

## Chapitre 3

# Calibrage de modèle

---

*Un écart important peut être constaté entre les performances énergétiques prévues par la simulation et les consommations réelles. Le calibrage du modèle de bâtiment à partir de relevés in situ est alors indispensable pour s'assurer de disposer d'un modèle fiable permettant de calculer les stratégies de gestion énergétique les plus performantes possibles. Ce chapitre a pour objectif de présenter une nouvelle méthodologie de calibrage de modèle. Nous en présenterons les principes généraux, avant de nous intéresser tout particulièrement au calibrage bayésien. Enfin, une nouvelle méthodologie sera exposée et testée sur un cas d'application.*

---



## PLAN DU CHAPITRE 3

<b>1.</b>	<b><u>LE CALIBRAGE DE MODÈLE</u></b>	<b>99</b>
1.1	LES LOGICIELS DE SIMULATION THERMIQUE DYNAMIQUE	99
1.2	LES STANDARDS INTERNATIONAUX	100
1.3	LES INCERTITUDES DANS LE BÂTIMENT	101
1.4	LES DIFFÉRENTES MÉTHODES DE CALIBRAGE EXISTANTES	102
1.4.1	LES MÉTHODES HEURISTIQUES	102
1.4.2	LES MÉTHODES AVANCÉES	102
1.4.3	LES MÉTHODES BAYÉSIENNES	103
<b>2.</b>	<b><u>LE CALIBRAGE BAYÉSIEN</u></b>	<b>103</b>
2.1	INTRODUCTION : L'APPROCHE STATISTIQUE	103
2.2	LES MÉTHODES MCMC	105
2.3	LES MÉTHODES MCMC UTILISANT UN MÉTAMODÈLE	106
2.4	LE CALCUL BAYÉSIEN APPROCHÉ	108
2.4.1	PRINCIPE	108
2.4.2	INTERPRÉTATION DE L'ALGORITHME DE REJET APPROXIMATIF	109
2.4.3	LES ALGORITHMES ABC-MCMC	111
2.4.4	LES ALGORITHMES ABC-SMC	111
<b>3.</b>	<b><u>MÉTHODOLOGIE</u></b>	<b>113</b>
3.1	ÉTAPE 1 : RÉCUPÉRATION DES DONNÉES	113
3.2	ÉTAPE 2 : CRÉATION DU MODÈLE	114
3.3	ÉTAPE 3 : ANALYSE DE SENSIBILITÉ	114
3.3.1	PLAN D'ÉCHANTILLONNAGE DE LA MÉTHODE DE CIBLAGE	114
3.3.2	CALCUL DES INDICES POUR L'ANALYSE DE SENSIBILITÉ	115
3.4	ÉTAPE 4 : CALIBRAGE DU MODÈLE	116
3.5	ÉTAPE 5 : ÉVALUATION DES RÉSULTATS	117
3.5.1	PROPAGATION DES INCERTITUDES	117
3.5.2	ÉVALUATION DE L'EFFICACITÉ DU CALIBRAGE	118
<b>4.</b>	<b><u>CAS D'APPLICATION</u></b>	<b>118</b>
4.1	COLLECTE DES DONNÉES (ÉTAPE 1)	118
4.1.1	PRÉSENTATION DU BÂTIMENT	118

### Chapitre 3 : Calibrage de modèle

4.1.2	PROTOCOLE EXPÉRIMENTAL .....	120
<b>4.2</b>	<b>HYPOTHÈSES DE MODÉLISATION DU BÂTIMENT (ÉTAPE 2) .....</b>	<b>122</b>
4.2.1	ZONAGE DE LA MAISON .....	122
4.2.2	DÉBITS DE VENTILATION .....	123
4.2.3	CONSIGNES DE TEMPÉRATURE .....	123
4.2.4	INFILTRATIONS .....	123
4.2.5	PONTS THERMIQUES .....	123
4.2.6	PUISSANCE DISSIPÉE PAR LES SYSTÈMES .....	123
4.2.7	OCCULTATION DES VITRAGES .....	123
4.2.8	ÉTATS DE SURFACE .....	124
4.2.9	MASQUES PROCHES .....	124
4.2.10	ALBÉDO .....	124
4.2.11	INITIALISATION DE LA SIMULATION .....	124
<b>4.3</b>	<b>ANALYSE DE SENSIBILITÉ : MÉTHODE DE MORRIS (ÉTAPE 3) .....</b>	<b>124</b>
<b>4.4</b>	<b>CALIBRAGE DU MODÈLE (ÉTAPE 4) .....</b>	<b>128</b>
<b>4.5</b>	<b>EVALUATION DES RÉSULTATS (ÉTAPE 5) .....</b>	<b>132</b>
<b>5.</b>	<b><u>CONCLUSION DU CHAPITRE .....</u></b>	<b><u>138</u></b>

# 1. LE CALIBRAGE DE MODÈLE

## 1.1 Les logiciels de simulation thermique dynamique

La simulation numérique est un outil utilisé dans de nombreux domaines de recherche et développement tels que la mécanique, la mécanique des fluides ou en encore la climatologie. Elle a pour objectif de représenter un phénomène physique réel et permet d'étudier son fonctionnement ainsi que de prédire son évolution.

Dans le cadre du bâtiment, les outils de simulation étaient initialement utilisés lors de la phase de conception. Les logiciels de simulation thermique dynamique permettaient alors d'optimiser l'enveloppe et de dimensionner les équipements du bâtiment.

Cependant, dans le contexte du suivi et de l'évaluation des performances énergétiques des bâtiments existants, leur utilisation s'est étendue à d'autres phases de vie du bâtiment. Ces logiciels peuvent ainsi être employés pour réaliser des audits énergétiques, garantir la performance énergétique des bâtiments, mettre au point les systèmes de régulation ou encore évaluer économiquement et techniquement les rénovations énergétiques envisagées (Coakley et al. 2011).

Les modèles utilisés dans les logiciels de simulation énergétique des bâtiments étant des modèles physiques (modèle de type boîte blanche), ils présentent l'avantage de permettre des variations paramétriques. En particulier, l'utilisateur peut modifier l'enveloppe ou le fonctionnement d'un bâtiment et évaluer directement l'impact de ces modifications sur les performances et le comportement de ce bâtiment.

Cependant, le développement d'un modèle fiable nécessite de connaître précisément le bâtiment à modéliser (géométrie du bâtiment, systèmes installés), et de renseigner de nombreux paramètres (qui peuvent être difficilement accessibles ou mesurables). Ainsi, plusieurs études ont souligné les différences qui peuvent exister entre les performances prévues par le modèle et les mesures sur le terrain (Ahmad et Culp 2006 ; C. Turner et Frankel 2008). En effet, étant donné la complexité des bâtiments et les nombreux paramètres incertains à renseigner, il est difficile d'obtenir une modélisation précise du comportement réel de ces bâtiments. Il devient ainsi indispensable de calibrer ces modèles pour assurer la fiabilité de leurs prédictions. En conciliant mesures in-situ et sorties du modèle, l'objectif est alors d'obtenir une représentation plus fiable et plus précise du comportement réel du bâtiment. Plus précisément, il s'agit de résoudre le problème inverse suivant : « à partir de données d'observations, estimer les paramètres du modèle qui expliquent ces données d'observations ».

Le calibrage de modèle présente plusieurs difficultés. Tout d'abord, le problème inverse posé est fortement sous-déterminé, c'est-à-dire que sa résolution amène à une multitude de solutions au lieu d'une solution unique. De plus, le calibrage des logiciels de simulation énergétique nécessite

l'identification de centaines de paramètres. Enfin, il n'existe pas une méthodologie clairement établie et suivie par l'ensemble de la communauté. Ainsi, les approches où l'utilisateur règle manuellement les différents paramètres jusqu'à ce qu'un critère de précision soit atteint sont couramment utilisées (Coakley et al. 2014). Cependant, différents standards internationaux suggèrent l'utilisation du calibrage de modèle et proposent quelques bonnes pratiques à suivre.

## 1.2 Les standards internationaux

Plusieurs standards internationaux fournissent les orientations à suivre pour mettre en place un plan de mesure et de vérification de la performance énergétique d'un bâtiment. On peut citer par exemple :

- le protocole IPMVP (*International Performance Measurement and Verification Protocol*) (EVO 2009) ;
- l'ASHRAE Guideline 14-2002 (ASHRAE 2002).

Ces protocoles proposent, entre autres, l'utilisation de la simulation numérique pour calculer les gains réalisés suite à la mise en œuvre de solutions d'efficacité énergétique. Par exemple, l'option D du protocole IPMVP suggère l'utilisation d'un modèle de bâtiment calibré sur des données mesurées pour évaluer la performance de ces solutions.

L'emploi d'un modèle de simulation nécessite l'utilisation d'indicateurs pour évaluer sa fiabilité par rapport à des données mesurées. Les indicateurs couramment utilisés dans ce contexte par les standards internationaux sont les suivants :

- le biais moyen *MBE* (*Mean Biased Error*) (%) : entre les données mesurées et les données simulées.

$$MBE(\%) = \frac{\sum_{i=1}^N (m_i - s_i)}{\sum_{i=1}^N (m_i)} \quad (3.1)$$

où  $m_i$  et  $s_i$  sont respectivement les données mesurées et simulées à chaque instant  $i$  et  $N$  le nombre de pas de temps considérés ( $N = 12$  si l'on souhaite calibrer le modèle sur des données mensuelles et  $N = 8760$  si l'on utilise des données horaires).

- le coefficient de variation de l'erreur quadratique moyenne *CVRMSE* (*Coefficient of Variation of Root Mean Square Error*) : permettant d'évaluer l'ajustement du modèle aux données mesurées sans la compensation des erreurs.

$$CVRMSE(\%) = \frac{\sqrt{(\sum_{i=1}^N (m_i - s_i)^2 / N)}}{\bar{m}} \quad (3.2)$$

où  $\bar{m}$  est la moyenne des données mesurées.

### Chapitre 3 : Calibrage de modèle

L'ASHRAE Guideline 14-2002 fournit les critères d'acceptation à atteindre par ces indicateurs pour considérer le modèle comme calibré. Ils sont résumés dans le tableau suivant. Cependant, ces critères ne sont pas applicables lorsque l'on souhaite calibrer le modèle à partir de données de température mesurées.

Tableau 8 : Critères d'acceptation pour le calibrage des modèles

Critères d'acceptation	Critère mensuel (%)		Critère horaire (%)	
	<i>MBE</i>	<i>CVRMSE</i>	<i>MBE</i>	<i>CVRMSE</i>
ASHRAE Guideline 14-2002	5	15	10	30

Il est à noter que le modèle calibré respectant ces critères n'est pas unique et que, par conséquent, il peut exister plusieurs modèles d'un même bâtiment considérés comme calibrés.

## 1.3 Les incertitudes dans le bâtiment

Les sorties des logiciels de simulation thermique dynamique sont soumises à de nombreuses incertitudes qu'il est indispensable de prendre en compte. Ainsi, de nombreuses études se sont portées sur l'évaluation de l'impact des incertitudes sur les sorties des logiciels de simulation thermique dynamique (Lomas et Eppel 1992 ; Macdonald 2002). Cette évaluation permet d'expliquer les différences pouvant exister entre la simulation et les mesures. Par exemple, Spitz (2012) dresse dans sa thèse un aperçu des différentes sources d'incertitudes qui peuvent expliquer ce biais. Par exemple, lors de la phase de conception, de nombreuses incertitudes portent sur le choix du fichier météorologique utilisé, ou encore sur la définition des scénarios d'occupation. D'autres sources d'erreurs sont également possibles lors de la phase de mise en œuvre. Par exemple, une différence peut exister entre les caractéristiques physiques des matériaux choisis lors de la phase de conception et celles des matériaux réellement utilisés. De même, une mauvaise mise en œuvre peut amener à des ponts thermiques plus importants que prévus. Enfin, lors de la phase d'exploitation, un mauvais fonctionnement des systèmes peut également amener à une détérioration des performances.

Ces sources d'incertitudes peuvent être classées en quatre grandes catégories (de Wit et Augenbroe 2002; Coakley et al. 2014) :

- Incertitudes de spécification (des matériaux, des systèmes, de la géométrie) correspondant au manque d'informations sur les propriétés exactes des matériaux utilisés, sur la géométrie du bâtiment ou encore sur le fonctionnement des systèmes énergétiques.
- Incertitudes de modélisation survenant lors du développement du modèle à partir des hypothèses, approximations et simplifications considérées (zonage, simplification de la géométrie, phénomènes physiques simplifiés ou négligés, etc.).
- Incertitudes de simulation introduites lors de la discrétisation et de la résolution numérique du modèle.

- Incertitudes sur les sollicitations et les scénarios correspondant à la définition des différents scénarios d'usage ainsi que des conditions environnementales.

## 1.4 Les différentes méthodes de calibrage existantes

De nombreuses méthodes de calibrage existent dans la littérature. Il est possible de les classer en trois grandes familles (Heo et al. 2014). Un état de l'art complet sur les différentes méthodes existantes dans la littérature a été réalisé par Coakley et al. (2014)

### 1.4.1 Les méthodes heuristiques

Les méthodes heuristiques correspondent aux méthodes ne prenant pas en compte les incertitudes existant dans les logiciels de simulation thermique dynamique. L'objectif de ces méthodes consiste à calibrer manuellement les différents paramètres du modèle jusqu'à respecter un critère de précision désiré (Pedrini et al. 2002). Le processus de calibrage nécessite une expertise non négligeable de l'utilisateur pour sélectionner les paramètres à calibrer et les valeurs des paramètres à tester. Par exemple, ces méthodes peuvent reposer sur une analyse graphique (utilisation d'histogrammes) pour comparer les résultats de la simulation aux mesures (Bou-Saada et Haberl 1995 ; McCray et al. 1995).

### 1.4.2 Les méthodes avancées

Des études plus récentes prennent en compte les incertitudes des paramètres du modèle dans le processus de calibrage. Les paramètres à calibrer sont alors affectés d'une borne d'incertitudes et d'une estimation la plus probable de leur valeur. Des techniques de type Monte-Carlo sont ensuite utilisées pour propager ces incertitudes et déterminer un ensemble de solutions respectant un critère de précision donné. Ainsi, le processus de calibrage ne calcule plus une solution unique pour les différents paramètres à calibrer, mais un ensemble de solutions (Reddy et al. 2007). Des techniques d'analyse de sensibilité peuvent également être utilisées pour déterminer les paramètres les plus influents dans le modèle, réduire le nombre de paramètres à calibrer et ainsi affiner leurs distributions de probabilités. Ces méthodes sont indispensables par exemple pour évaluer avec plus de confiance les solutions de rénovation énergétique envisagées.

Plus récemment, Raftery et al. (2011) ont suggéré une méthodologie appelée « *Evidence-based calibration* » qui consiste en un processus de calibrage à partir de données factuelles. Toute modification d'un paramètre doit alors être clairement justifiée dans l'objectif d'améliorer le comportement du modèle. Ainsi, les différentes sources d'information utilisées pour calibrer les paramètres sont hiérarchisées selon leurs origines. Par exemple, une information obtenue à partir d'une mesure sera jugée plus fiable qu'une information tirée d'une documentation. Pour assurer la reproductibilité des résultats, les modifications réalisées lors du processus de calibrage doivent être indiquées et conservées en mémoire (à l'aide d'un logiciel de gestion de versions par exemple). De

nombreuses études utilisant cette méthode ainsi que des analyses de sensibilité et d'incertitudes peuvent être trouvées dans la littérature (Coakley et al. 2011 ; Bertagnolio et al. 2012 ; Coakley et al. 2012).

#### 1.4.3 Les méthodes bayésiennes

Les méthodes avancées ont l'inconvénient de ne pas considérer l'ensemble des sources d'incertitudes et d'erreurs pouvant expliquer les différences entre les données simulées et les données mesurées. En effet, les erreurs de mesures et les erreurs de modélisation et de simulation ne sont pas considérées dans le processus de calibrage. Une alternative consiste à utiliser des méthodes bayésiennes. Ces dernières ont été employées dans de nombreux domaines pour calibrer des modèles complexes (Pritchard et al. 1999 ; A. E. Raftery et al. 2005 ; Van Oijen et al. 2005 ; Wood 2010). Dans le contexte du bâtiment, l'application du calibrage bayésien à des logiciels de simulation thermique dynamique a été formalisée par Heo et al. (2012). La présentation de ces méthodes fait l'objet de la partie suivante.

## 2. LE CALIBRAGE BAYÉSIEN

### 2.1 Introduction : l'approche statistique

Le calibrage bayésien s'inscrit dans le cadre de l'approche statistique du calibrage où les incertitudes pesant sur les paramètres à calibrer sont modélisées par des distributions de probabilité *a priori* (ou PDF pour *probability density function*). L'objectif est alors d'affiner ces distributions de probabilité en comparant les sorties du modèle numérique avec des mesures *in situ*. Des distributions de probabilité *a posteriori* des paramètres sont alors estimées. Cette approche statistique se différencie de l'approche classique (ou non statistique) où l'objectif est d'estimer une valeur unique pour les différents paramètres à calibrer (par la méthode des moindres carrés par exemple).

Dans un processus de calibrage, il est important de distinguer les variables d'entrée des paramètres du modèle (Campbell 2006). Les variables d'entrée, notées «  $u$  », correspondent à des données connues (ou au moins observables) et variables dans le temps. Ces variables permettent de décrire et de modéliser différents scénarios observés (par exemple des scénarios météorologiques). Les paramètres du modèle, notés «  $\theta$  », correspondent quant à eux à des données inconnues mais constantes, et sont modélisés par des distributions de probabilité *a priori* représentant l'incertitude épistémique (dû à un manque de connaissance) pesant sur ces paramètres. Ces derniers sont ainsi supposés invariables sur l'ensemble de la simulation. Cette distinction n'est pas importante dans une analyse de sensibilité mais indispensable dans un processus de calibrage où seuls les paramètres du modèle, et non les variables d'entrée, sont à calibrer.

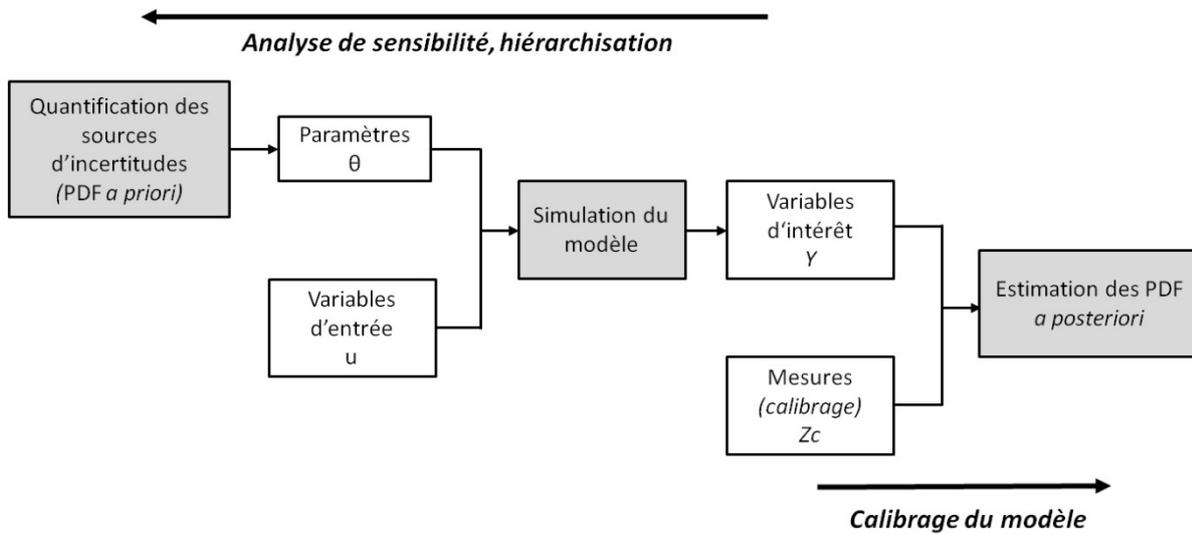


Figure 25 : Représentation de l'analyse de sensibilité et du calibrage de modèle

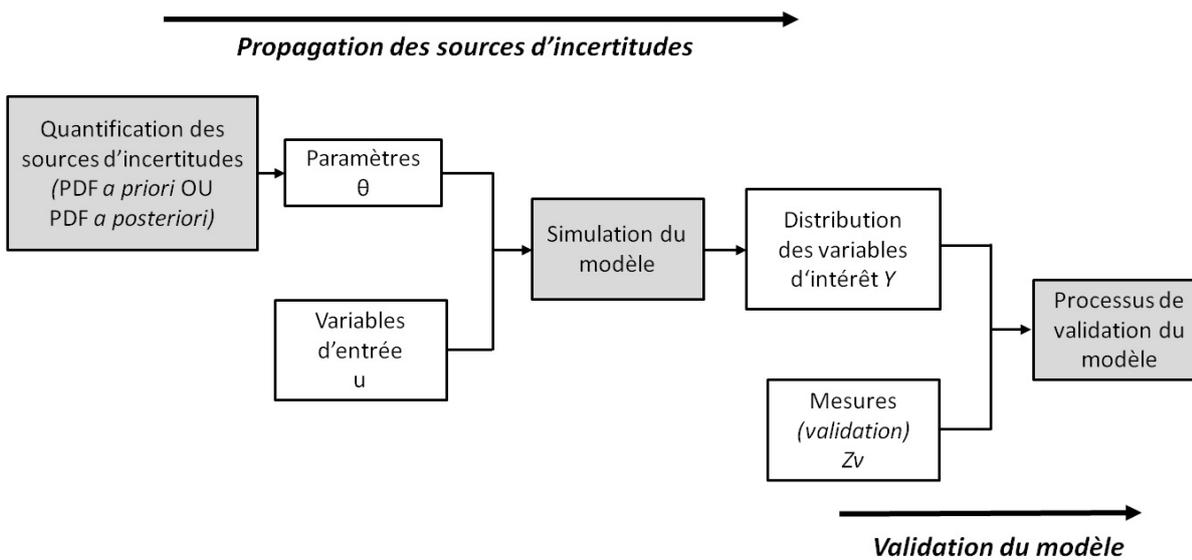


Figure 26 : Représentation de la propagation des incertitudes et de la validation de modèle

De nombreuses méthodes, intimement liées, peuvent être utilisées pour étudier un modèle numérique. Une étape importante correspond à l'analyse de sensibilité (Figure 25). Son objectif est d'identifier les paramètres du modèle et les variables d'entrée qui ont une influence forte sur les sorties du modèle. Cette analyse permet ainsi de hiérarchiser les différents paramètres et variables et d'orienter, par exemple, les campagnes de mesures à mener (il est inutile de mener une campagne de mesures, potentiellement longue et coûteuse, sur un paramètre non influent).

Ensuite, la méthode de calibrage permet de mettre à jour (dans le contexte bayésien) les fonctions de densité de probabilité des paramètres «  $\theta$  » à partir d'une comparaison entre les sorties

du modèle et les données expérimentales (observations «  $Z$  »). Le processus minimise les écarts dans l'objectif d'améliorer les capacités de prédiction du modèle.

Enfin, une troisième méthode, liée à l'analyse de sensibilité, correspond à l'analyse d'incertitudes. Cette dernière permet, à travers la propagation des différentes sources d'incertitudes, d'estimer la distribution des variables d'intérêt «  $Y$  » et de donner des informations sur l'incertitude associée aux prédictions du modèle. Cette méthode est utilisée par exemple dans le cadre plus large de la validation de modèle consistant à quantifier la « confiance » qu'un utilisateur peut avoir dans les capacités prédictives de son modèle en comparaison à un ensemble de données expérimentales. On mesure alors la capacité du modèle numérique à représenter de manière adéquate les phénomènes physiques mesurés.

Dans le processus de calibrage, la mise à jour des distributions de probabilité est réalisée selon le théorème de Bayes qui stipule que la distribution de probabilité *a posteriori*,  $p(\theta|Z)$ , est proportionnelle au produit de la distribution de probabilité *a priori*,  $p(\theta)$ , et de la fonction de vraisemblance,  $p(Z|\theta)$  :

$$p(\theta|Z) = \frac{p(\theta) \times p(Z|\theta)}{p(Z)} \quad (3.3)$$

Le dénominateur  $p(Z)$ , indépendant de  $\theta$ , peut être interprété comme un facteur permettant de normer  $(\theta|Z)$ . La fonction de vraisemblance permet ici de caractériser les sorties du modèle en fonction des valeurs prises par les paramètres «  $\theta$  ». Elle mesure à quel point les sorties du modèle sont proches des observations «  $Z$  » en fonction des paramètres testés. Ainsi, la probabilité d'obtenir les observations à partir du modèle pilote le processus de mise à jour des fonctions de densité de probabilité. Différentes méthodes peuvent être employées pour estimer la distribution *a posteriori*,  $p(\theta|Z)$  et se distinguent selon la connaissance de la fonction de vraisemblance  $p(Z|\theta)$ .

Tout d'abord, si la fonction de vraisemblance est connue analytiquement, alors des méthodes de Monte-Carlo par Chaînes de Markov ou méthodes MCMC pour *Markov Chain Monte Carlo* peuvent être utilisées (§2.2). Ensuite, si elle n'est pas connue analytiquement mais qu'une approximation analytique peut être obtenue, des méthodes MCMC utilisant des métamodèles peuvent être employées (§2.3). Enfin, si l'obtention d'une expression analytique de la fonction de vraisemblance est impossible, des méthodes basées sur le calcul bayésien approché (ou ABC pour *Approximate Bayesian Computation*) existent (§2.4).

## 2.2 Les méthodes MCMC

Les méthodes de Monte-Carlo par chaînes de Markov sont couramment utilisées pour estimer la distribution *a posteriori* des paramètres  $p(\theta|Z)$  sans avoir à déterminer sa forme analytique. Ces méthodes permettent de générer des échantillons à partir d'une loi de probabilité dont la forme n'est pas connue complètement de manière analytique (la loi est connue à une constante près). Plus précisément dans l'équation (3.3), la distribution de probabilité *a priori*,  $p(\theta)$ , et la fonction de

vraisemblance,  $p(Z|\theta)$  sont connues, tandis que le dénominateur  $p(Z)$  est inconnu mais peut être considéré comme une constante puisque ne dépendant pas de  $\theta$ .

Dans le cadre de l'estimation bayésienne, un algorithme MCMC est une méthode itérative qui propose de construire une chaîne de Markov convergeant vers la loi *a posteriori*,  $p(\theta|Z)$ . Une chaîne de Markov est un processus stochastique qui vérifie les conditions de Markov selon lesquelles toute l'information nécessaire pour prévoir l'état futur est contenue dans l'état présent du processus. Il existe de nombreux algorithmes MCMC dont le plus couramment utilisé est l'algorithme de Metropolis-Hastings. La première étape consiste à tirer une valeur initiale  $\theta_0$  de la distribution *a priori*,  $p(\theta)$ . On va ensuite tirer aléatoirement un candidat  $\theta^*$  dans une distribution  $q(\cdot|\theta_0)$  conditionnée par  $\theta_0$  (par exemple une loi normale centrée en  $\theta_0$  et de variance  $\sigma^2$ ). Selon une probabilité dépendant de la vraisemblance (appelé ratio d'acceptation-rejet, ou probabilité d'acceptation), le candidat  $\theta^*$  est accepté (on pose alors  $\theta_1 = \theta^*$ ) ou rejeté (on pose alors  $\theta_1 = \theta_0$ ). Cette procédure est enfin répétée jusqu'à obtenir une chaîne de valeur  $\{\theta_0, \theta_1, \dots, \theta_n\}$  qu'on peut supposer être un échantillon de la distribution *a posteriori*,  $p(\theta|Z)$ .

Les algorithmes MCMC peuvent être très efficaces pour estimer la distribution *a posteriori*,  $p(\theta|Z)$ , mais ils nécessitent la connaissance de la fonction de vraisemblance, ce qui est difficile voire impossible dans le cas des modèles numériques complexes (comme les logiciels de simulation thermique dynamique). L'utilisation d'un métamodèle pour approximer l'expression analytique de la fonction de vraisemblance peut alors être une solution.

## 2.3 Les méthodes MCMC utilisant un métamodèle

Le calibrage bayésien par métamodèle repose sur l'approche développée par Kennedy et O'Hagan (2001) permettant de relier les observations aux sorties du modèle. Cette approche prend en compte trois types d'erreurs : les incertitudes pesant sur les paramètres du modèle, l'écart existant entre les sorties du modèle et la réalité, et les erreurs de mesure. Ainsi, la relation entre les observations et les sorties du modèle dans les conditions connues  $u$  peut être modélisée par :

$$Z = \eta(u, \theta) + \delta(u) + e(u) \quad (3.4)$$

où  $Z$  sont les observations,  $\eta(u, \theta)$  correspond aux sorties du modèle calculées avec les conditions connues  $u$  et les paramètres de calibrage  $\theta$ , et  $e(u)$  l'erreur de mesure. L'erreur de modèle  $\delta(u)$  permet de modéliser l'écart existant entre les sorties du modèle et la réalité. En effet, les logiciels de simulation (et en particulier les logiciels de simulation thermique dynamique) sont basés sur des approximations des phénomènes physiques et sont donc incapables de reproduire fidèlement la réalité même en balayant l'ensemble des valeurs de  $\theta$  possibles. Par conséquent, l'erreur de modèle est indépendante des paramètres de calibrage. L'objectif du calibrage est de mettre à jour les fonctions de densité de probabilité *a priori* des paramètres  $\theta$  en comparant les prédictions du modèle et les observations (Campbell 2006).

### Chapitre 3 : Calibrage de modèle

Une remarque peut être faite quant aux méthodes avancées présentées dans le §1.4.2. Ces dernières comparent les sorties du modèle avec les observations pour définir une distribution de solutions concernant les paramètres à calibrer. Ces méthodes peuvent alors être considérées comme des méthodes de calibrage statistique. En effet, même si elles ne définissent pas clairement la relation qui existe entre les observations, la réalité et les sorties du modèle, cela ne signifie pas qu'il n'existe pas un tel modèle implicite. Généralement, le modèle correspond à l'équation (3.4) sans le modèle d'erreur  $\delta(u)$  (ou de manière équivalente, en considérant que l'espérance de  $\delta(u)$  pour tout  $u$  est nulle). Ainsi, le modèle sous-jacent est en général :

$$Z = \eta(u, \theta) + e'(u) \quad (3.5)$$

où  $e'(u)$  ne correspond plus simplement à une erreur de mesure, mais à une erreur de moyenne nulle incluant d'autres sources d'erreur.

Dans l'objectif d'estimer la fonction de vraisemblance, Kennedy et O'Hagan (2001) proposent d'utiliser des processus gaussien pour modéliser les sorties du modèle  $\eta(u, \theta)$  et l'erreur de modèle  $\delta(u)$ . Un processus aléatoire  $f(\cdot)$  est un processus gaussien si la distribution jointe de  $f(s_1), \dots, f(s_n)$  est une loi normale multidimensionnelle. Un processus gaussien a de plus la propriété d'être entièrement caractérisé par sa moyenne et sa fonction de covariance :

$$f(s) \sim PG(m(s), c(s, s')) \quad (3.6)$$

où  $m(s) = E[f(s)]$  et  $c(s, s') = Cov(f(s), f(s'))$ .

Ainsi la sortie du logiciel de simulation énergétique sous des conditions connues et des paramètres de calibrage choisis est supposée suivre une distribution normale (bien qu'elle soit déterministe) (Heo et al. 2014). Les sorties du logiciel sous différentes conditions suivent alors une loi normale multidimensionnelle avec une structure de covariance spécifique. Il en est de même pour la modélisation de l'erreur de modèle. La définition de ces processus gaussien implique l'utilisation d'hyper-paramètres dans les matrices de covariance qui doivent également être identifiés. Les détails sur la définition des structures de covariance et l'identification des hyper-paramètres sont résumés dans Kennedy et O'Hagan (2001). Cette formulation mathématique permet alors de calculer la fonction de vraisemblance  $p(Z|\theta)$  pour n'importe quelle valeur de  $\theta$  testée (Riddle et Muehleisen 2014). En utilisant le théorème de Bayes, il est ainsi possible de mettre à jour les distributions *a priori* des paramètres à calibrer en prenant en compte l'erreur de modèle et les erreurs de mesure. Comme la distribution *a posteriori* ne peut être calculée complètement analytiquement ( $p(Z)$  est inconnue), des méthodes de Monte-Carlo par chaîne de Markov (§2.2) peuvent être utilisées pour approximer numériquement cette distribution (Campbell 2006 ; Heo et al. 2012 ; Riddle et Muehleisen 2014).

Dans le contexte du bâtiment, cette méthodologie a été appliquée avec succès par Heo dans ses travaux de thèse (Heo 2011). L'objectif était de développer une méthodologie pour calibrer les logiciels de simulation énergétique des bâtiments en limitant l'utilisation du jugement expert dans le processus de calibrage. Appliquée dans le cadre de l'évaluation de l'efficacité des solutions de rénovation énergétique, cette méthodologie repose sur les étapes suivantes (Heo et al. 2014) :

### Chapitre 3 : Calibrage de modèle

- quantifier les sources d'incertitudes et déterminer les distributions *a priori* des paramètres à calibrer ;
- réaliser un criblage pour déterminer les paramètres les plus influents ;
- explorer l'espace des entrées des paramètres à calibrer et générer les sorties associées ;
- estimer les distributions *a posteriori* des paramètres par le calibrage bayésien ;
- valider statistiquement le modèle calibré.

Les auteurs ont démontré l'intérêt de l'utilisation du calibrage bayésien pour la prise en compte des incertitudes dans le processus de calibrage. Cette méthodologie permet d'automatiser le processus, de quantifier le risque lié à chaque décision de rénovation énergétique et de limiter l'utilisation du jugement expert. Cependant, ce processus de calibrage repose sur une formulation mathématique sophistiquée, nécessitant des développements informatiques conséquents et un changement de paradigme par rapport aux méthodes de calibrage classiques. De plus, cette méthode ne s'applique qu'à des données énergétiques agrégées (comme des consommations mensuelles) qui ne tiennent pas compte des corrélations temporelles dans les données (Heo et al. 2014). Par conséquent, son extension à des mesures horaires de température n'est pour l'instant pas possible. Or, dans le cadre de la régulation, nous souhaitons améliorer la précision de notre modèle concernant la simulation des températures horaires des zones de notre bâtiment. Une alternative consiste alors à utiliser une autre méthode, le calcul bayésien approché (ou ABC), utilisant uniquement les simulations numériques pour réaliser l'inférence et pouvant s'appliquer à des séries temporelles.

## 2.4 Le calcul bayésien approché

### 2.4.1 Principe

Le calcul bayésien approché (ABC) rassemble l'ensemble des algorithmes développés ne nécessitant pas l'évaluation explicite de la fonction de vraisemblance pour réaliser l'inférence bayésienne (*likelihood-free inference*). Ces méthodes sont particulièrement intéressantes dans le cadre des logiciels de simulation numérique où la fonction de vraisemblance n'est pas connue. Originellement développé par Pritchard et al. (1999), le calcul bayésien approché permet de réaliser l'inférence bayésienne en n'utilisant que les simulations générées par le modèle sans requérir à l'évaluation de la fonction de vraisemblance.

Le premier algorithme ABC se base sur l'algorithme de rejet et se comporte comme suit (algorithme de rejet approximatif) :

1. tirer aléatoirement  $\theta^*$  de la distribution a priori  $p(\theta)$  ( $\theta^* \sim p(\theta)$ ) ;
2. simuler le modèle en fonction de  $\theta^*$  et générer les variables d'intérêt  $Y$  ( $Y \sim \eta(u, \theta^*)$ ) ;

### Chapitre 3 : Calibrage de modèle

3. accepter  $\theta^*$  si la distance entre les données simulées  $Y$  et les données observées  $Z$  est inférieure à un seuil cible  $\epsilon_c$  ( $\rho(Y, Z) \leq \epsilon_c$ ).

La Figure 27 illustre le fonctionnement de l'algorithme de rejet approximatif. Il est important de noter ici que l'ensemble des paramètres acceptés  $\theta^*$  n'appartiennent pas à la véritable distribution de probabilité *a posteriori*  $p(\theta|Z)$  mais à une approximation de cette dernière  $p(\theta|\rho(Y, Z) \leq \epsilon_c)$ . Lorsque  $\epsilon_c = 0$ , cet algorithme donne des résultats exacts et permet de tirer des paramètres dans la véritable distribution *a posteriori*  $p(\theta|Z)$ . A l'opposé, si  $\epsilon_c \rightarrow +\infty$ , les paramètres sont alors tirés dans la distribution *a priori*  $p(\theta)$ . Ainsi, la distribution *a posteriori* sera d'autant mieux approximée que la valeur de  $\epsilon_c$  sera petite, mais engendrera un taux de rejet plus important à l'étape 3 de l'algorithme et donc des temps de calcul plus importants pour une taille d'échantillon voulue. Le seuil cible  $\epsilon_c$  représente donc un compromis entre précision et temps de calcul.

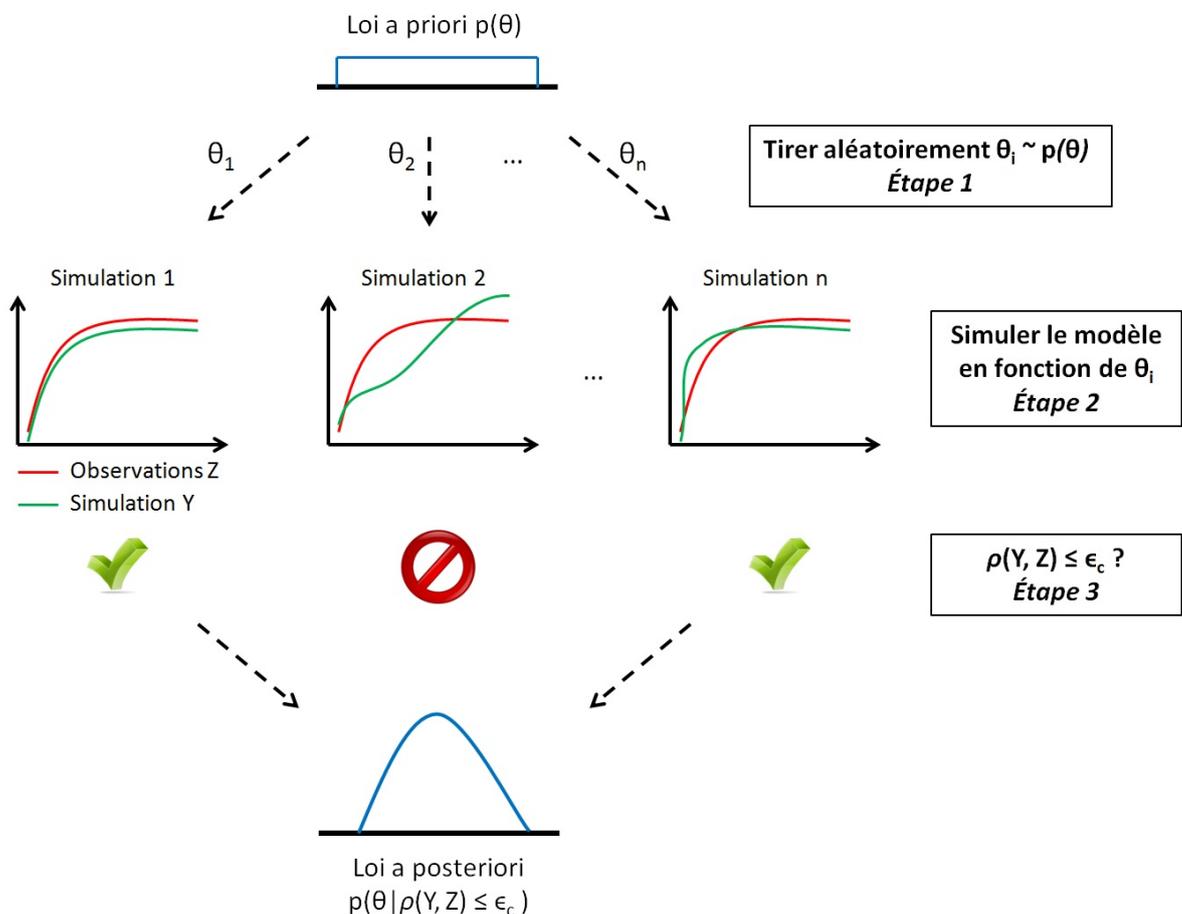


Figure 27 : Principe de l'algorithme de rejet approximatif

#### 2.4.2 Interprétation de l'algorithme de rejet approximatif

Wilkinson (2013) a démontré que l'algorithme de rejet approximatif pouvait être interprété comme réalisant l'inférence exacte en présence d'une erreur de modèle (ou de mesure) uniforme. Plus précisément, il est possible de considérer que la distribution *a posteriori* trouvée par

### Chapitre 3 : Calibrage de modèle

l'algorithme peut être interprétée comme la distribution exacte *a posteriori* d'un nouveau modèle où l'on considère en plus une distribution uniforme des termes d'erreur.

L'idée essentielle est de supposer qu'il existe un écart entre la meilleure prédiction possible du modèle et les données mesurées. Cet écart peut alors représenter une erreur de mesure, une erreur de modèle ou encore les deux à la fois. Ainsi, dans l'objectif de relier les sorties du modèles aux mesures, il est important de prendre en compte ce modèle d'erreur lors de l'inférence. Le cadre théorique est donc très proche de celui développé par Kennedy et O'Hagan (2001). On suppose alors que les données mesurées  $Z$  correspondent à la simulation du modèle à sa meilleure valeur de paramètre  $\hat{\theta}$  (meilleure prédiction possible) plus un terme d'erreur indépendant  $e$  (Wilkinson 2013) :

$$Z = \eta(u, \hat{\theta}) + e(u) \quad (3.7)$$

Par exemple, dans l'algorithme de rejet approximatif, si  $\rho(Y, Z) = |Y - Z|$ , cela revient à supposer que l'erreur  $e$  dans l'équation (3.7) suit une loi uniforme sur  $[-\epsilon_C, \epsilon_C]$ . L'algorithme permet alors de réaliser l'inférence exacte du modèle statistique décrit dans l'équation (3.7) et de déterminer la distribution *a posteriori*  $p(\theta|Z, e \sim U[-\epsilon_C, \epsilon_C])$ .

Wilkinson (2013) propose alors une généralisation de l'algorithme de rejet approximatif en modélisant la distribution de probabilité de l'erreur  $e$  par  $p_e(\cdot)$ . L'objectif de l'algorithme est de déterminer la distribution *a posteriori* des meilleurs paramètres  $\hat{\theta}$ , en prenant en compte l'erreur  $e$ , les données mesurées  $Z$  et la distribution *a priori* des paramètres  $\theta$ . L'algorithme se déroule comme suit (algorithme de rejet approximatif probabiliste) :

1. tirer aléatoirement  $\theta^* \sim p(\theta)$  ;
2. simuler  $Y \sim \eta(u, \theta^*)$  ;
3. accepter  $\theta^*$  avec la probabilité  $r = \frac{p_e(Z-Y)}{c}$ .

La constante  $c$  permet de s'assurer que  $\frac{p_e(Z-Y)}{c}$  correspond à une probabilité. Cet algorithme est équivalent à l'algorithme de rejet approximatif si l'on suppose que la probabilité d'acceptation correspond à :

$$\frac{p_e(Z - Y)}{c} = \begin{cases} 1 & \text{si } \rho(Y, Z) \leq \epsilon_C \\ 0 & \text{sinon} \end{cases} \quad (3.8)$$

Ainsi, si la distribution de probabilité  $p_e(\cdot)$  est modélisée avec attention et qu'elle représente fidèlement l'écart entre le modèle et la réalité, on peut supposer que  $\hat{\theta}$  représente la valeur « réelle » (physique) de  $\theta$ , et que  $p(\hat{\theta}|Z, e \sim p_e(\cdot))$  est la distribution *a posteriori* de  $\hat{\theta}$  à la lumière des données et du modèle (Wilkinson 2013).

De nombreux algorithmes ont été développés par la suite pour améliorer le taux d'acceptation de l'algorithme de rejet approximatif. Nous présentons dans la suite de ce chapitre les deux principales méthodes : les algorithmes ABC utilisant les méthodes MCMC et les algorithmes ABC s'inspirant des méthodes de Monte-Carlo séquentielles (ou SMC pour *Sequential Monte Carlo*) (Csilléry et al. 2010 ; B. M. Turner et Van Zandt 2012).

### 2.4.3 Les algorithmes ABC-MCMC

Dans l'objectif de diminuer dans l'algorithme de rejet approximatif le taux de rejet, Marjoram et al. (2003) ont proposé d'utiliser une approche MCMC ne nécessitant pas l'utilisation ou l'évaluation de la fonction de vraisemblance dans le calcul de la probabilité d'acceptation contrairement aux méthodes MCMC classiques.

L'algorithme proposé, reposant sur l'algorithme de Metropolis-Hastings, est alors le suivant :

1. tirer une valeur initiale  $\theta_0$  de la distribution *a priori*  $p(\theta)$  ;
2. pour  $i = 0, 1, 2, \dots$ 
  - a) tirer aléatoirement un candidat  $\theta^*$  dans une distribution  $q(\cdot | \theta_i)$  conditionnée par  $\theta_i$  ;
  - b) simuler  $Y \sim \eta(u, \theta^*)$  ;
  - c) calculer la probabilité d'acceptation :

$$\alpha = \begin{cases} \min \left( 1, \frac{p(\theta^*)q(\theta_i | \theta^*)}{p(\theta_i)q(\theta^* | \theta_i)} \right) & \text{si } \rho(Y, Z) \leq \epsilon_c \\ 0 & \text{si } \rho(Y, Z) > \epsilon_c \end{cases}$$

- d) tirer aléatoirement  $\beta$  dans une distribution uniforme sur  $[0, 1]$  :  
 si  $\beta < \alpha$ , on pose  $\theta_{i+1} = \theta^*$ , sinon  $\theta_{i+1} = \theta_i$ .

Le choix de la distribution  $q(\cdot)$  est important en pratique pour assurer la convergence de la méthode. En effet, si le « saut » généré dans l'espace des valeurs possibles des paramètres par la distribution  $q(\cdot)$  est trop important, cela génèrera des candidats souvent refusés pouvant laisser la chaîne « bloquée », ce qui entraîne un nombre élevé d'itérations. De même, si le « saut » est trop petit, cela amènera la chaîne à bouger trop lentement et donc nécessitera également un nombre important d'itérations. De plus, dans le cas spécifique des algorithmes ABC-MCMC, les candidats doivent satisfaire deux conditions (les candidats doivent tout d'abord respecter la distance minimale entre les données simulées et les données d'observation, et ensuite être acceptés avec une probabilité  $\alpha$ ) pouvant amener à un taux de rejet des candidats très important. Enfin, l'algorithme ABC-MCMC ne peut être parallélisé pour réduire les temps de calculs (B. M. Turner et Van Zandt 2012). En conséquence, les algorithmes ABC-MCMC ne sont pas retenus pour réaliser le calibrage du modèle de simulation thermique dynamique. D'autres algorithmes reposant sur les méthodes de Monte-Carlo séquentielles peuvent alors être utilisés.

### 2.4.4 Les algorithmes ABC-SMC

Comme pour les algorithmes ABC-MCMC, l'objectif des algorithmes ABC-SMC est de déterminer  $N$  valeurs de paramètres satisfaisant le seuil de tolérance  $\epsilon_c$ . La démarche développée par ces derniers est de définir une séquence décroissante de seuils de tolérance ( $\epsilon_1 \geq \epsilon_2 \geq \dots \geq \epsilon_c$ ) pour atteindre le seuil de tolérance cible  $\epsilon_c$ . L'algorithme détermine alors à chaque seuil de tolérance une distribution intermédiaire de  $N_{ABC}$  valeurs de paramètres satisfaisant ce seuil de tolérance. Ainsi, au lieu d'explorer l'espace complet des valeurs de paramètres comme dans l'algorithme de rejet

approximatif, ces algorithmes se concentrent sur les parties de l'espace où les valeurs de paramètres ont une grande probabilité de respecter le seuil de tolérance fixé.

Contrairement aux algorithmes ABC-MCMC, les algorithmes ABC-SMC ne travaillent pas sur un seul candidat  $\theta^*$  à la fois, mais testent un ensemble de candidats (appelés également « particules ») simultanément. Ces particules sont ensuite perturbées et filtrées à chaque itération de l'algorithme, les amenant à se rapprocher de plus en plus de la distribution *a posteriori* recherchée. Par exemple, pour passer de la distribution satisfaisant le seuil de tolérance  $\epsilon_1$  à la distribution à  $\epsilon_2$ , on va tirer aléatoirement une valeur de paramètre dans la distribution  $\epsilon_1$ , puis la faire bouger à l'aide d'un noyau de transition et enfin vérifier si les données simulées avec cette nouvelle valeur de paramètre respectent le seuil de tolérance  $\epsilon_2$ . Cependant, ce processus amène à un biais. En effet, les valeurs de paramètre satisfaisant le seuil de tolérance  $\epsilon_2$  n'ont pas été tirées dans la distribution *a priori*  $p(\theta)$  mais dans la distribution correspondant au seuil de tolérance  $\epsilon_1$ . Pour éviter ce biais, on associe à chaque valeur de paramètre un poids  $\omega_i$  qui est égal à la probabilité que la valeur de paramètre ait été tirée dans la distribution *a priori* divisée par la probabilité que la valeur de paramètre ait été tirée dans la distribution précédente.

La première application de ce type de méthode a été proposée par Sisson et al. (2007) avec l'algorithme ABC-PRC (*Partial Rejection Control*). Cependant, Beaumont et al. (2009) ont démontré que cet algorithme présentait un biais dans l'approximation de la distribution *a posteriori*. Ils ont alors proposé un nouvel algorithme corrigeant ce biais, appelé *Population Monte Carlo ABC* (ABC-PMC). La définition du poids et du noyau de transition (noyau gaussien) utilisés dans l'algorithme ainsi que les démonstrations associées se trouvent dans l'article de Beaumont et al. (2009).

L'algorithme ABC-PMC se déroule comme suit :

Soit une séquence décroissante de seuils de tolérance  $\epsilon_1 \geq \epsilon_2 \geq \dots \geq \epsilon_C$

1. A l'itération  $j = 1$

**Pour**  $i = 1, \dots, N_{ABC}$ ,

**Répéter**

Tirer  $\theta_i^{(1)} \sim p(\theta)$  et simuler  $Y \sim \eta(x, \theta_i^{(1)})$

**jusqu'à**  $\rho(Y, Z) \leq \epsilon_j$

Poser  $\omega_i^{(1)} = 1/N_{ABC}$

**fin**

Poser  $\sigma_1^2$  correspondant au double de la variance empirique de  $(\theta_i^{(1)})_{1 \leq i \leq N}$

2. A l'itération  $2 \leq j \leq C$

**Pour**  $i = 1, \dots, N_{ABC}$ ,

**Répéter**

Choisir  $\theta_i^*$  dans l'ensemble  $\theta_j^{(t-1)}$  avec la probabilité  $\omega_j^{(t-1)}$

Générer  $\theta_i^t | \theta_i^* \sim \mathcal{N}(\theta_i^*, \sigma_t^2)$ <sup>6</sup> et simuler  $Y \sim \eta(x, \theta_i^t)$

**jusqu'à**  $\rho(Y, Z) \leq \epsilon_j$

<sup>6</sup> où  $\mathcal{N}(\theta_i^*, \sigma_t^2)$  est la loi normale d'espérance  $\theta_i^*$  et de variance  $\sigma_t^2$

$$\text{Poser } \omega_i^{(t)} \propto \frac{p(\theta_i^{(t)})}{\sum_{j=1}^N \omega_j^{(t-1)} \sigma_t^{-1} \varphi(\sigma_t^{-1}(\theta_i^{(t)} - \theta_j^{(t-1)}))} \text{ avec } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

fin

Poser  $\sigma_{t+1}^2$  correspondant au double de la variance empirique de  $(\theta_i^{(t)})_{1 \leq i \leq N}$

Cet algorithme a l'avantage d'être facilement implémentable, robuste et présente de bons résultats (Beaumont et al. 2009; Lenormand et al. 2013). De plus, il peut être parallélisé permettant ainsi de diminuer fortement les temps de calcul. Par conséquent, cet algorithme sera utilisé dans la méthodologie développée pour calibrer le logiciel de simulation thermique dynamique COMFIE.

## 3. MÉTHODOLOGIE

Dans ce chapitre, nous souhaitons calibrer le modèle de simulation thermique dynamique à partir de la comparaison des températures horaires simulées par le modèle thermique et des températures mesurées.

La méthodologie développée pour réaliser le calibrage s'inspire de celle développée par Coakley et al. (2011) et se déroule en cinq étapes :

1. récupération de l'ensemble des données disponibles sur le bâtiment ;
2. création du modèle de bâtiment ;
3. analyse de sensibilité ;
4. calibrage du modèle (estimations des PDF *a posteriori*) ;
5. évaluation du modèle calibré.

Ces étapes sont détaillées dans les sections suivantes.

Pour mener à bien le processus de calibrage, les données mesurées à disposition doivent être séparées en données utilisées par l'analyse de sensibilité (étape 3, §3.3) et par le calibrage du modèle (étape 4, §3.4), et en données de validation utilisées pour l'évaluation de l'efficacité du processus de calibrage (étape 5, §3.5).

### 3.1 Étape 1 : récupération des données

L'objectif de cette étape est de minimiser les incertitudes de spécification dans le processus de calibrage. Pour cela, il est nécessaire de récolter l'ensemble des sources d'information disponibles pour la modélisation du bâtiment. Il peut s'agir par exemple des différents documents décrivant le bâtiment (données de construction, géométrie du bâtiment, propriétés physiques des matériaux), de données de fonctionnement des systèmes, de données mesurées dans le bâtiment (taux d'infiltration par exemple), de données d'occupation (capteurs de présence), de données météorologiques (station météorologique) ou encore de données sur les incertitudes pesant sur les différents paramètres ou les mesures. Il est important que les différentes sources d'information soient répertoriées et classées pour faciliter le suivi de la modélisation.

## 3.2 Étape 2 : création du modèle

L'objectif de cette étape est de minimiser les incertitudes de modélisation ainsi que les incertitudes pesant sur la définition des sollicitations et des scénarios. Il est important de prêter une attention toute particulière à la modélisation du bâtiment avec le logiciel de simulation thermique dynamique. Les différentes hypothèses de modélisation doivent être clairement identifiées, répertoriées et justifiées. Ces hypothèses peuvent correspondre par exemple à la définition du nombre de zones thermiques, à la modélisation des systèmes énergétiques, à la modélisation des différents scénarios d'usage (privilégier les mesures aux données par défaut) ainsi qu'à la modélisation des sollicitations. Cette étape doit être réalisée de manière itérative pour s'assurer que toutes les données recensées dans l'étape 1 ont été utilisées.

## 3.3 Étape 3 : analyse de sensibilité

L'objectif de cette étape est de déterminer les paramètres incertains les plus influents dans le modèle de bâtiment. En effet, il est inenvisageable du point de vue du temps de calcul de réaliser le calibrage du modèle en considérant l'ensemble des paramètres incertains. L'objectif de l'analyse de sensibilité est alors d'identifier les paramètres incertains les plus influents qui seront alors retenus pour réaliser le calibrage dans l'étape 4.

Une méthode de criblage peut être utilisée pour déterminer ces paramètres influents. La méthode de Morris (Morris 1991 ; Campolongo et al. 2007) a été retenue car elle ne nécessite pas d'hypothèses sur le modèle, elle est peu coûteuse en temps de calcul et elle permet de prendre en compte les interactions entre les différents paramètres (contrairement aux analyses de sensibilité locales).

La méthode de Morris fait partie des méthodes dites « un-par-un » ou « *one at a time* (OAT) ». En effet, elle réalise d'une part une discrétisation régulière de l'espace des paramètres (ainsi seuls des points appartenant à la grille peuvent être échantillonnés) et d'autre part elle ne fait varier qu'un paramètre à la fois entre les simulations successives pour évaluer l'influence de ces paramètres.

### 3.3.1 Plan d'échantillonnage de la méthode de criblage

On suppose que notre modèle dispose de  $K$  paramètres incertains. On estime tout d'abord l'incertitude pesant sur les différents paramètres en fixant une valeur haute et basse. Ensuite, la gamme de variation de chaque facteur est divisée en  $Q$  niveaux  $\{0, \frac{1}{Q-1}, \frac{2}{Q-1}, \dots, 1\}$ . Cette transformation linéaire permet ainsi de comparer des paramètres de natures différentes. Enfin, l'échantillonnage de la méthode de Morris consiste à répéter  $r$  fois une trajectoire aléatoire passant par  $K + 1$  nœuds de l'espace discrétisé des paramètres  $\hat{\Omega}$  de telle sorte que chaque paramètre ne

### Chapitre 3 : Calibrage de modèle

varie qu'une seule fois par trajectoire. Plus précisément, chaque trajectoire doit respecter les propriétés suivantes :

- le nœud de départ est choisi au hasard dans  $\widehat{\Omega}$  ;
- le saut entre deux nœuds successifs de la trajectoire est proportionnel à  $\delta = \frac{1}{Q-1}$  ;
- l'ordre dans lequel les paramètres sont modifiés est tiré de manière aléatoire.

La Figure 28 illustre le fonctionnement de l'échantillonnage dans la méthode de Morris.

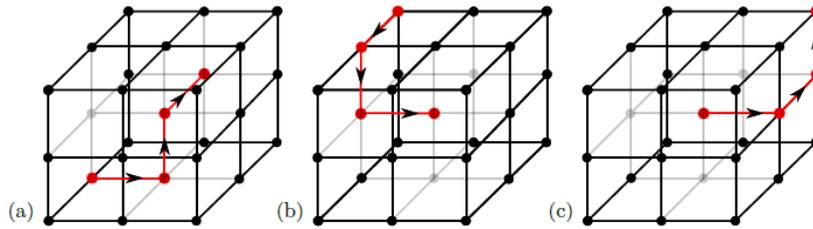


Figure 28 : Exemple d'échantillonnage par la méthode de Morris avec  $K = 3$  facteurs,  $Q = 3$  niveaux, et  $r = 3$  trajectoires (a), (b), et (c) (tiré de Munaretto (2014))

Pour garantir l'uniformité dans la distribution des points échantillonnés, Morris propose de choisir une longueur de saut égale à  $\Delta = \frac{Q}{2(Q-1)}$  avec un niveau de discrétisation  $Q$  pair (Faivre et al. 2013).

### 3.3.2 Calcul des indices pour l'analyse de sensibilité

Pour chaque trajectoire  $i$ , la méthode de Morris calcule un effet élémentaire pour chaque paramètre  $\theta_j$  correspondant à la variation de la sortie du modèle  $\eta$  entre deux points successifs de la trajectoire (pour simplifier les notations,  $\eta(u, \theta)$  est remplacé ici par  $\eta(\theta)$ ) :

$$E_j^{(i)} = \pm \frac{\eta(\theta_1^{(i)}, \dots, \theta_j^{(i)} \pm \Delta, \dots, \theta_K^{(i)}) - \eta(\theta_1^{(i)}, \dots, \theta_j^{(i)}, \dots, \theta_K^{(i)})}{\Delta} \quad (3.9)$$

Une fois les  $r$  répétitions réalisées, on détermine :

- la moyenne des effets élémentaires :

$$\mu_j = \frac{1}{r} \sum_{i=1}^r E_j^{(i)} \quad (3.10)$$

- la moyenne absolue des effets élémentaires :

$$\mu_j^* = \frac{1}{r} \sum_{i=1}^r |E_j^{(i)}| \quad (3.11)$$

- l'écart-type des effets élémentaires :

$$\sigma_j = \sqrt{\frac{1}{r-1} \sum_{i=1}^r (E_j^{(i)} - \mu_j)^2} \quad (3.12)$$

Pour éviter que des effets élémentaires de signe opposé ne se compensent,  $\mu_j^*$  est préféré à  $\mu_j$ .  $\mu_j^*$  permet ainsi de représenter le degré d'influence qu'à le paramètre  $\theta_j$  sur la sortie du modèle, alors que  $\sigma_j$  donne des informations quant aux potentielles implications des paramètres d'entrées dans des interactions ou des non-linéarités. Les paramètres incertains peuvent alors être tracés dans un graphique de Morris (Figure 29) et classés selon trois catégories :

- paramètres ayant des effets négligeables (groupe n°1) ;
- paramètres ayant des effets linéaires (groupe n°2) ;
- paramètres ayant des effets non-linéaires et/ou avec interactions (groupe n°3).

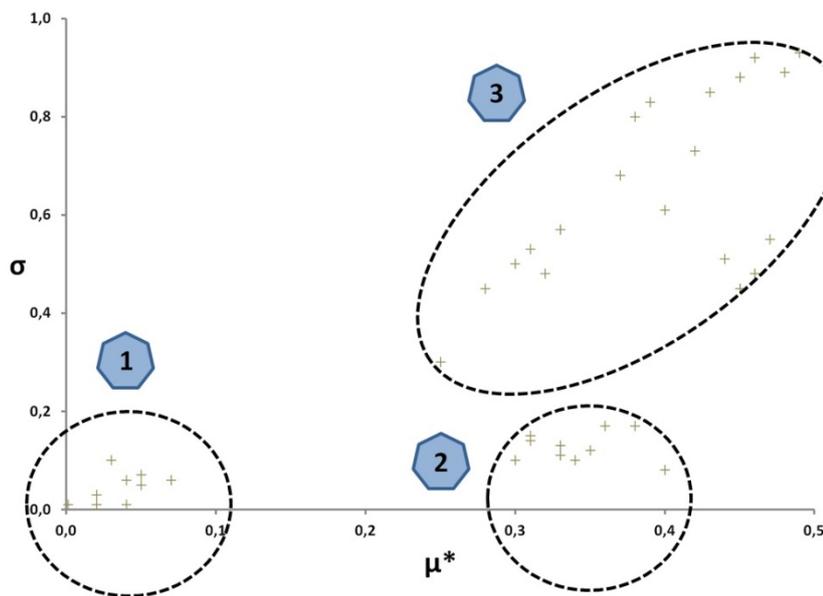


Figure 29 : Exemple de résultats par la méthode de Morris, adapté de Iooss (2009)

Lors de cette étape, les paramètres sont classés à partir de la distance entre l'origine et le point relatif au paramètre  $\theta_j$  dans le graphique  $\mu^*$  vs  $\sigma$  (Munaretto 2014) dans l'objectif de déterminer et de sélectionner les paramètres les plus influents :

$$d_j^* = \sqrt{\mu_j^{*2} + \sigma_j^2} \quad (3.13)$$

### 3.4 Étape 4 : calibrage du modèle

L'objectif de cette étape est de minimiser les incertitudes épistémiques pesant sur les paramètres incertains les plus influents identifiés à l'étape 3.

Dans un premier temps, il est important de définir avec précision la distribution de probabilité *a priori* des paramètres influents, représentant la connaissance (supposée) que l'on a sur ces paramètres. Ensuite, l'algorithme ABC-PMC (§2.4.4) est mis en œuvre pour réaliser le calibrage du modèle. Le cadre théorique utilisé est celui développé par Wilkinson (2013). Nous considérons ainsi que les données mesurées  $T^{mes}$  correspondent à la meilleure prédiction du modèle plus un terme d'erreur indépendant  $e$ , représentant l'erreur du modèle (cf. §2.4.2). Les erreurs de mesure seront quant à elles considérées lors de l'évaluation des résultats et de la propagation des incertitudes (étape 5). Dans l'objectif de calculer une distance entre les données simulées par le modèle et les données observées, la définition d'une grandeur d'intérêt est nécessaire. Ainsi, pour chaque simulation  $i$ , la distance quadratique moyenne notée  $\rho(T_i^{sim}, T^{mes})$  est calculée par :

$$\rho(T_i^{sim}, T^{mes}) = RMSE_i = \sqrt{\left( \sum_{k=1}^N (T_i^{sim}(k\Delta t) - T^{mes}(k\Delta t))^2 / N \right)} \quad (3.14)$$

avec  $\Delta t$  le pas de temps de simulation,  $T_i^{sim}(k\Delta t)$  et  $T^{mes}(k\Delta t)$  la température respectivement simulée et mesurée à l'instant  $k\Delta t$ . La définition de cette distance a l'avantage de pouvoir pénaliser les écarts importants entre les données simulées et les données observées. L'algorithme ABC-PMC permet alors de réaliser l'inférence du modèle statistique décrit dans l'équation (3.7) et de déterminer la distribution *a posteriori*,  $p(\theta|T^{mes}, \rho(T_i^{sim}, T^{mes}) \leq \epsilon_c)$ , des paramètres les plus influents.

## 3.5 Étape 5 : évaluation des résultats

L'objectif de cette étape est d'évaluer, par la propagation des incertitudes, le degré d'incertitudes sur les résultats du modèle avant et après calibrage dans le but d'apprécier l'efficacité du processus de calibrage.

### 3.5.1 Propagation des incertitudes

La propagation des incertitudes permet d'estimer la distribution des sorties du modèle et ainsi donner des informations sur l'incertitude associée aux prédictions du modèle. Plus précisément, cette méthode vise à évaluer le degré d'incertitude relatif aux résultats du modèle induit par les incertitudes pesant sur les paramètres du modèle. La Figure 30 illustre le fonctionnement de cette méthode où l'objectif est de déterminer le profil d'incertitudes du modèle (en vert) paramétré par deux paramètres dont les profils d'incertitudes sont connus (en gris).

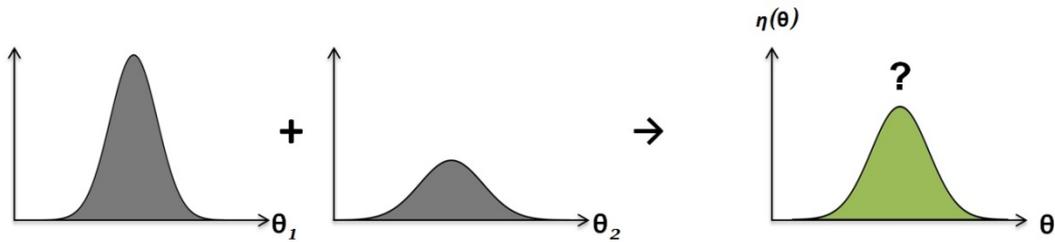


Figure 30 : Illustration de la propagation des incertitudes, adapté de Makowski (2014)

Cette méthode se déroule en quatre étapes :

- définir la fonction de densité de probabilité de chaque paramètre ;
- échantillonner les paramètres incertains ;
- simuler le modèle avec les paramètres échantillonnés ;
- évaluer le profil d'incertitudes de la sortie du modèle à partir des résultats des simulations.

### 3.5.2 Évaluation de l'efficacité du calibrage

Suite à la propagation des incertitudes sur les modèles calibré et non calibré, il est possible de calculer pour chaque modèle une RMSE moyenne sur les données de validation correspondant à :

$$RMSE = \frac{\sum_{i=1}^{N_{pro}} RMSE_i}{N_{pro}} \quad (3.15)$$

où  $N_{pro}$  est le nombre de simulations réalisées lors de la propagation des incertitudes. La comparaison entre la RMSE obtenue par le modèle calibré et le modèle non calibré permettra alors d'évaluer l'efficacité du processus de calibrage.

## 4. CAS D'APPLICATION

### 4.1 Collecte des données (étape 1)

#### 4.1.1 Présentation du bâtiment

Le bâtiment étudié correspond à la maison I-BB (Béton banché) de la plateforme « INCAS » de l'INES (Institut National de l'Énergie solaire), située au Bourget-du-Lac. D'une architecture simple et compacte, elle dispose d'une surface intérieure de 89 m<sup>2</sup> et comporte deux étages. La maison a été conçue pour correspondre à la performance du label « PassivHaus », grâce notamment à une forte

### Chapitre 3 : Calibrage de modèle

isolation, de très faibles ponts thermiques, des vitrages performants, ou encore une étanchéité à l'air de l'enveloppe soignée.



Figure 31 : Maison I-BB (façades sud et ouest)

Les dimensions intérieures sont de 7,5 m en longueur et de 6,5 m en largeur. La maison est bâtie sur un vide sanitaire de 80 cm de hauteur et est surmontée de combles indépendants du volume chauffé. Son orientation est décalée de 15° dans le sens trigonométrique par rapport à l'axe nord-sud. La toiture est à deux pans avec une orientation Nord/Sud et un débord de 60 cm à l'est, à l'ouest et au nord. La façade sud du bâtiment comprend une grande surface vitrée (28 % de la façade) protégée du soleil par un balcon d'une largeur de 1,3 m et une avancée de toiture de 1 m (Figure 32). La paroi nord ne comprend que deux petites fenêtres (2,5 % de la façade) de type triple vitrage afin de minimiser les pertes thermiques pendant la période hivernale. Les façades est et ouest disposent quant à elles d'un ratio de surface vitrée de 5 % et 10 % respectivement.

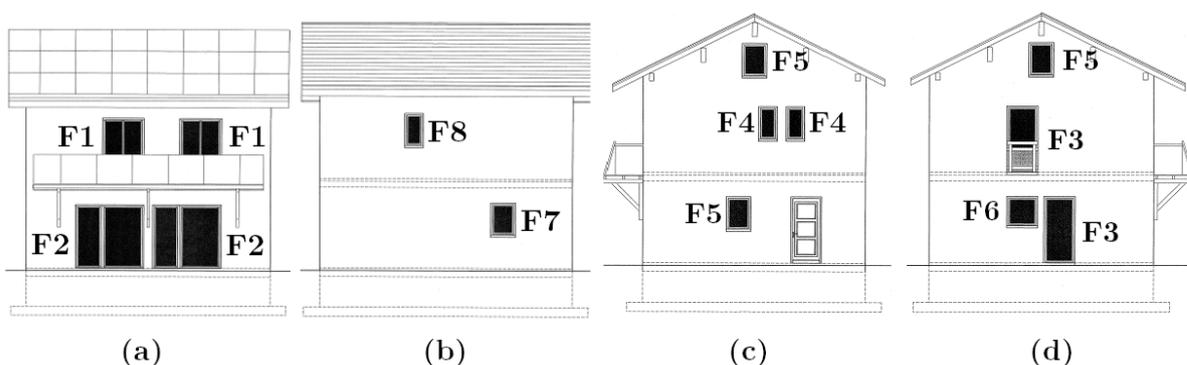


Figure 32 : Plan des façades de la maison I-BB, (a) sud, (b) nord, (c) ouest, (d) est. Source (CEA 2010)

Les menuiseries ont des cadres PVC dont le coefficient de déperdition thermique est égal à 1,45 W/(m<sup>2</sup>.K). Les façades Sud, Est et Ouest possèdent des doubles vitrages 4-16-4 à faible émissivité avec remplissage argon. Le coefficient de déperdition thermique de ces vitrages est égal à 1,1 W/(m<sup>2</sup>.K) avec un facteur solaire de 0,6. La façade nord est quant à elle équipée par des triples vitrages 4-12-4-12-4 à faible émissivité avec remplissage argon. Ces vitrages ont un coefficient de déperdition thermique égal à 0,7 W/(m<sup>2</sup>.K) et un facteur solaire égal à 0,45.

### Chapitre 3 : Calibrage de modèle

La maison I-BB est fortement isolée (isolation par l'extérieur) avec une inertie lourde comme montré dans le Tableau 9.

Tableau 9 : Composition des parois de la maison étudiée

	Murs extérieurs	Combles	Plancher bas	Plancher intermédiaire	Cloisons intérieures
Composition (ext → int)	20 cm de polystyrène extrudé 15 cm de béton banché	40 cm de laine de verre 1,5 cm de plaque de plâtre	20 cm de polystyrène extrudé Béton coulé sur place de 16 cm Chape en béton de 8 cm	Plancher poutrelles et entrevous de 16 cm 4 cm de dalle 8 cm de chape	1,5 cm de plâtre 4 cm de laine de verre 1,5 cm de plâtre
U (W/(m <sup>2</sup> .K))	0,15	0,09	0,15	2,2	0,96

La maison est équipée d'une VMC (Ventilation Mécanique Contrôlée) de type double flux à deux ventilateurs : l'un pour le circuit d'air neuf et l'autre pour le circuit d'air vicié. L'air neuf est soufflé dans le séjour et les chambres tandis que des bouches situées dans les toilettes et la salle de bain assurent l'extraction de l'air vicié. L'échangeur à plaques permet une récupération de chaleur jusqu'à 90 %. Les ventilateurs sont dimensionnés pour imposer des débits nominaux de 135 ( $Q_{v,nom-1}$ ) ou 270 m<sup>3</sup>/h ( $Q_{v,nom-2}$ ). Un by-pass permet le rafraîchissement par ventilation directe. Le chauffage est assuré par une résistance électrique d'une puissance de 1,2 kW située au départ du réseau de distribution de l'air neuf. Pour plus de détails sur la maison I-BB, il est conseillé au lecteur de se rapporter à la thèse de Munaretto (2014).

#### 4.1.2 Protocole expérimental

L'étude consiste à comparer des mesures de température d'air intérieur réalisées dans la maison I-BB avec les températures évaluées par le logiciel de simulation thermique dynamique COMFIE. L'objectif est d'améliorer les prédictions du modèle grâce au calibrage de ce dernier. Cette étude s'inscrit dans le prolongement des études réalisées par Munaretto (2014) et Recht et al. (2014) sur la validation empirique de l'outil COMFIE.

##### 4.1.2.1 Description des scénarios

La période d'étude considérée, qui s'étend du 1<sup>er</sup> janvier au 22 avril 2012, correspond à un protocole expérimental subdivisé en six scénarios, avec une prise en compte séparée ou combinée de différents phénomènes physiques. Ainsi, selon les scénarios, les expérimentateurs de l'INES ont modifié un voire plusieurs paramètres d'exploitation, tels que la consigne de chauffage, l'ouverture/fermeture des volets, ou encore le régime et le débit de la VMC.

### Chapitre 3 : Calibrage de modèle

Les six scénarios sont décrits ci-dessous (Figure 33) :

- **scénario n°1** : consigne de chauffage à 20°C ; by-pass de l'échangeur ; débit nominal de renouvellement d'air  $Q_{v,nom-1}$  ; volets fermés ; séparation physique entre l'étage et le rez-de-chaussée (RDC) ;
- **scénario n°2** : température en évolution libre ; système de chauffage et de ventilation éteints (off) ;
- **scénario n°3** : séparation physique entre les deux niveaux retirée ;
- **scénario n°4** : mêmes conditions que le scénario n°1 mais avec une consigne de chauffage à 24°C ;
- **scénario n°5** : chauffage éteint (off), ouverture des volets, séparation physique entre les deux niveaux retirée ;
- **scénario n°6** : renouvellement d'air avec un débit nominal augmenté noté  $Q_{v,nom-2}$ .

	Scénario n°1	Scénario n°2	Scénario n°3	Scénario n°4	Scénario n°5	Scénario n°6
Consigne	20°C	Off	Off	24°C	Off	Off
Régime VMC	BP	Off	Off	BP	BP	BP
Débit VMC	$Q_{v,nom-1}$	0	0	$Q_{v,nom-1}$	$Q_{v,nom-1}$	$Q_{v,nom-2}$
Volets	Fermés	Fermés	Fermés	Fermés	Ouverts	Ouverts
Séparation RDC/R+1	Avec	Avec	Sans	Avec	Sans	Sans

Timeline: 01/01/2012, 16/01/2012, 09/02/2012, 20/02/2012, 15/03/2012, 31/03/2012, 22/04/2012

Figure 33 : Description du protocole expérimental

Dans notre cas d'application, nous avons choisi de séparer en deux catégories les différents scénarios. Ainsi, les scénarios n°1, 2, 3 et 5 ont été utilisés pour identifier les paramètres les plus influents (étape 3, analyse de sensibilité) et déterminer leur lois de distribution *a posteriori* (étape 4, calibrage). Les scénarios n°4 et 6 ont quant à eux été utilisés pour comparer les données simulées par l'outil COMFIE avant et après calibrage (étape 5, propagation d'incertitudes) et ainsi évaluer et quantifier l'apport du calibrage du modèle. Les scénarios n°4 et 6 ont été choisis comme données de validation car ils sont proches respectivement du scénario n°1 et du scénario n°5. On peut ainsi évaluer le comportement du modèle calibré dans des conditions expérimentales relativement proches. De plus, les scénarios n°4 et 6 ont l'intérêt de mettre en avant des processus physiques différents avec l'ouverture ou non des volets.

#### 4.1.2.2 Données météorologiques

Les données météorologiques utilisées correspondent aux données mesurées (température extérieure, rayonnement solaire global et diffus horizontal) au pas de temps horaire à l'aéroport du Bourget-du-Lac situé à quelques centaines de mètres de la maison I-BB.

Les informations relatives au site sont :

- longitude : 5,8814° ;
- latitude : 45,6876° ;
- altitude : 233 m ;
- température moyenne du sol à 10 m : 9°C.

Le Bourget-du-Lac est entouré de massifs montagneux induisant des masques lointains. Cependant, ces derniers n'ont pas été modélisés car ils sont déjà pris en compte par les données météorologiques : la station météorologique est à quelques centaines de mètres de la maison I-BB.

#### 4.1.2.3 Mesures des températures intérieures

A l'intérieur de la maison, les températures d'air sont mesurées par des sondes de platine protégées par un bouclier thermique pour éviter que le rayonnement n'influence la mesure. Elles sont mesurées dans chacune des pièces, au centre du volume, à une hauteur de 1,10 m. Pour effectuer la comparaison entre l'outil de simulation COMFIE et l'expérimentation, une bande d'incertitudes de  $\pm 1^\circ\text{C}$  a été considérée sur la mesure, en raison des incertitudes intrinsèques des capteurs, des effets induits par le bouclier thermique et de l'inhomogénéité de température dans les pièces (Spitz 2012).

## 4.2 Hypothèses de modélisation du bâtiment (étape 2)

Il est conseillé au lecteur intéressé de se reporter à la thèse de Munaretto (2014) s'il souhaite plus de précision sur les hypothèses de modélisation.

### 4.2.1 Zonage de la maison

Nous avons considéré dans cette étude un modèle thermique du bâtiment monozone. En effet, le cadre général de cette thèse concerne la régulation et l'application de différentes méthodes d'optimisation. Il est donc nécessaire de disposer d'un modèle de bâtiment rapide à simuler dans l'objectif de minimiser les temps de calcul. Cependant, ce choix amène à une diminution de la précision des températures simulées. Par conséquent, les températures intérieures mesurées dans les pièces de la maison ont été agrégées au prorata des surfaces intérieures de ces pièces afin de les comparer avec la température intérieure simulée par le modèle.

L'outil de simulation dynamique COMFIE est basé sur un modèle 1D au niveau des parois. Le choix d'une convention de surface est donc essentiel lorsque les dimensions d'un bâtiment sont décrites sur l'un de ces modèles numériques. En convention extérieure, le volume chauffé sera fortement surévalué tandis qu'en convention intérieure, la surface des parois déperditives sera sous-estimée. Afin de réduire l'erreur de modélisation, une convention au milieu du béton a été considérée pour la maison I-BB, ce qui permet de représenter plus fidèlement l'inertie thermique des locaux.

### 4.2.2 Débits de ventilation

Les informations constructeurs affirment que les débits nominaux sont égaux à  $Q_{v,nom-1} = 135 \text{ m}^3/\text{h}$  et  $Q_{v,nom-2} = 270 \text{ m}^3/\text{h}$ . Cependant, suite à une campagne de mesures menée par le CEA, les débits nominaux  $Q_{v,nom-1}$  et  $Q_{v,nom-2}$  ont été estimés à 110 et 160  $\text{m}^3/\text{h}$  respectivement et ont été retenus dans l'étude.

### 4.2.3 Consignes de température

Sur la période de l'étude, les températures de consigne n'ont jamais été atteintes en pratique (l'échangeur de chaleur de la VMC était by-passé), la puissance de chauffage était donc maximale sur tous les scénarios où le chauffage était allumé.

### 4.2.4 Infiltrations

Suite aux tests de porte soufflante effectués in situ par le CETE de Lyon en février 2010, nous avons fixé le débit d'infiltration à 0,03 vol/h.

### 4.2.5 Ponts thermiques

Les ponts thermiques ont été calculés à l'aide d'un logiciel basé sur une discrétisation en éléments finis (TRISCO).

### 4.2.6 Puissance dissipée par les systèmes

Les différentes puissances dissipées à l'intérieur de la maison par les multiples appareils électriques (commutateur, transformateurs, centrale de traitement des mesures, capteurs divers, moteur de la VMC) ont été prises en compte. Un total de 208 W pour l'ensemble de la zone a alors été considéré.

### 4.2.7 Occultation des vitrages

Des volets roulants sont installés sur l'ensemble des menuiseries. La résistance additionnelle des volets a été fixée à 0,20  $\text{m}^2 \cdot \text{K}/\text{W}$ . Cette valeur reste cependant très incertaine car nous disposons de peu d'informations sur ce sujet.

### 4.2.8 États de surface

Toutes les surfaces intérieures à l'exception des planchers sont de couleur blanche, de même que les enduits extérieurs (Tableau 10). Les planchers sont quant à eux de couleur brun clair (carrelage).

Tableau 10 : Propriétés radiatives des revêtements de la maison I-BB

Type de surface	$\epsilon_{GLO}$ [-]	$\alpha_{CLO}$ [-]
Enduit extérieur	0,9	0,65
Peinture intérieure	0,9	0,2
Carrelage	0,9	0,7

### 4.2.9 Masques proches

Les différents masques proches induits par le balcon (avancée du balcon, garde corps), l'avancée de la toiture et le retrait de toutes les menuiseries ont été pris en compte.

### 4.2.10 Albédo

Le sol jouxtant la maison est du sable blanc ; de l'herbe se trouve à proximité. Or la valeur de l'albédo est différente entre ces deux revêtements. Un albédomètre a été mis en place afin de mesurer la réflectivité moyenne du sol environnant et une valeur de 0,35 a été mesurée. Une incertitude relativement élevée pèse sur ce paramètre car la valeur de l'albédo est sujette à des variations au cours de l'année.

### 4.2.11 Initialisation de la simulation

La température de zone a été initialisée à la température mesurée expérimentalement à la première heure (le bâtiment ayant été chauffé à 20°C les semaines précédentes) et les profils de température des parois correspondent à ceux en régime permanent avec les températures d'air extérieur et de zone comme conditions aux limites.

## 4.3 Analyse de sensibilité : méthode de Morris (étape 3)

Pour identifier les paramètres à calibrer les plus influents, la méthode de Morris a été utilisée. Comme expliqué dans le §3.3.2, les différents paramètres  $\theta_j$  ont été classés selon leurs distances  $d_j^*$ . Plus précisément, la distance  $d_j^*$  a été calculée sur chaque scénario  $s$  (scénarios n°1, 2, 3 et 5), puis normalisée selon l'équation suivante :

$$D_{j,s}^* = \frac{d_{j,s}^* - \min_j(d_{j,s}^*)}{\max_j(d_{j,s}^*) - \min_j(d_{j,s}^*)} \quad (3.16)$$

### Chapitre 3 : Calibrage de modèle

Le maximum des  $D_{j,s}^*$  sur les quatre phases du protocole est choisi pour classer les paramètres. Les phénomènes physiques mis en avant dans chaque scénario étant différents, le but est de retenir tout facteur qui serait influent, ne serait-ce que sur une seule phase du protocole.

L'influence de 102 paramètres incertains a ainsi été évaluée dans ce cas d'application. Ces paramètres concernent toutes les hypothèses de modélisation prises dans l'étape de modélisation. Les incertitudes considérées sur chaque paramètre ainsi que les abréviations utilisées se trouvent dans le Tableau 30 de l'Annexe C.

Concernant les paramètres de la méthode de criblage, nous avons considéré :  $r = 100$ ,  $Q = 6$  et  $\Delta = \frac{Q}{2 \times (Q-1)} = 0,6$ .

Les résultats du criblage sont présentés de la Figure 34 à la Figure 36. De manière générale, on peut noter l'influence de la puissance dissipée ( $Pdi$ ) sur l'ensemble des scénarios.

On constate dans le scénario n°1 (Figure 34) la forte influence du débit de ventilation  $Q_{v,nom-1}$ . Ceci est normal car il s'agit du débit de ventilation appliqué lors de ce scénario. De même, étant donné que le chauffage était actif, la puissance de chauffage ( $Pch$ ) est également très influente. Enfin, on peut remarquer la résistance thermique des volets roulants  $Roc$  qui peut s'expliquer par l'incertitude forte qui pèse sur cet indicateur.

Le modèle de bâtiment considéré étant monozone, les scénarios n°2 et 3 sont identiques. Par conséquent, les résultats du criblage sont rassemblés pour ces deux scénarios (Figure 35). De manière logique, on constate l'apparition de la capacité thermique volumique du béton banché ( $Cvl - \text{Béton banché}$ ) et de la chape béton ( $Cvl - \text{Chape béton}$ ). En effet, le chauffage est coupé à la fin du scénario n°1. Par conséquent, l'évolution de la température du bâtiment dépend fortement de l'inertie thermique du bâtiment, ce dernier étant isolé par l'extérieur. On peut également remarquer que les scénarios n°2 et 3 sont encore influencés par la puissance de chauffage et le débit de ventilation. Ceci est dû à la persistance de leur influence après le scénario n°1. Le chauffage étant éteint lors de cette période expérimentale, on note l'augmentation de l'influence de la résistance des volets roulants.

Le scénario n°5 (Figure 36), correspondant à l'ouverture des volets, fait émerger les paramètres relatifs aux apports solaires, à savoir l'albédo ( $Alb$ ), le facteur solaire de la fenêtre F2 ( $Sow - F2$ ) ainsi que l'absorptivité de l'enduit ( $Alp - \text{Enduit}$ ). Les volets étant ouverts, la résistance thermique des occultations disparaît. Bien que le chauffage soit coupé, le scénario n°5 est encore influencé par la puissance de chauffage utilisée lors du scénario n°4.

Enfin, sur l'ensemble des scénarios, on peut noter l'influence de la puissance dissipée ( $Pdi$ ).

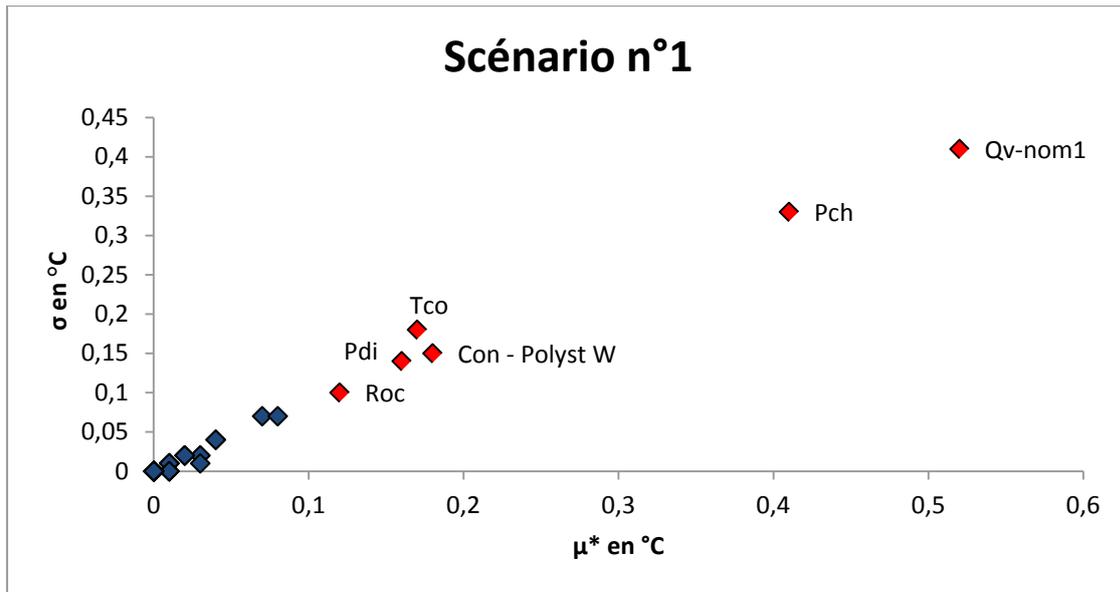


Figure 34 : Graphe de Morris, scénario n°1

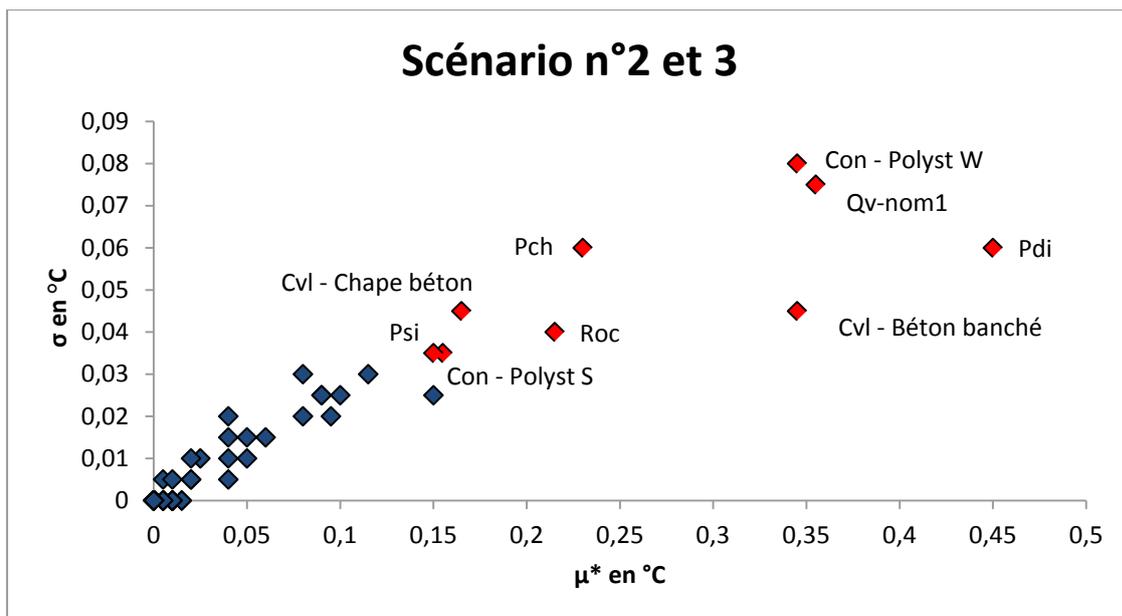


Figure 35 : Graphe de Morris, scénario n°2 et 3

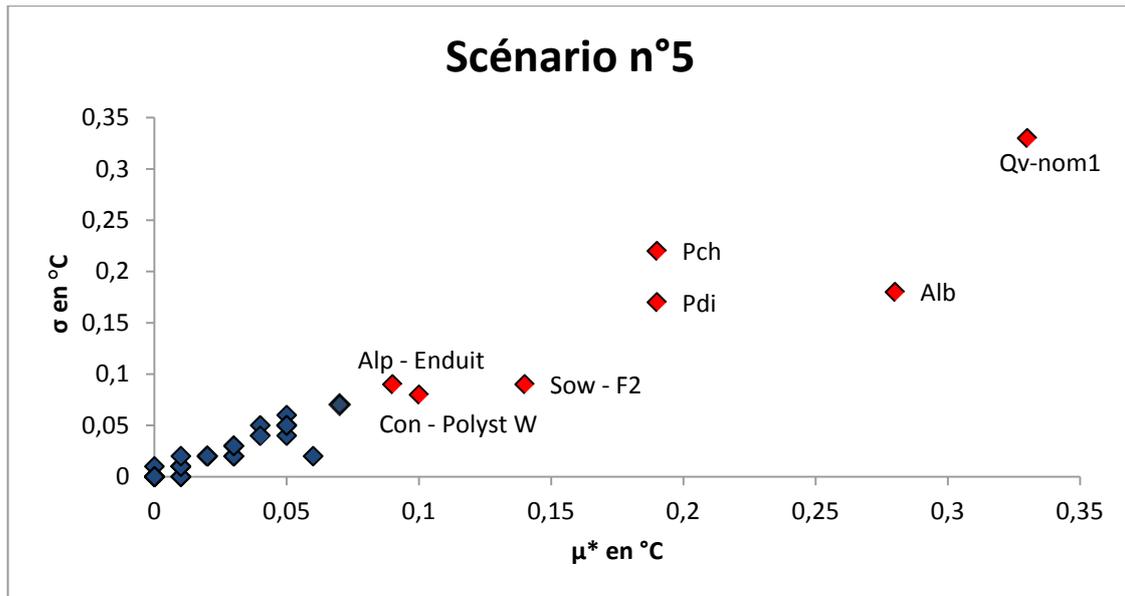


Figure 36 : Graphe de Morris, scénario n°5

Suite à la présentation des résultats du criblage, l'ensemble des paramètres ont été classés selon leur distance  $D_{j,s}^*$  dans l'objectif de sélectionner les paramètres les plus influents sur l'ensemble des scénarios. Les résultats de ce classement sont présentés dans la Figure 37. Suite à ce classement, seuls les six premiers paramètres ont été sélectionnés pour l'étape de calibrage, car d'une part ils se détachent des autres paramètres, et d'autre part ce nombre de paramètre permet d'éviter des temps de calcul prohibitifs lors du calibrage. Ces paramètres correspondent au débit de ventilation  $Q_{v,nom-1}$ , à la puissance dissipée  $Pdi$ , à la puissance de chauffage  $Pch$ , à la capacité thermique volumique du béton banché  $Cvl - \text{Béton banché}$ , à l'albédo  $Alb$ , et à la conductivité thermique du polystyrène extrudé  $Con - \text{Polyst W}$ .

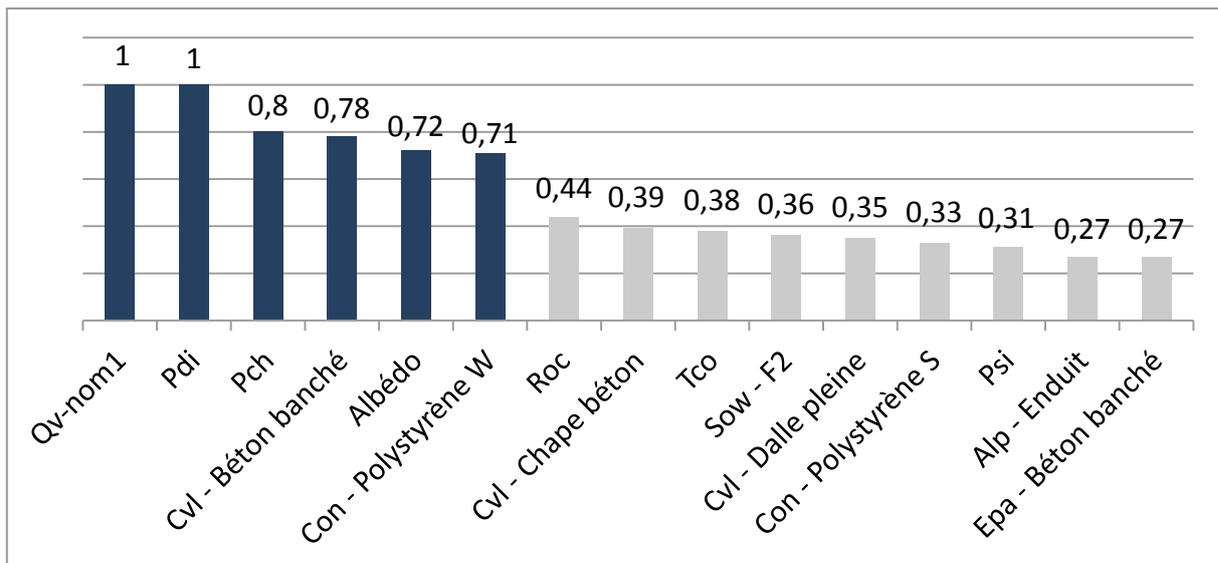


Figure 37 : Classement des facteurs incertains selon  $\max_s(D_{j,s}^*)$

## 4.4 Calibrage du modèle (étape 4)

Dans un premier temps, il est nécessaire de définir les distributions de probabilité *a priori* des paramètres sélectionnés à l'étape 3. Les distributions considérées dans ce cas d'application sont résumées dans le Tableau 11 (Munaretto 2014).

Tableau 11 : Distributions de probabilité *a priori* considérées

Paramètre	Loi <i>a priori</i>	Unité
$Q_{v,nom-1}$	Loi normale (150,7 ; 15,07)	[m <sup>3</sup> /h]
$P_{di}$	Loi normale (208 ; 20,8)	[W]
$P_{ch}$	Loi normale (1200 ; 20)	[W]
Cvl – Béton banché	Loi normale (2120 ; 212)	[kJ/(m <sup>3</sup> .K)]
$Alb$	Uniforme (0,28 ; 0,42)	[-]
Con – Polyst W	Loi normale (0,03 ; 0,003)	[W/(m.K)]

Suite à la définition des distributions de probabilité *a priori*, il est possible d'estimer, à partir des données mesurées, les distributions de probabilité *a posteriori* en utilisant l'algorithme ABC-PMC présenté au §2.4.4. Tout d'abord, pour chaque simulation  $i$ , la distance  $\rho(T_i^{sim}, T^{mes}) = RMSE_i$  a été calculée sur les données de calibrage correspondant aux scénarios 1, 2, 3 et 5. Ensuite, une séquence décroissante de seuils de tolérance de  $\epsilon_1 = 20^\circ C$  à  $\epsilon_C = 0,4^\circ C$  a été considérée. Enfin, le nombre de valeurs de paramètres satisfaisant à chaque étape le seuil de tolérance ( $N_{ABC}$ ) a été fixé à 300. L'utilisation de l'algorithme de calibrage a nécessité six heures de calcul.

Les résultats sont présentés de la Figure 38 à la Figure 43. Sur chaque figure sont représentées la distribution de probabilité *a priori*, les 300 valeurs du paramètre sélectionnées par l'algorithme et la distribution de probabilité *a posteriori* estimée. Pour chaque paramètre, un test d'ajustement à une loi normale (test du Khi deux) a été réalisé sur les 300 valeurs du paramètre retenues. Seules les valeurs de paramètre obtenues pour l'albédo ne satisfont pas ce test.

On peut tout d'abord constater que les distributions de probabilité *a posteriori* du débit de ventilation  $Q_{v,nom-1}$ , et de la puissance dissipée  $P_{di}$  sont très proches de leurs distributions *a priori* (Figure 38 et Figure 39) avec un ajustement de 2 % de l'espérance dans les deux cas. Pour la puissance de chauffage (Figure 40) on constate que l'espérance de la distribution de probabilité passe de 1,2 kW à 1,23 kW (ajustement de 3 %) avec une légère augmentation de l'écart-type. Le constat est différent en ce qui concerne la capacité thermique volumique du béton banché (Figure 41) où l'on note que l'espérance de la distribution de probabilité passe de 2120 kJ/(m<sup>3</sup>.K) à 2586 kJ/(m<sup>3</sup>.K) (20 % d'écart) avec une légère diminution de l'écart-type. Cette différence peut s'expliquer par le fait que le calibrage cherche à compenser des erreurs non prises en compte. Par exemple, les capacités thermiques volumiques de la chape béton et de la dalle pleine n'ont pas été retenues comme paramètres à calibrer dans le processus de calibrage. Les résultats pour l'albédo ne sont quant à eux pas probants (Figure 42). En effet, les valeurs du paramètre retenues par l'algorithme ne satisfont pas le test d'ajustement à une loi normale. L'utilisation d'un intervalle d'incertitude trop restreint peut expliquer ce résultat. On peut par exemple supposer que la loi

### Chapitre 3 : Calibrage de modèle

*a posteriori* de l'albédo aurait pu être une loi normale si l'intervalle d'incertitude avait été compris entre 0,2 et 0,4. On peut cependant remarquer que toutes les valeurs du paramètre sont inférieures à 0,38. Par conséquent, la distribution de probabilité *a posteriori* de l'albédo considérée dans la suite de l'étude sera une loi uniforme entre 0,28 et 0,38. Enfin, pour la conductivité thermique du polystyrène extrudé, on peut noter une diminution de l'espérance de la distribution de probabilité *a posteriori* par rapport à la distribution *a priori*, avec respectivement 0,027 W/(m.K) et 0,03 W/(m.K), avec une légère augmentation de l'écart-type.

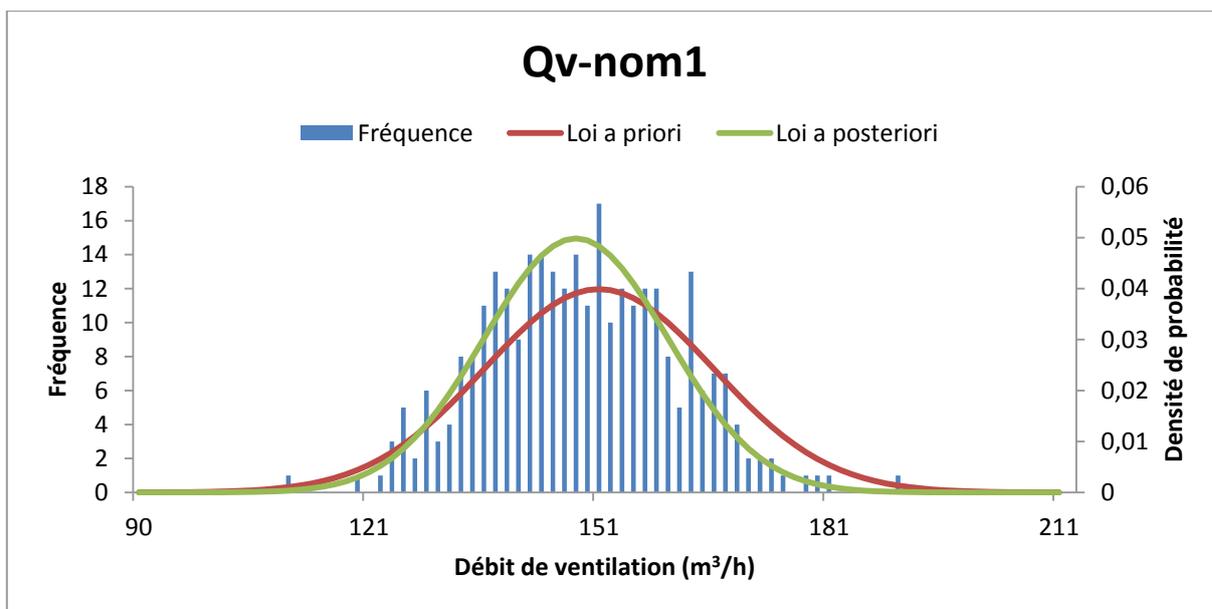


Figure 38 : Distributions de probabilité *a priori* et *a posteriori* de  $Q_{v,nom-1}$

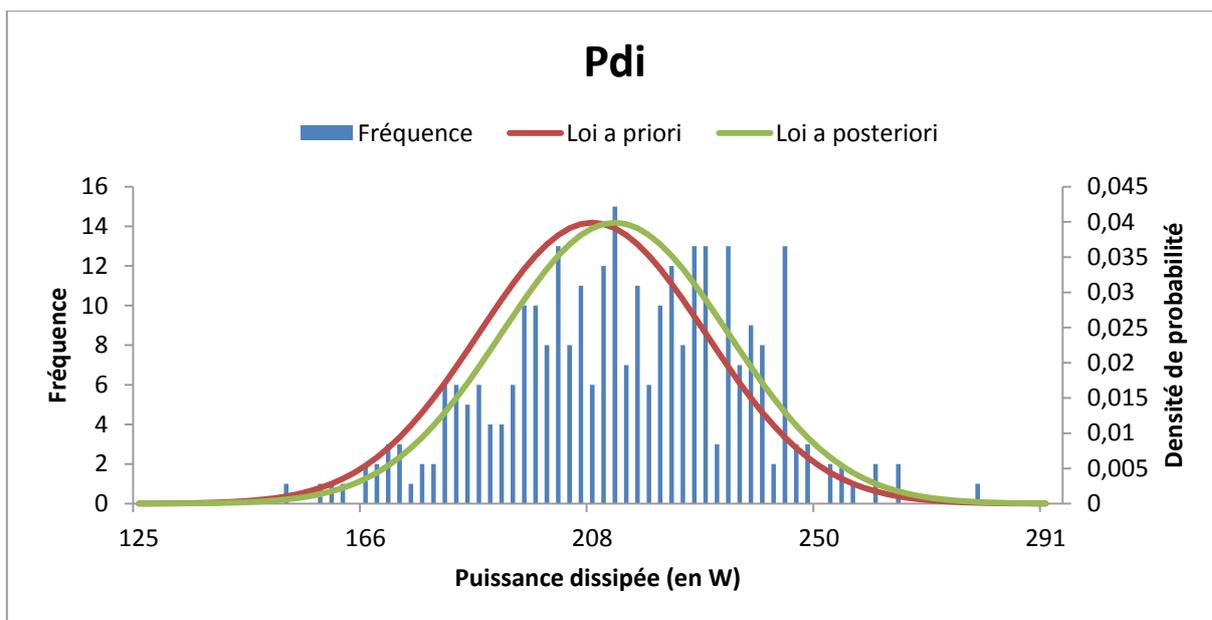


Figure 39 : Distributions de probabilité *a priori* et *a posteriori* de  $P_{di}$

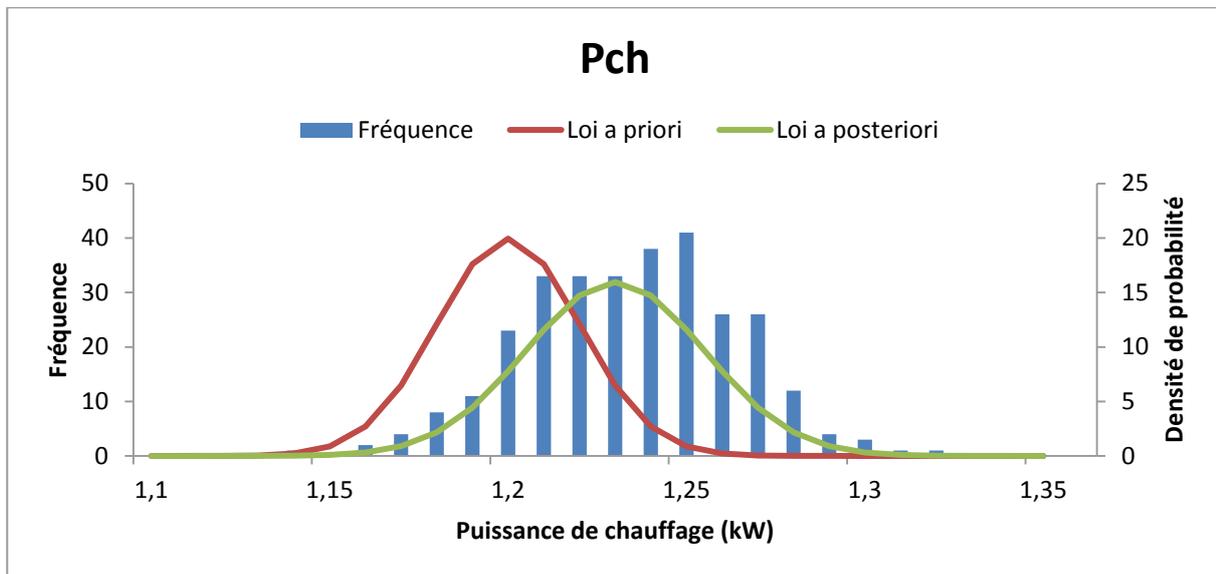


Figure 40 : Distributions de probabilité *a priori* et *a posteriori* de *Pch*

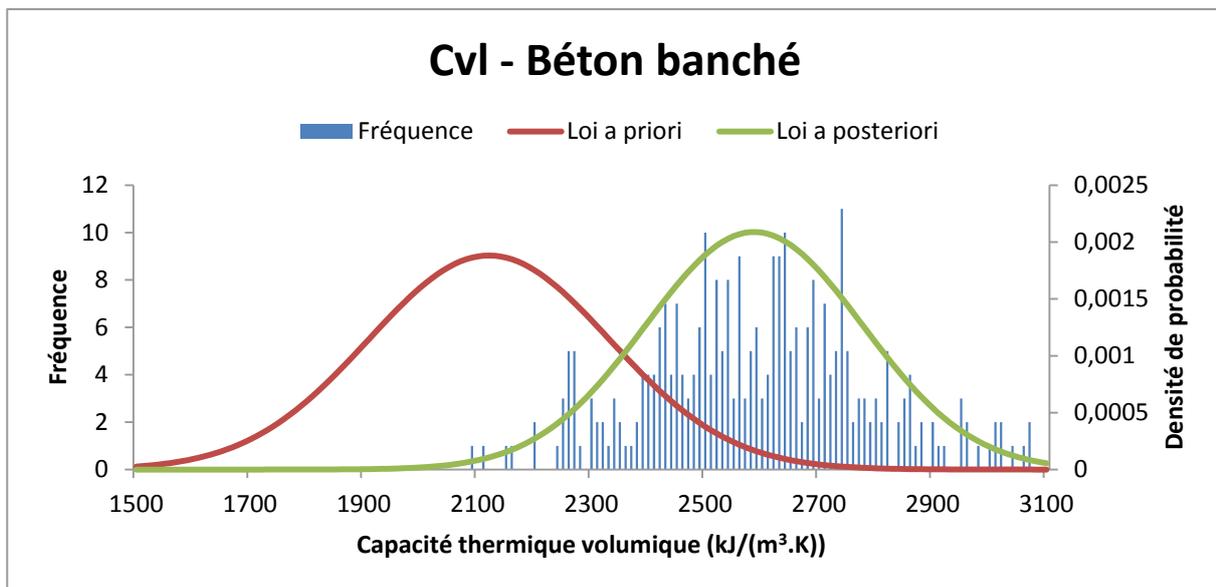


Figure 41 : Distributions de probabilité *a priori* et *a posteriori* de *Cv1* – Béton banché

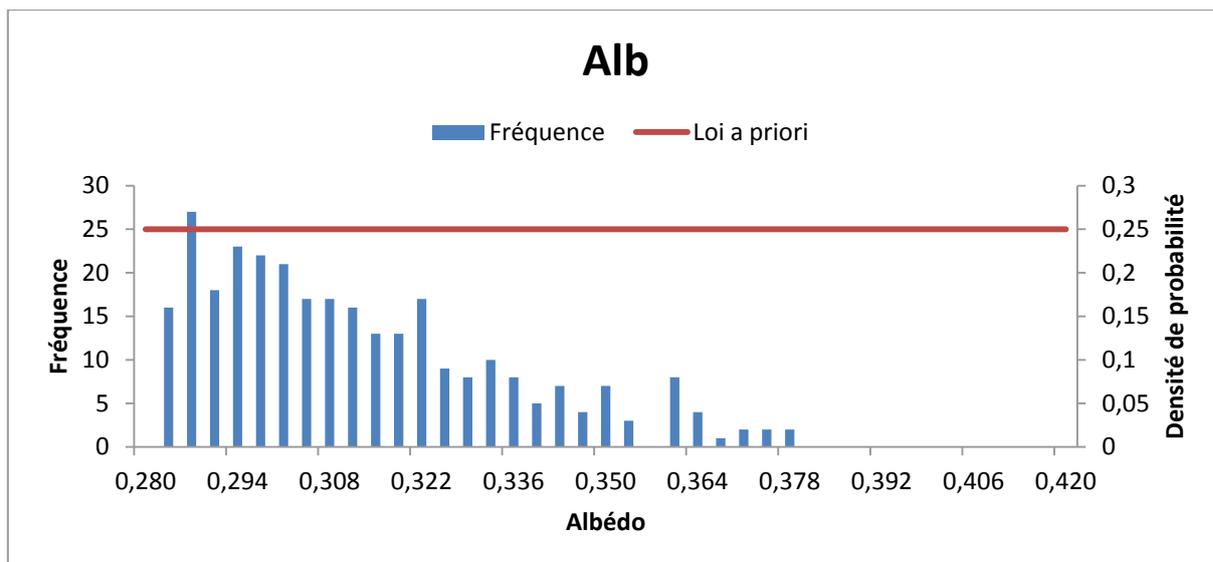


Figure 42 : Distributions de probabilité *a priori* et *a posteriori* de *Alb*

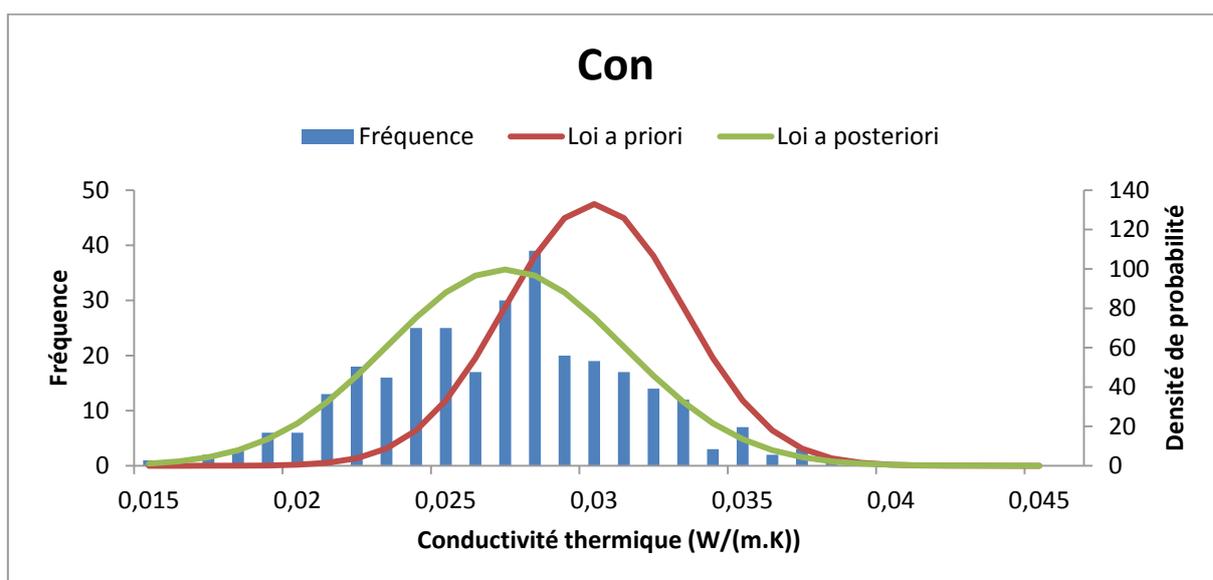


Figure 43 : Distributions de probabilité *a priori* et *a posteriori* de *Con – Polyst W*

L'ensemble des distributions de probabilité *a posteriori* déterminées lors de cette phase de calibrage est présenté dans le Tableau 12. Il est important de noter que les distributions de probabilité *a posteriori* estimées sont jointes, il n'est pas possible de les considérer séparément. Plus précisément, les distributions de probabilité *a posteriori* sont issues d'une distribution commune multivariée et sont par conséquent corrélées. Cela signifie qu'une valeur d'un paramètre coïncide avec des valeurs spécifiques des autres paramètres. Par conséquent, les distributions de probabilité *a posteriori* doivent être appliquées conjointement lors de la propagation des sources d'incertitudes. Il faut également rappeler que les distributions de probabilité *a posteriori* estimées par l'algorithme peuvent ne pas être physiques. On peut par exemple obtenir une puissance de chauffage estimée supérieure à la documentation technique.

Tableau 12 : Distributions de probabilité *a posteriori* estimées

Paramètre	Loi <i>a posteriori</i>	Unité
$Q_{v,nom-1}$	Loi normale (147,7 ; 11,8)	[m <sup>3</sup> /h]
$Pdi$	Loi normale (212 ; 21)	[W]
$Pch$	Loi normale (1230 ; 29)	[W]
Cvl – Béton banché	Loi normale (2586 ; 191)	[kJ/(m <sup>3</sup> .K)]
$Alb$	Uniforme (0,28 ; 0,38)	[-]
Con – Polyst W	Loi normale (0,027 ; 0,004)	[W/(m.K)]

Suite à l'estimation des distributions de probabilité *a posteriori*, il est possible d'évaluer la performance du processus de calibrage grâce à une propagation des sources d'incertitudes.

## 4.5 Évaluation des résultats (étape 5)

La définition des distributions de probabilité *a priori* et *a posteriori* permet de caractériser les incertitudes liées aux facteurs d'entrée du modèle avant calibrage et après calibrage. Ces dernières sont ensuite propagées selon une méthode de Monte-Carlo consistant à réaliser un grand nombre de simulations avec un tirage aléatoire des paramètres selon leur distribution. Pour son efficacité reconnue, un tirage stratifié latin hypercube dit « LHS » (*Latin Hypercube Sampling*) a été utilisé (Munaretto 2014). Les sorties du modèle (ici la température de la zone) sont relevées à chaque simulation afin d'obtenir leur distribution et d'analyser les résultats. Pour obtenir une représentation graphique des résultats de la propagation des sources d'incertitudes, pour chaque de simulation  $i$  le profil de distribution des températures simulées  $T_i^{sim}(k\Delta t)$  à un pas de temps fixé  $k\Delta t$  ( $k \in \mathbb{Z}$ ) est approché par une loi normale d'espérance  $\mu_{i,k}$  et d'écart-type  $\sigma_{i,k}$ . La moyenne sur toutes les simulations des  $\mu_{i,k}$  est une série temporelle représentée par la courbe  $C_\mu$ . De la même manière,  $C_\sigma$ , construite à partir des  $\sigma_{i,k}$ , représente la courbe des écarts-types à la moyenne. La courbe  $C_\mu$  est alors encadrée par les courbes  $C_{\mu+2\sigma}$  et  $C_{\mu-2\sigma}$  valant respectivement  $C_\mu$  plus ou moins deux fois  $C_\sigma$  (Recht et al. 2014).

Les résultats de la propagation des sources d'incertitudes sur le modèle avant calibrage et après calibrage sont représentés respectivement sur la Figure 45 et la Figure 47. Pour rappel, dans le but d'effectuer une comparaison entre l'outil de simulation COMFIE et l'expérimentation, une bande d'incertitudes de  $\pm 1^\circ C$  a été considérée sur la mesure.

Pour le modèle avant calibrage (Figure 45 et Figure 46), on constate que sur le scénario n°1 la simulation a tendance à sous-estimer la température intérieure en fin de simulation. Sur les scénarios n°2 et 3, on observe une tendance du modèle à sous-estimer la température intérieure bien que les dynamiques soient similaires. Sur le scénario n°4, on constate que la température simulée sous-estime la température mesurée sur l'ensemble du scénario. Cependant en fin de période, les bandes d'incertitudes sont confondues. Sur le scénario n°5, correspondant à l'ouverture des volets, l'amplitude journalière de variation des températures augmente sensiblement. On peut remarquer que celle des températures simulées est plus réduite que celle des températures

### Chapitre 3 : Calibrage de modèle

mesurées. Enfin, sur le dernier scénario, l'analyse est similaire : on observe des profils simulés et mesurés proches avec des amplitudes plus faibles concernant les résultats de simulation. Sur l'ensemble de ces résultats, on peut constater les bonnes performances du modèle non calibré qui permet de reproduire de manière satisfaisante l'évolution de la température au sein du bâtiment.

Pour le modèle après calibrage (Figure 47 et Figure 48), on constate que sur les scénarios utilisés pour le calibrage (scénarios n°1, 2, 3 et 5) les bandes d'incertitudes des profils simulés et mesurés sont proches voire quasiment confondues, en particulier pour les scénarios n°2 et 3. Sur les données de validation, correspondant aux scénarios n°4 et 6, le constat est cependant plus mitigé. Sur le scénario n°4, on peut observer que le calibrage n'apporte que des améliorations à la marge. On constate ainsi que les profils mesurés et simulés sont plus proches en début de scénario. Sur le scénario n°6 on remarque l'apport du calibrage qui permet de diminuer l'incertitude du profil simulé sur la fin du scénario.

Pour quantifier les résultats, une RMSE moyenne (correspondant à la moyenne des RMSE calculées sur chaque simulation) entre les températures simulées et les températures mesurées a été calculée sur les données de validation (scénarios n°4 et 6). Les résultats sont présentés dans la Figure 44. Comme précédemment, on peut constater que sur le scénario n°4 le calibrage n'apporte qu'une amélioration à la marge. Cependant, sur le scénario n°6 le calibrage permet d'obtenir une RMSE de 0,5°C après calibrage contre une RMSE de 0,6°C avant calibrage.

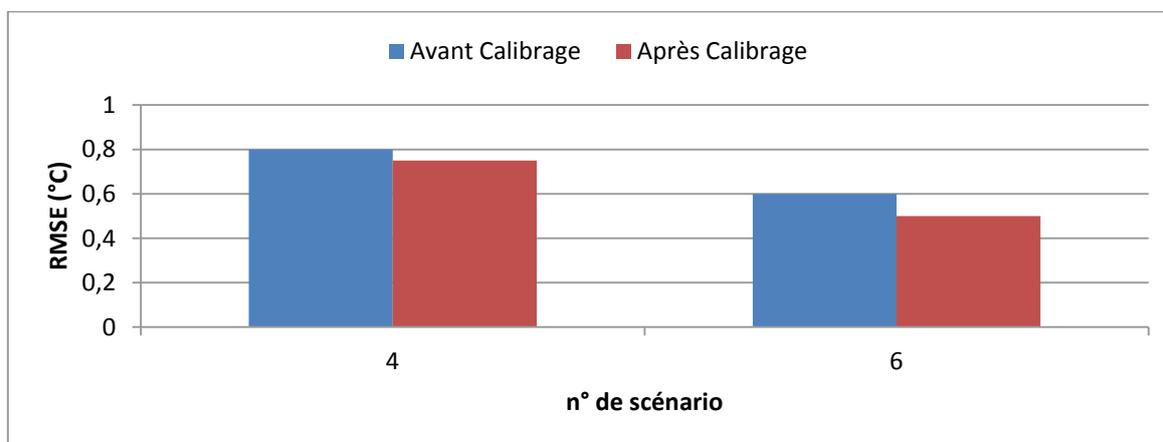


Figure 44 : Comparaison la RMSE moyenne avant et après calibrage

La faible diminution du RMSE suite à la phase de calibrage peut s'expliquer de plusieurs manières. Tout d'abord le modèle considéré dans cette étude est un modèle monozone dont la précision est inférieure à celle d'un modèle multizone. Ensuite, dans l'objectif de comparer les températures simulées avec les températures mesurées, ces dernières ont été agrégées au prorata des surfaces intérieures des pièces du bâtiment, amenant également à une diminution de la précision. Enfin, dans ce cas d'application, seules des lois normales ou uniformes ont été utilisées.

On peut envisager l'utilisation de lois normales tronquées qui permettrait d'une part d'obtenir des distributions de probabilité *a posteriori* plus fines et d'autre part de s'assurer d'obtenir des distributions de probabilité *a posteriori* plus réalistes en évitant des valeurs physiques aberrantes.

### Chapitre 3 : Calibrage de modèle

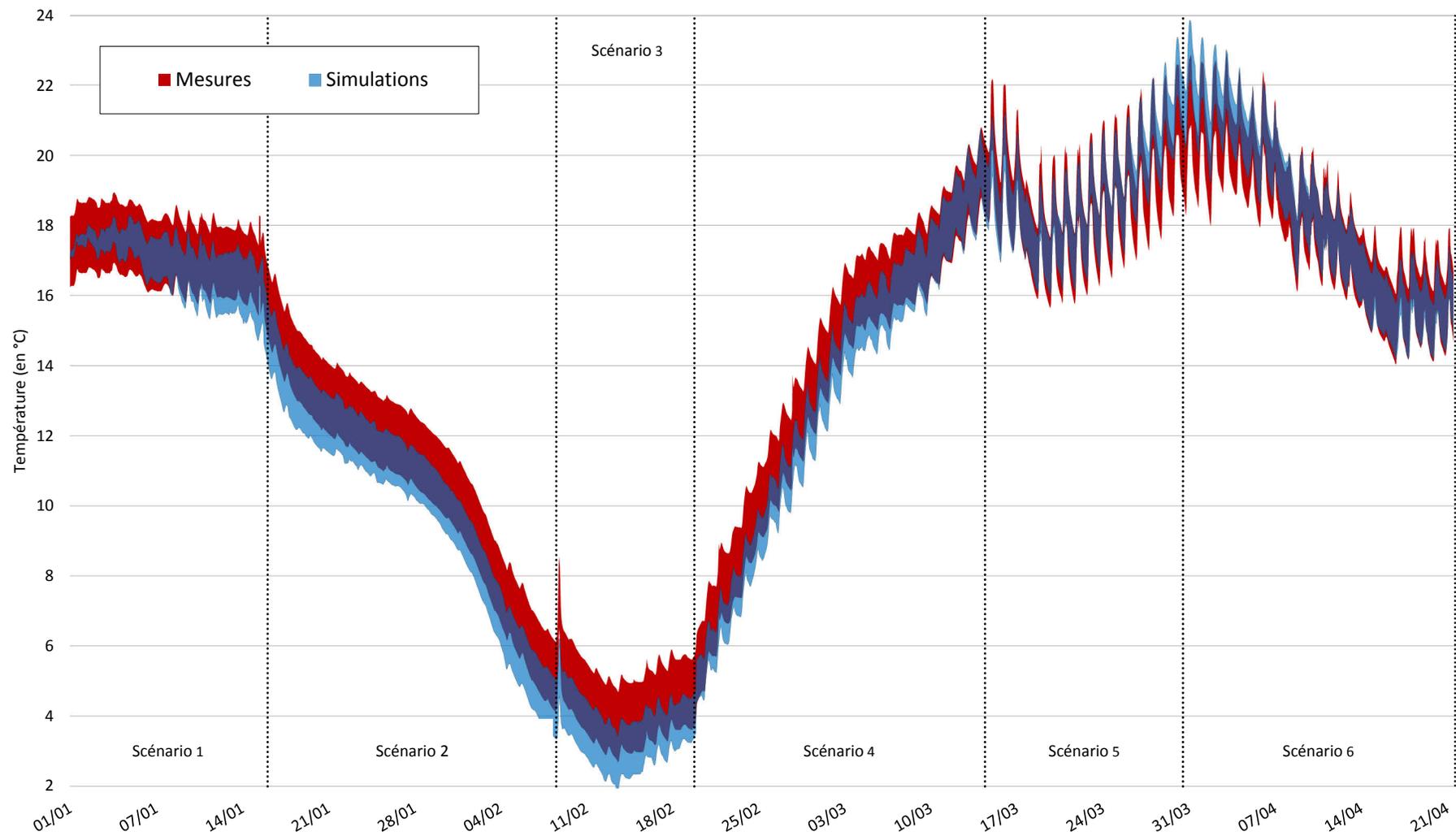


Figure 45 : Propagation des sources d'incertitudes avant calibrage

### Chapitre 3 : Calibrage de modèle

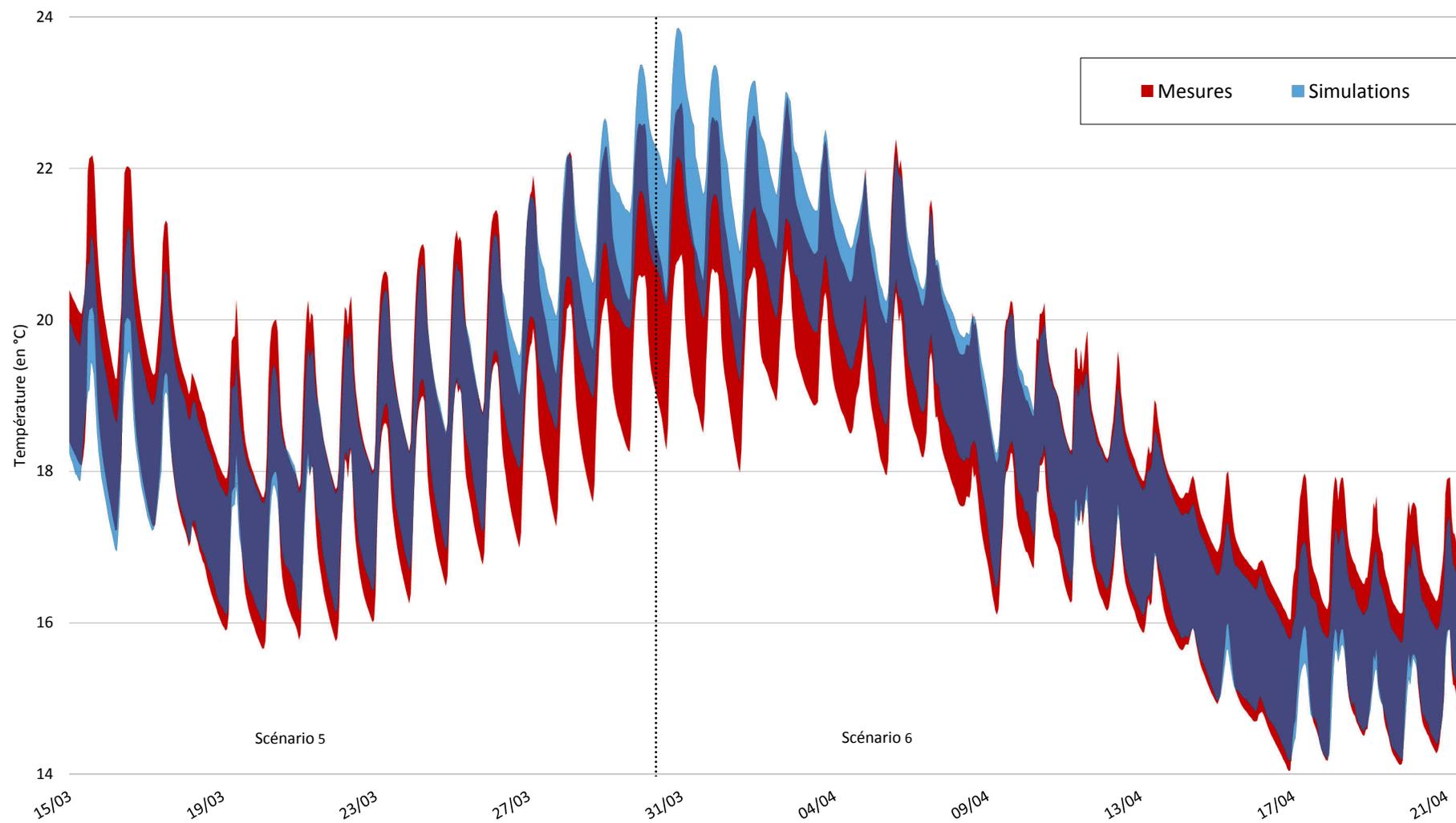


Figure 46 : Zoom sur les scénarios n°5 et n°6 (avant calibrage)

### Chapitre 3 : Calibrage de modèle

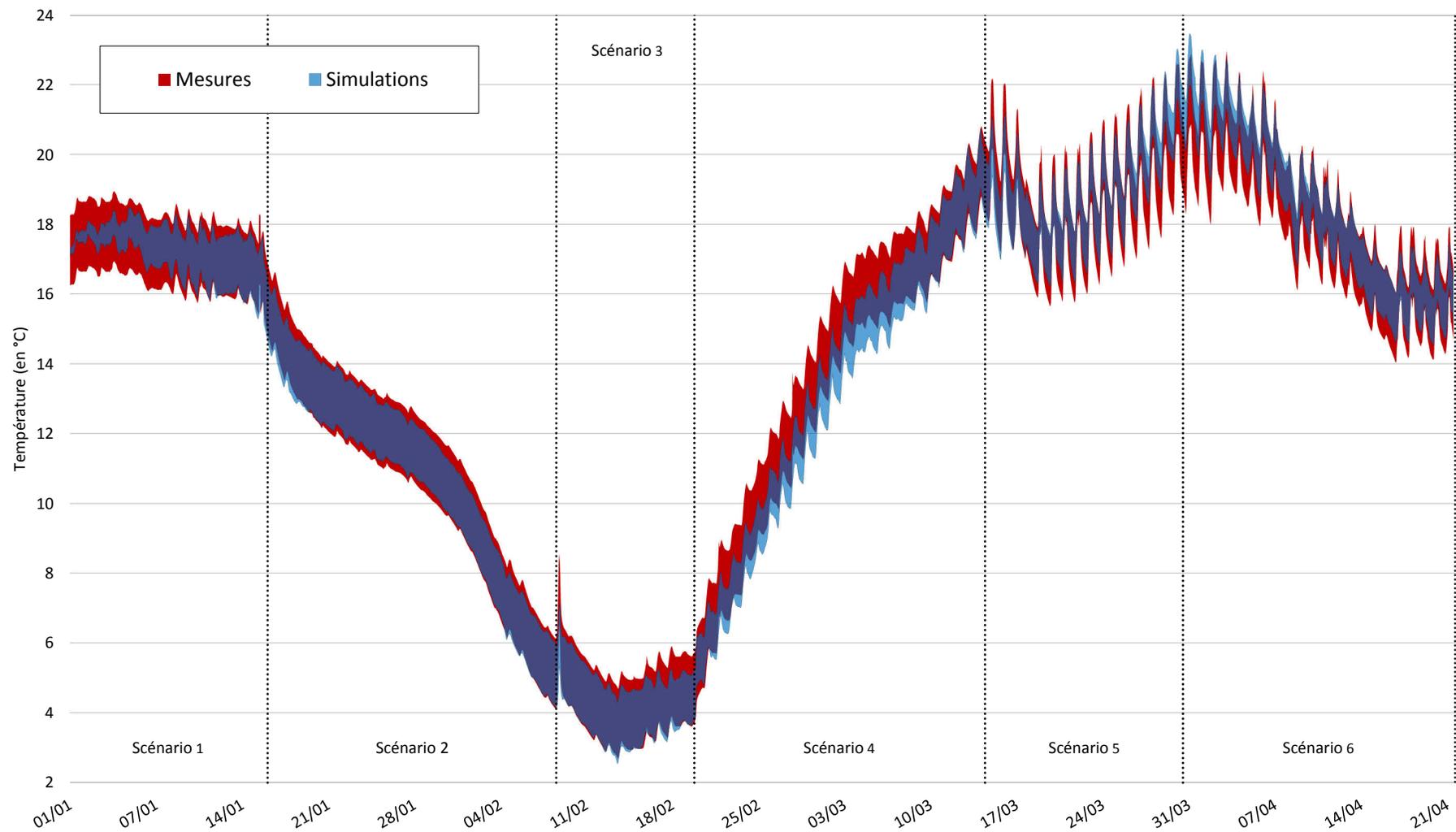


Figure 47 : Propagation des sources d'incertitudes après calibrage

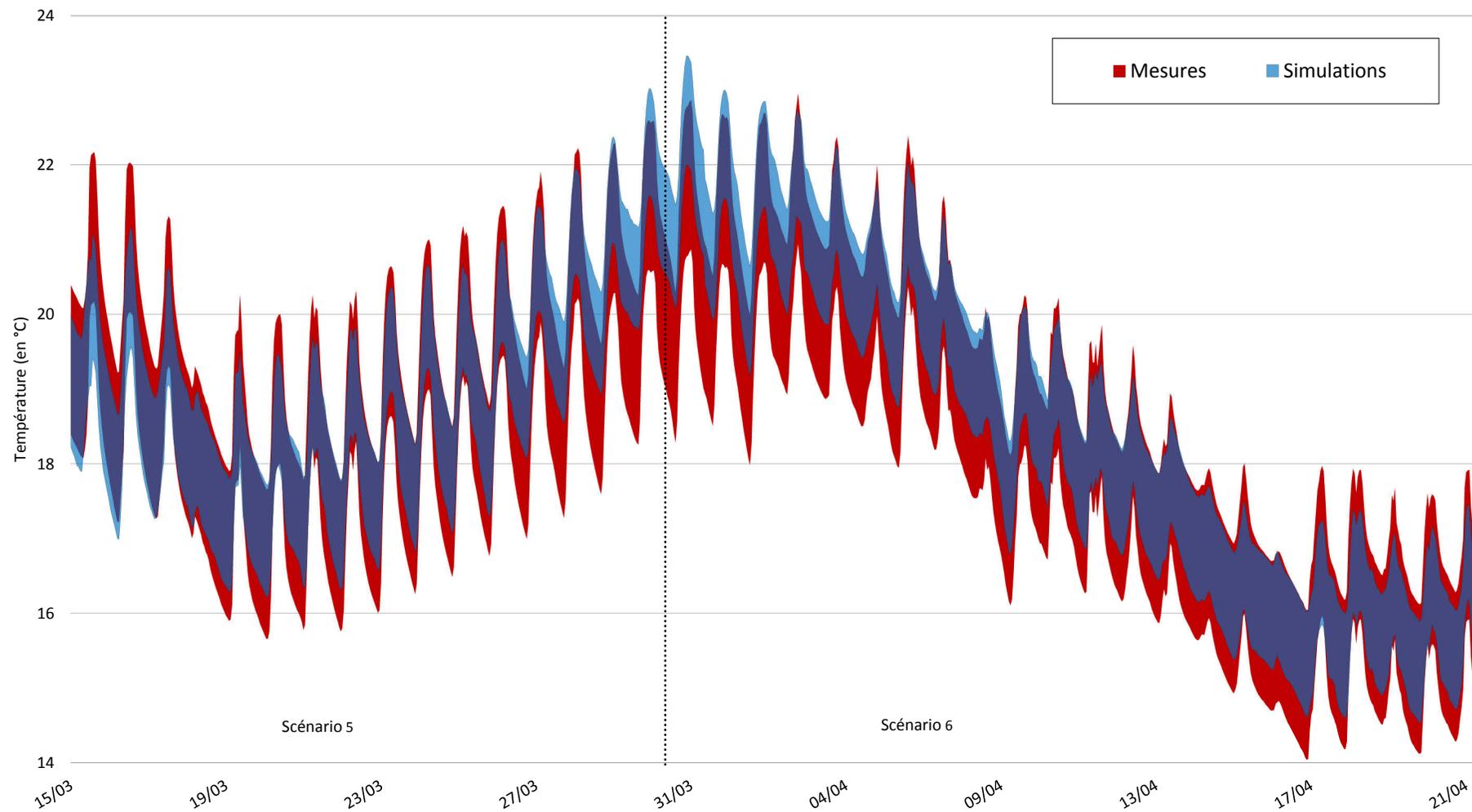


Figure 48 : Zoom sur les scénarios n°5 et n°6 (après calibrage)

## 5. CONCLUSION DU CHAPITRE

Dans ce chapitre, un rapide état de l'art a permis de présenter les principales méthodes de calibrage existantes, à savoir les méthodes heuristiques, les méthodes avancées et les méthodes bayésiennes. Nous nous sommes ensuite concentrés sur les méthodes bayésiennes car elles présentent l'avantage de considérer l'ensemble des sources d'incertitudes et d'erreurs pouvant expliquer les différences entre les données simulées et les données mesurées. Ces méthodes se placent dans le cadre de l'approche statistique du calibrage dont l'objectif est d'estimer des distributions de probabilité *a posteriori* des paramètres incertains. Les méthodes bayésiennes reposent sur le théorème de Bayes permettant d'estimer la distribution de probabilité *a posteriori* et se différencient selon la connaissance ou non de la fonction de vraisemblance. Dans l'objectif de définir une méthode de calibrage la plus générique possible (s'appliquant par exemple à des données simulées temporellement corrélées), notre choix s'est porté sur l'utilisation du calcul bayésien approché permettant d'éviter de connaître analytiquement la fonction de vraisemblance. Par la suite, une méthodologie a été présentée consistant à collecter l'ensemble des données nécessaires à la modélisation, à déterminer les paramètres les plus influents par une analyse de sensibilité, à estimer les distributions de probabilité *a posteriori* des paramètres influents grâce à l'algorithme ABC-PMC, et enfin à évaluer les résultats par le biais d'une propagation d'incertitudes.

Les résultats ont montré qu'il était possible d'estimer des distributions de probabilité *a posteriori* des paramètres influents, et d'améliorer le comportement du modèle après calibrage. Cependant, les améliorations sont faibles par rapport au modèle non calibré. Des études complémentaires seraient nécessaires pour évaluer les performances de la méthode développée. Cependant, plusieurs raisons peuvent expliquer ces faibles améliorations, telles que l'utilisation d'un modèle monozone du bâtiment et de données agrégées de la température intérieure.

Les avantages de la méthode de calibrage définie dans ce chapitre sont multiples. Tout d'abord, dans le cadre de la garantie de performance, cette méthode fournit d'une part des distributions de probabilité *a posteriori* des paramètres influents et non des valeurs uniques estimées par moindre carré. Ces distributions peuvent alors être utilisées pour propager les sources d'incertitudes et évaluer leur impact sur la grandeur d'intérêt considérée (consommation de chauffage, évolution de la température intérieure, etc.). D'autre part, elle permet de s'assurer de disposer d'un modèle fiable pour évaluer par exemple l'impact de rénovations énergétiques sur le comportement du bâtiment. De plus, cette méthode de calibrage peut également être utilisée pour calibrer un modèle de bâtiment utilisé dans le cadre de la régulation. Il suffit alors de considérer l'espérance mathématique de la distribution de probabilité *a posteriori* de chaque paramètre. Enfin, cette méthode de calibrage est générique et peut être utilisée sur n'importe quel modèle de bâtiment tout en utilisant des mesures diverses (puissance de chauffage, facture d'énergie, température intérieure, etc.). En perspectives, l'emploi de lois normales tronquées pour caractériser les distributions de probabilité *a priori* permettrait de s'assurer du respect des contraintes physiques des paramètres à calibrer. La

### Chapitre 3 : Calibrage de modèle

méthode présentée dans ce chapitre permet d'envisager d'utiliser ce type de loi. De même, cette méthode pourrait être utilisée pour de la détection de défauts en orientant les campagnes de mesure à mener sur les paramètres dont la distribution de probabilité *a posteriori* ne semble pas correspondre avec la réalité.

Suite au calibrage du modèle thermique du bâtiment, nous nous intéresserons dans le chapitre 4 au développement de méthodes d'optimisation hors-ligne permettant d'approximer les résultats de la commande prédictive dans le but d'identifier les lois de commande simplifiées pouvant être embarquées dans les systèmes de régulation couramment utilisés par les professionnels du bâtiment.

