAcc.
58
59   logalpha <- logPrior1 + logL1 - (logPrior0 + logL0) # metropolis ratio: likelihood for the candidate and likelihood for current point
60
61   if (log(runif(1,0,1)) < logalpha) # If the Metropolis ratio is larger than 1 (i.e. the candidate point has a higher posterior probability
62   #then the current point) ==>it is accepted
63   {
64     pValues <- candidatepValues
65     logPrior0 <- logPrior1
66     logL0 <- logL1
67
68     if (j > nBI)
69     {
70       if ((logPrior0 + logL0) > logMAP)
71       {
72         logMAP <- logPrior0 + logL0
73         psetMAP <- pValues
74       }
75       kc <- kc + 1
76       jChAcc[kc,] <- j
77       pChAcc[kc,] <- pValues
78     }
79   }
80   pChain[j,] <- pValues # store the new values of parameters
81 }
```

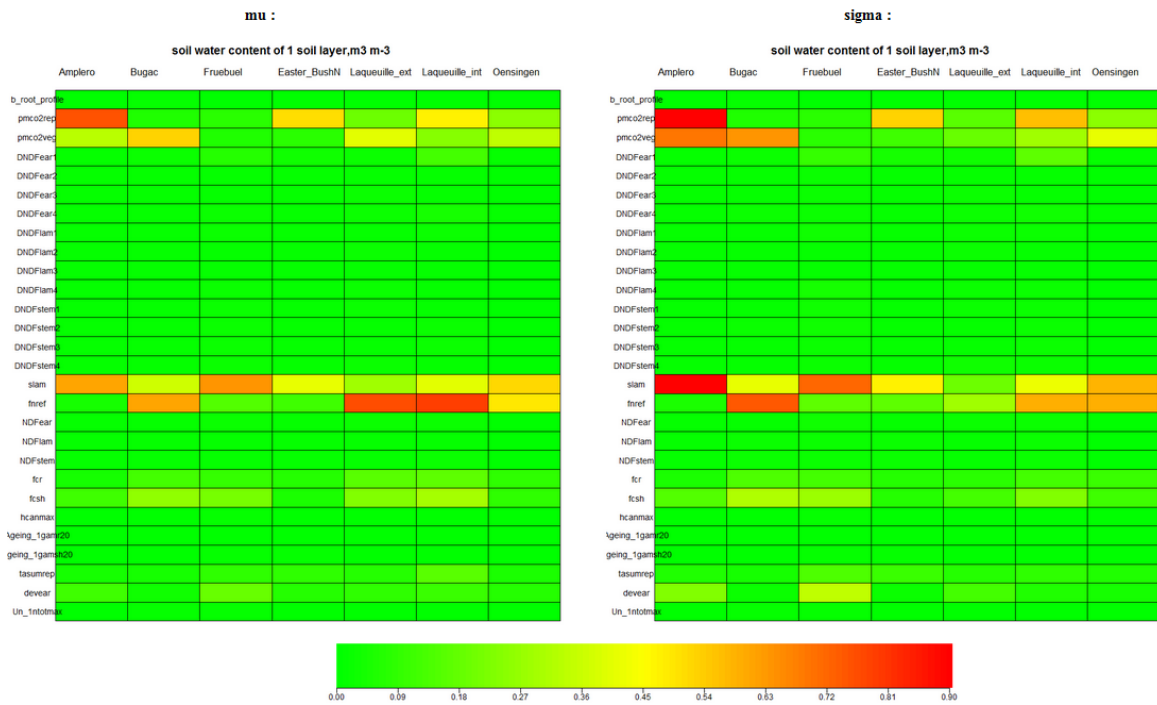
II. Résultats de l'analyse de sensibilité



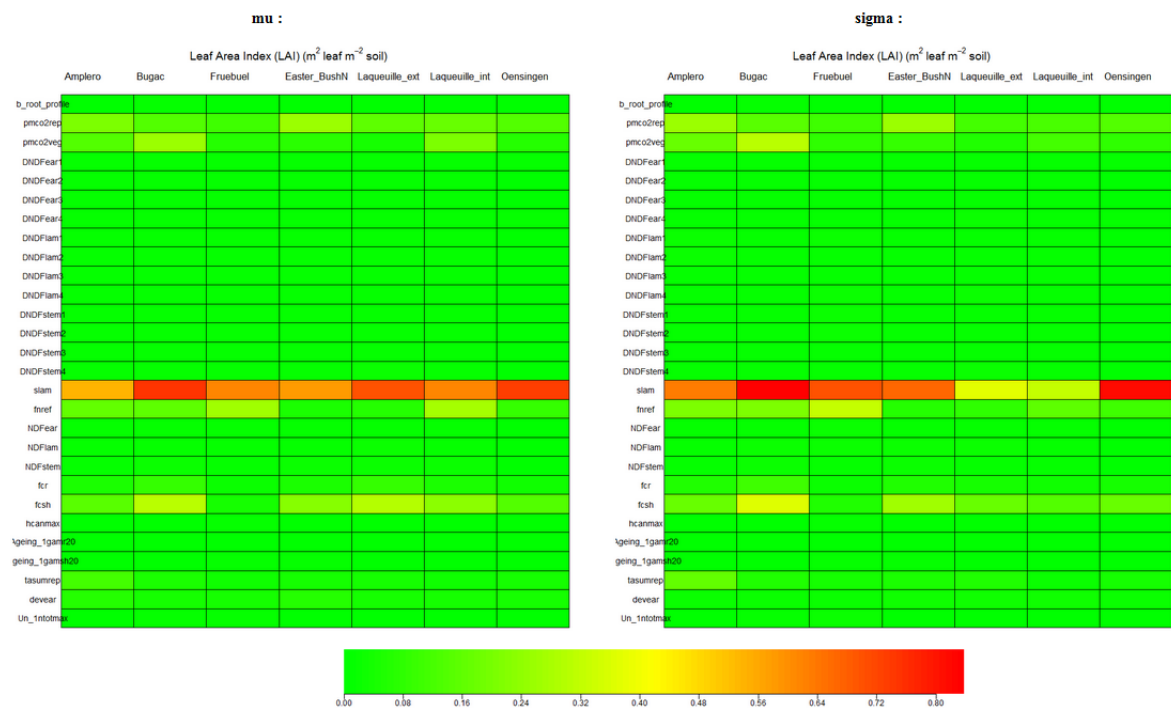
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la concentration du substrat de carbone en plante (C) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



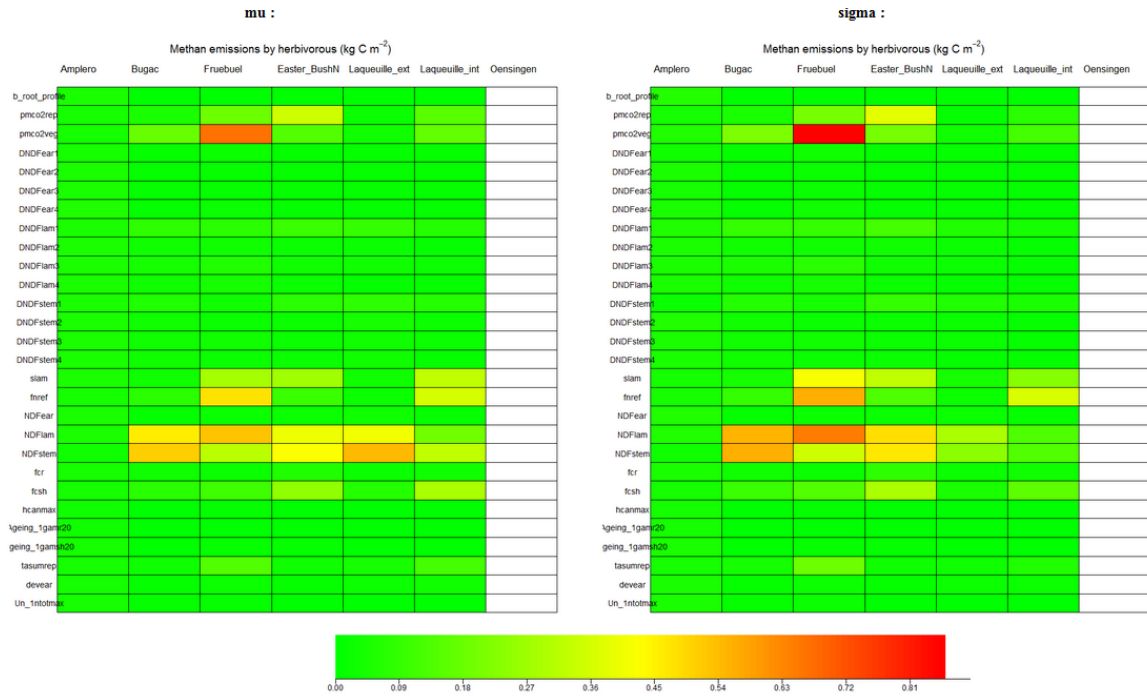
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la température du sol à la première couche (T_{soil}) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



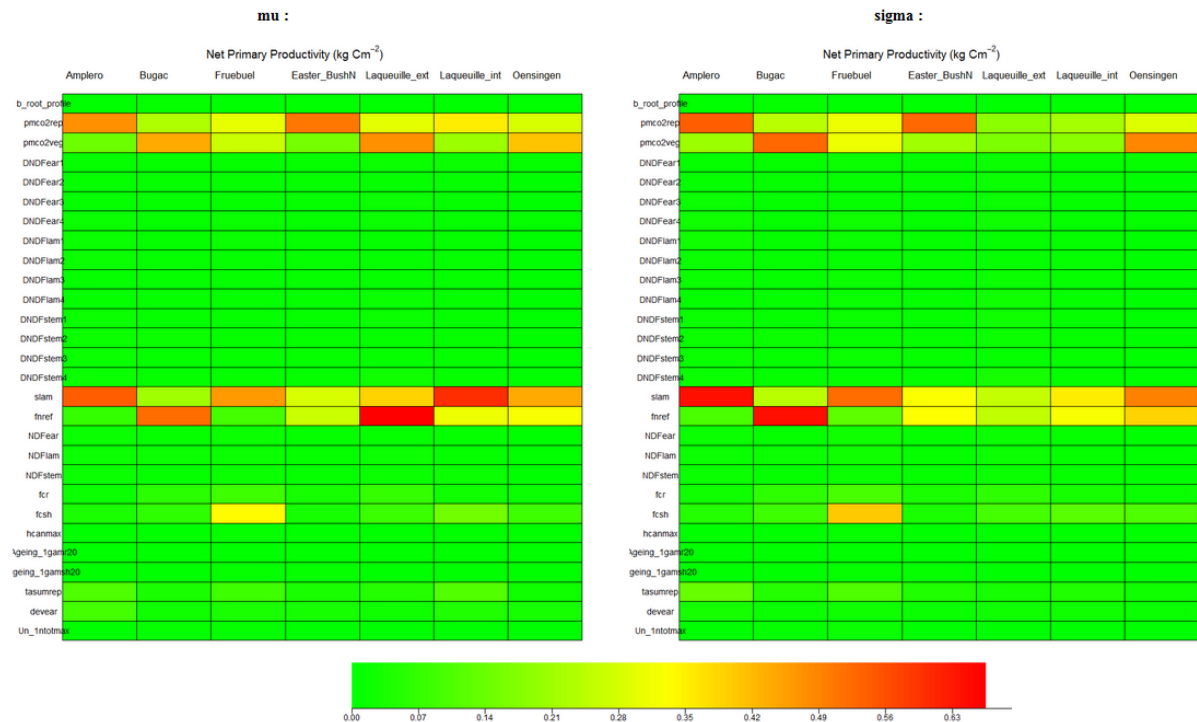
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la teneur en eau du sol à la première couche (SWC) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



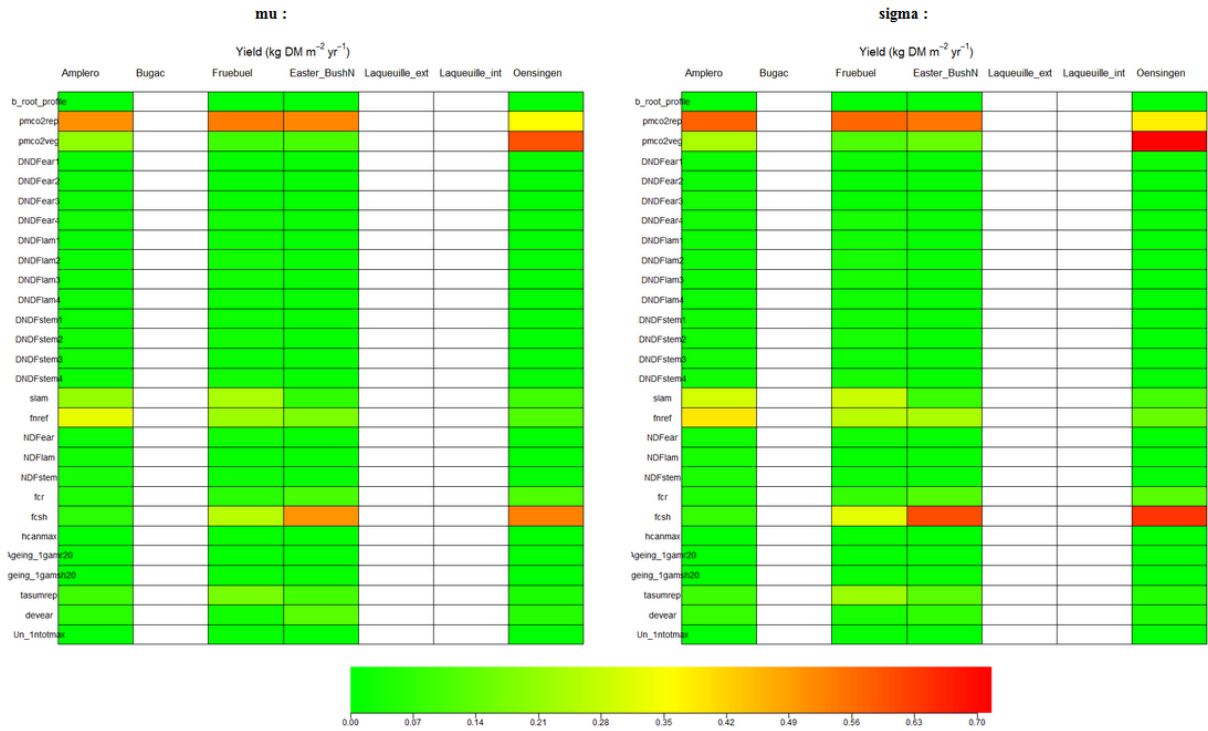
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de l'indice de surface foliaire (LAI) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) du méthane (CH₄) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



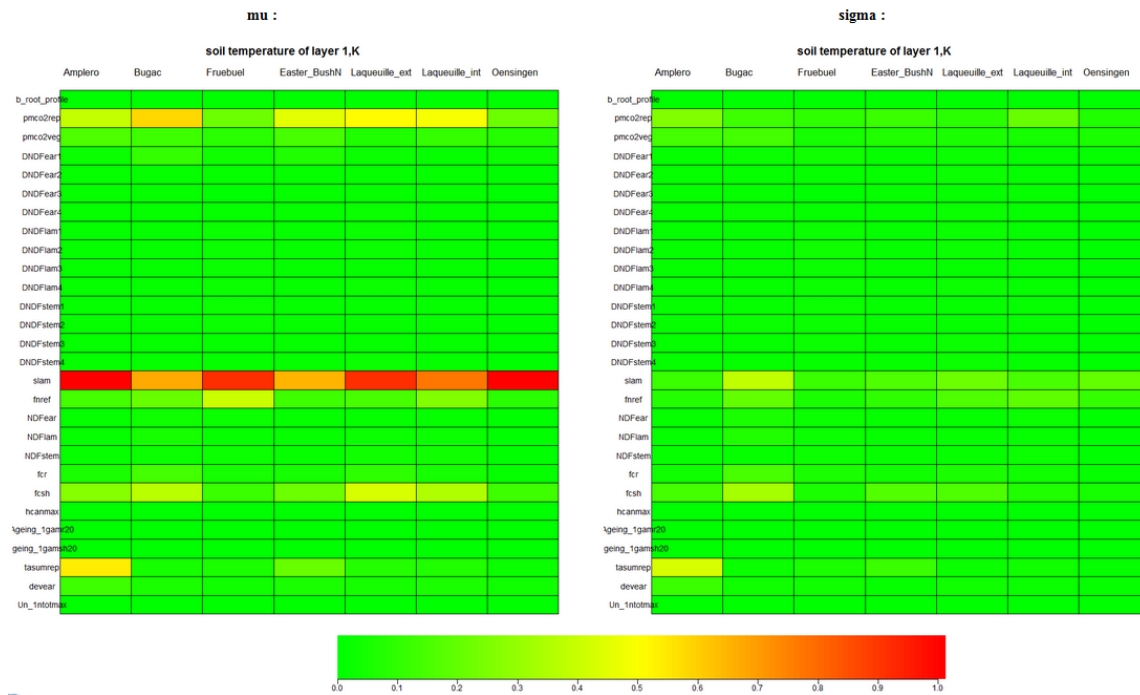
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire nette (NPP) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) du rendement (Y) pour chaque site en utilisant une distribution Gaussienne pour les valeurs des paramètres



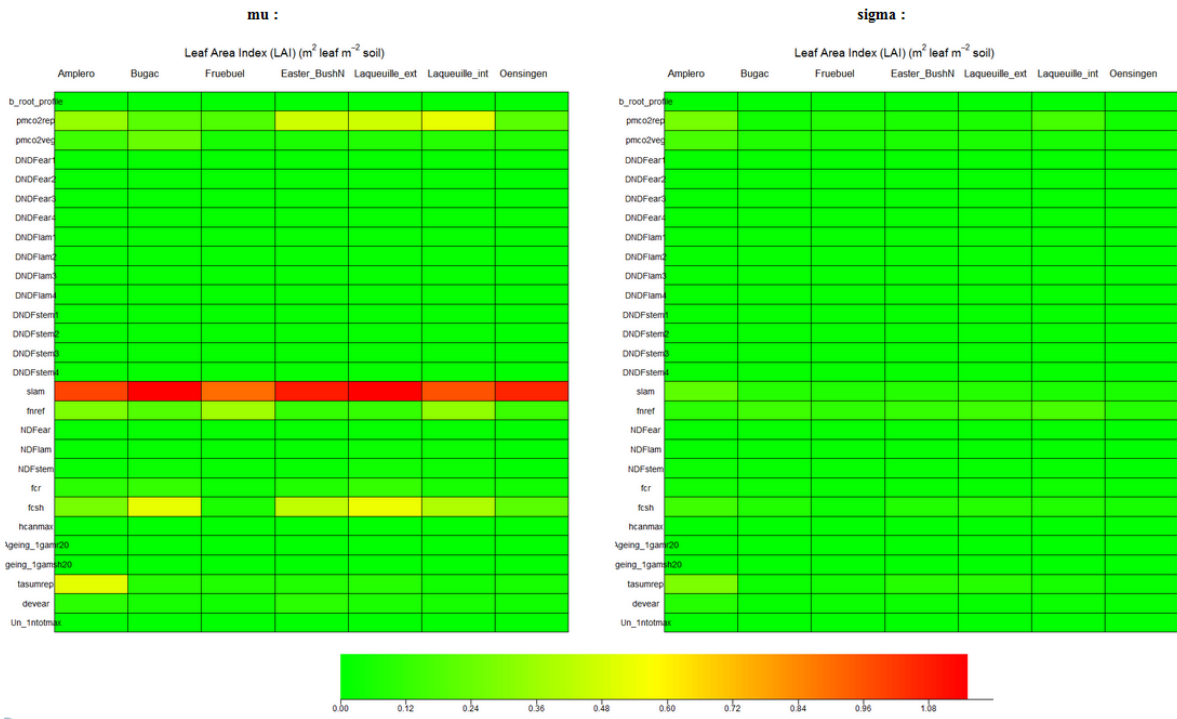
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la concentration du substrat de carbone en plante (C) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



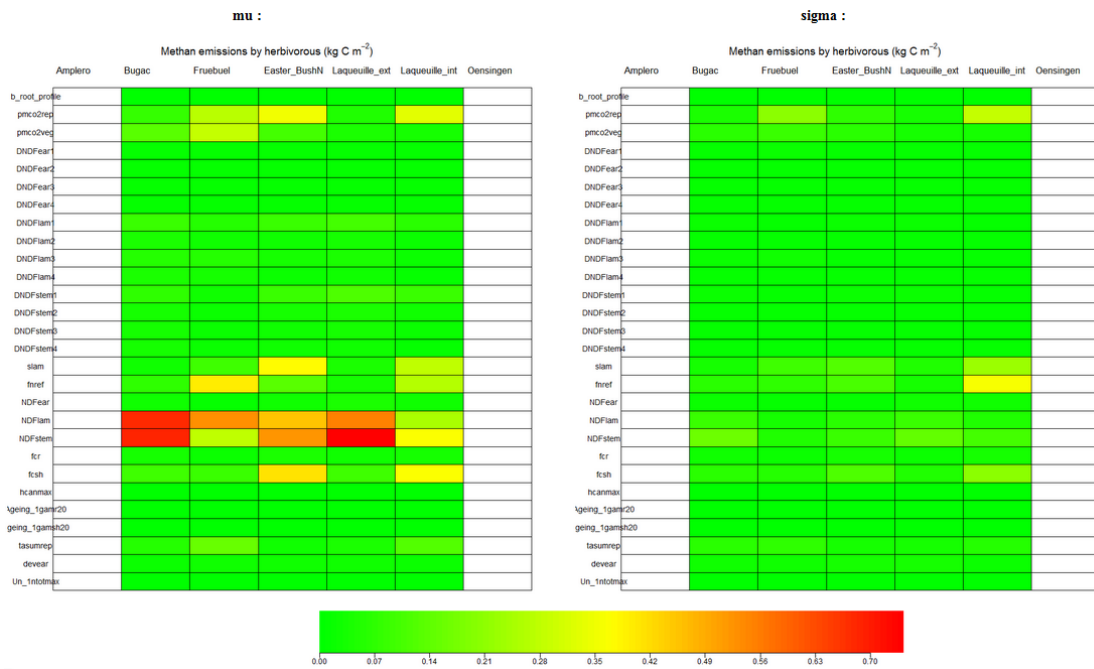
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la température du sol à la première couche (T_{soil}) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



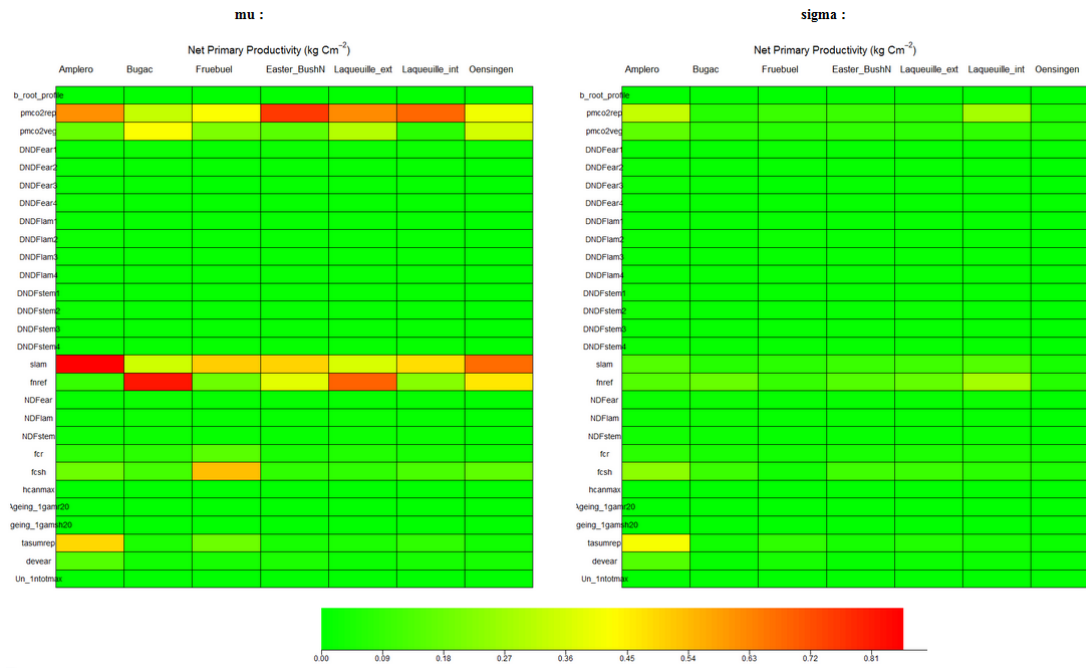
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la teneur en eau du sol à la première couche (SWC) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



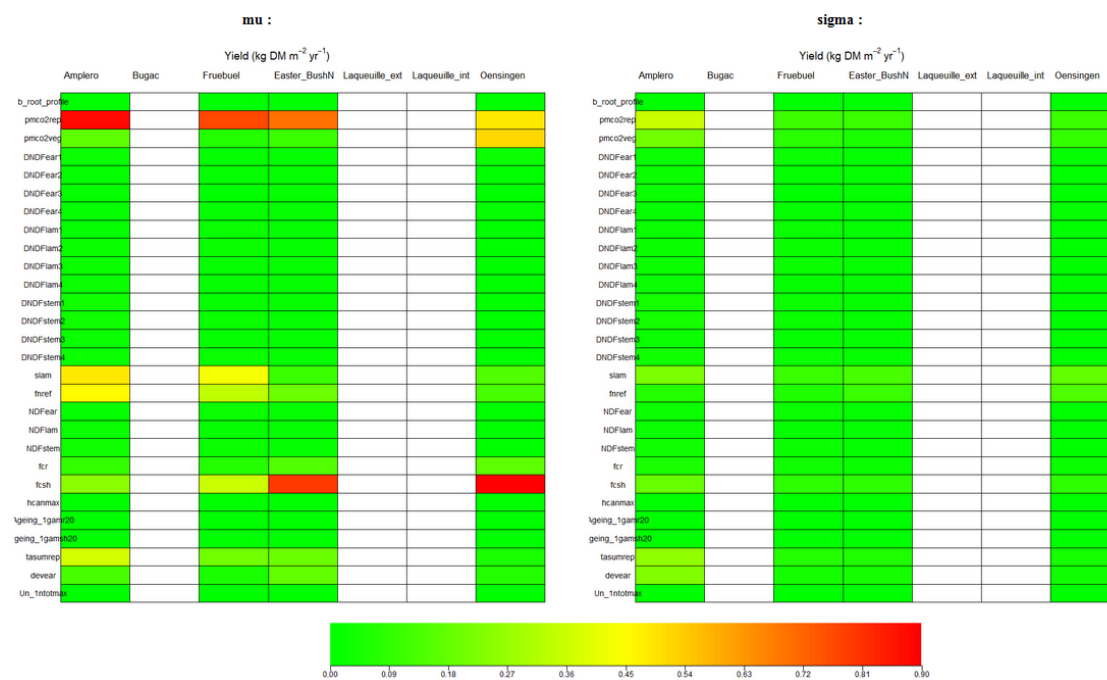
Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de l'indice de surface foliaire (LAI) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) du méthane (CH₄) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) de la production primaire nette (NPP) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres



Matrice de mesures quantitatives des résultats de l'analyse de sensibilité (μ , à gauche; σ , à droite) du rendement (Y) pour chaque site en utilisant une distribution uniforme pour les valeurs des paramètres

III. Relevant parameters into the modelling equations

The vegetation parameters identified as most important by the sensitivity analysis are documented hereafter by reproducing the main equations of PaSim in which they are used. Here are the actual formulas implemented in the source code, mostly documented in Riedo et al. (1998, 2000), Vuichard (2005) and Graux (2011) and in the seminal sources cited by these publications.

A - tasumrep: normalized temperature sum

To model grassland phenology a base temperature (T_{base} , set to 5 °C) is used above which plant development proceeds proportionally to temperature until an optimum level may be reached. A condition is the occurrence of mean soil temperature, $T_{soil,mean}$, higher than or equal to T_{base} . The temperature sum is calculated by accumulating daily differences between the mean air temperature (T_a) and T_{base} .

A developmental stage index, $devstage$, is initialized to 0 at the 1st of January (northern hemisphere) and calculated by a normalized temperature sum, as follows:

$$Devstage = \frac{\sum(T_a - T_{base})}{T_{asumrep}} \quad (A.1)$$

Where the parameter $T_{asumrep}$ (°C-d) normalizes the index in such a way that $devstage = 1$ marks the end of ear emergence, and the transition from the reproductive to the vegetative stage. The calculation is performed under the conditions of $\sum(T_a - T_{base}) \leq 2 \cdot T_{asumrep}$ and that plant development stage progresses until $devstage=2$.

During the period of the year before the growing season, the plant canopy is in the vegetative stage until the condition for the start of the new growing season is reached (mean air temperature above a given threshold, set equal to T_{base} , for a period of at least seven consecutive days). The phenologic development of the canopy is not only determined by air temperature but is also influenced by the cutting regime, and a second reproductive period can be an option during the growing season. The phenologic development is assumed to continue until the first cut, or at the latest until $devstage=2$. If the first cut occurs before $devstage=0.77$ ($devsecond$), then the development is set back to zero (which would allow a second reproductive period), otherwise the canopy is assumed to be in the vegetative development and $devstage$ is fixed at 2.

B - devear: ear developmental stage

The parameter *devear* represents an intermediate stage ($0 < devstage < 1$) between the beginning and the end of ear emergence, that is, when the vegetation reaches the stage of “ear at 10 cm”. This is a key parameter for shoot and root growth rates, and plays a role on the allocation of assimilates between aerial and root parts, as well as among aerial sub-compartments (laminas, sheaths and stems, ears). In fact, the shoot and root growth rates are based on partitioning functions, referred to as $G_{sh,veg}$ and $G_{r,veg}$ (kg DM m⁻² d⁻¹), and they influence shoot and root growth rates used, G_{sh} and G_r (kg DM m⁻² d⁻¹) according to:

$$G_{sh} = G_{sh,veg} + f_G \cdot (1 - \lambda_{rep}) \cdot G_{r,veg} \quad (B.1)$$

$$G_r = G_{r,veg} - f_G \cdot (1 - \lambda_{rep}) \cdot G_{r,veg} \quad (B.2)$$

For $devstage \geq 1$, G_{sh} is equal to $G_{sh,veg}$, and G_r is equal to $G_{r,veg}$ (i.e. $f_G=0$) while for $devstage < 1$, λ_{rep} and f_G lead to a shift in shoot/root partitioning. The term f_G is set to 1 (first $devstage < 1$) or 2/3 (second, optional $devstage < 1$). The term λ_{rep} accounts for the decrease in root partitioning, as follows:

$$\lambda_{reph} = \begin{cases} 1 - (1 - \lambda_{reph,ear}) \cdot \frac{devstage}{devear} & 0 < devstage \leq devear \\ \lambda_{reph,ear} + (1 - \lambda_{reph,ear}) \cdot \frac{devstage - devear}{1 - devear} & devear < devstage < 1 \\ 1 & otherwise \end{cases} \quad (B.3)$$

where $\lambda_{reph,ear}$ is set to a constant (the reference value being 0.25).

During the vegetative stage, the fraction f_{lam} of shoot structural growth partitioned to the lamina is given by the constant $f_{lam,veg}$, and there is no partitioning to ears. Otherwise, the fractions f_{lam} , f_{ss} and f_{ear} , representing the fractions partitioned to lamina, sheath and stem, and ears, respectively, depend on the developmental stage. The fraction f_{lam} decreases linearly from $f_{lam,a}$ at the beginning of the phenological development to the value $f_{lam,min}$ at *devear*, at which ear emergence starts, and remains constant during the rest of the reproductive stage. The fraction f_{ear} partitioned to ears is 0 until *devear* is reached. Thereafter, f_{ear} is given by:

$$f_{earh} = \max \left(f_{ear,max}; f_{ear,a} \cdot \frac{Devstage - devear}{1 - Devear} \cdot \frac{f_{ear,max} \cdot \frac{W_{ear}}{W_{lam} + W_{ss} + W_{ear}}}{f_{ear,max}} \right) \quad (B.4)$$

in which $f_{ear,max}$ is the maximum value of ear fraction of shoot structural dry matter (set equal to 0.1 kg DM m⁻²). W_{ear} , W_{lam} and W_{ss} (kg DM m⁻²) are ear, lamina and sheath/stem components of shoot structural dry matter.

C - slam: maximum specific leaf (lamina) area

The production of new biomass is accompanied by new lamina, sheath and stem area, given by $sla \cdot f_{lam} \cdot G_{sh}$ and $ssta \cdot f_{ss} \cdot G_{sh}$ where $ssta$ is specific sheath and stem area and sla is specific leaf area (m² (kg DM)⁻¹), and G_{sh} is shoot growth rate (kg DM m⁻² d⁻¹). The terms f_{lam} and f_{ss} represent the fractions partitioned to lamina and sheath + stem, respectively. The specific sheath and stem area term is set to a constant (reference value: 6.6 m² (kg DM)⁻¹), while the calculation of the specific leaf area (sla , m² (kg DM)⁻¹) is as follows:

$$sla = slam \cdot (1 - \zeta \cdot C) \quad (C.1)$$

The parameter ζ is the specific leaf area constant (set equal to 10 kg DM (kg C)⁻¹), while C is the substrate carbon per structural plant dry matter (kg C (kg DM)⁻¹) given by plant substrate C over structural dry matter (lamina + sheath/stem + ear + root) ratio.

D - pmco2rep, pmco2veg: light-saturated leaf photosynthetic rate for reproductive and vegetative stages

These parameters (hereafter indicated as $P_{m,dev}$, $\mu\text{mol m}^{-2} \text{s}^{-1}$) are defined at standard conditions of temperature (20 °C), atmospheric CO₂ concentration (350 $\mu\text{mol mol}^{-1}$) and total plant nitrogen concentration (0.035 kg N (kg DM)⁻¹). They represent the influence of developmental stage on the light saturated leaf photosynthetic rate (P_m , $\mu\text{mol m}^{-2} \text{s}^{-1}$):

$$P_m = P_{m,dev} \cdot P_{m,T} \cdot P_{m,N} \cdot P_{m,CO_2T} \quad (D.1)$$

$P_{m,T}$, $P_{m,N}$ and P_{m,CO_2T} are dimensionless factors controlling the effect of temperature, nitrogen and CO₂-temperature interaction on P_m . The latter is a component of the rate of canopy photosynthesis (P_c , $\mu\text{mol m}^{-2} \text{s}^{-1}$).

E - f_{csh} : fractional carbon content of shoot structural dry matter

This parameter (hereafter indicated as $f_{c,sh}$) multiplies the shoot structural growth rate (G_{sh} , kg DM m⁻² d⁻¹) to estimate the C substrate variation (dW_C) in a unit time step (dt) due to storage and remobilization processes (W_C , kg C m⁻² d⁻¹):

$$\frac{dW_C}{dt} = P_c - (f_{c,sh} \cdot G_{sh} + f_{c,r} \cdot G_r) - R_{plant} - F_{C,exudation} - F_{C,grazing} + F_{C,rec} + S_r \quad (E.1)$$

A similar parameter ($f_{c,r}$, kg C (kg DM)⁻¹) multiplies the root growth rate (G_r , kg DM m⁻² d⁻¹). P_c (kg C m⁻² d⁻¹) is the rate of canopy photosynthesis, given by the sum over all canopy layers of leaf photosynthetic rate, which in turn depends on the carbon concentration in structural dry matter, C (kg C (kg DM)⁻¹). A second source of carbon is represented by the flux from recycling of senescing shoot and root, $F_{C,rec}$ (kg C m⁻² d⁻¹).

The term R_{plant} (kg C m⁻² d⁻¹) is the plant respiration, represented by the respiratory fluxes associated with growth (shoot and root), maintenance and root nitrogen uptake. Other sinks are: root exudation of carbon, $F_{C,exudation}$ (kg C m⁻² d⁻¹), estimated as a fraction of plant substrate carbon, and in case of grazing, the carbon substrate removed by the animals, $F_{C,grazing}$ (kg C m⁻² d⁻¹), estimated by stocking rate of animals (animal m⁻²) and animal intake (kg C animal⁻¹ d⁻¹). S_r is the C flux (positive or negative) with plant storage.

F - f_{nref} : parameter controlling nitrogen concentration of structural dry matter

The fractional nitrogen concentration of plant structural dry matter, f_N (kg N (kg DM)⁻¹), is assumed to be constant. To allow for a changing C/N ratio of the plant structural dry matter, the nitrogen concentration of newly produced structural dry matter, $f_{N,growth}$ (kg N kg⁻¹), is expressed as:

$$f_{N,growth} = f_{N,ref} \cdot \frac{N}{N_{ref}} \quad (F.1)$$

N is substrate nitrogen per structural plant dry matter (kg N (kg DM)⁻¹) given by plant substrate N over structural dry matter (lamina + sheath/stem + ear + root) ratio. N_{ref} (set equal to 0.004 kg N (kg DM)⁻¹) is the reference N value to get $f_{N,growth}$ equal to $f_{N,ref}$.

G - NDFlam, NDFstem: fraction of fibres in ingested blades and sheaths/stems

These parameters influence the daily amount of digestible neutral detergent fibre (*DNDF*) in the animal's intake (*DNDFI*, kg *DNDF* d⁻¹), estimated as follows:

$$DNDFI = DNDF \cdot NDF \cdot I_{veg \rightarrow ani} \quad (G.1)$$

$I_{veg \rightarrow ani}$ is the amount of organic dry matter biomass ingested per animal per day (kg DM animal⁻¹ d⁻¹), given by:

$$I_{veg \rightarrow ani} = \begin{cases} I_{max} \cdot \frac{(W_{sh} - W_{sh,lim})^q}{(K - W_{sh,lim})^q + (W_{sh} - W_{sh,lim})^q} & W_{sh} > W_{sh,lim} \\ 0 & otherwise \end{cases} \quad (G.2)$$

$I_{veg \rightarrow ani}$ increases with shoot biomass (W_{sh} , kg DM m⁻²) until approaching an asymptotical value ($I_{max}=16$ kg DM m⁻²). Conversely, if shoot biomass falls below a minimum level, $W_{sh,lim}$, set equal to 0.03 kg DM m⁻², then animal grazing no longer takes place. Other parameters are: $K=0.1$ kg DM m⁻², $q=3$. *NDF* (Neutral Detergent Fiber) is the fraction of fibre in the intake and *DNDF* (Digestible Neutral Detergent Fiber) is the fraction of digestible fibre in the total fibres. Each of the components of shoot biomass (lamina, sheath/stem and ear) is divided into four state variables representing four age categories. The terms *NDF* and *DNDF* are thus the weighted means of the different leaf (*l*), sheath/stem (*s*) and ear (*e*) pools (*i*) of shoot biomass for each age class (*j*), expressed by:

$$NDF = \frac{\sum_{i=l,s,e} W_i \cdot NDF_i}{\sum_{i=l,s,e} W_i} \quad (G.3)$$

$$DNDF = \frac{\sum_{i=l,s,e} \sum_{j=1,\dots,4} W_{i,j} \cdot DNDF_{i,j}}{\sum_{i=l,s,e} \sum_{j=1,\dots,4} W_{i,j}} \quad (G.4)$$

$W_{i,j}$ (kg C m⁻²) is the amount of the pool *i* (age *j*) ingested by the animal, and $DNDF_{i,j}$ is the fraction of digestible fibres for pool *i* (age *j*). Table G1 provides a set of reference parameter values.

Table G1. Proportion of fibre in the ingested biomass (*NDF*) and fraction of digestible fibre in the total ingested fibre (*DNDF*) for lamina, sheath or ear tissues of different age (Vuichard, 2005)

Parameter	lamina				sheath/stem				ear			
	age 1	age 2	age 3	age 4	age 1	age 2	age 3	age 4	age 1	age 2	age 3	age 4
<i>DNDF</i>	0.92	0.82	0.76	0.74	0.84	0.65	0.53	0.50	0.76	0.48	0.30	0.26
<i>NDF</i>	60				70				80			

