

Méthode d'échantillonnage pour la prise en compte des incertitudes des données météorologiques pour l'analyse de sensibilité globale sur la performance énergétique

Jeanne Goffart^{*1,2}, Thierry Mara², Etienne Wurtz³

¹ LOCIE UMR 7152, Université de Savoie

Le Bourget-du-Lac

² PIMENT, Université de la Réunion

Saint Denis, Réunion

³ LEB, CEA-INES

Le Bourget-du-Lac

*jeanne.goffart@gmail.com

RESUME. Cet article propose une méthode originale d'échantillonnage des entrées temporelles corrélées afin de permettre leur prise en compte dans les analyses de sensibilité et d'incertitude en simulation thermique dynamique. La démarche s'appuie sur les données météorologiques de type TMY associées à EnergyPlus. Ces entrées sont le rayonnement direct normal et diffus horizontal, la température, la vitesse et la direction du vent ainsi que l'humidité relative. L'article présente les résultats du processus de génération ainsi qu'une comparaison avec des observations météo sur 10 ans. Une étude de cas est faite sur une maison passive (INCAS-DM). Les résultats de l'analyse d'incertitude, pour le climat de Lyon au mois de janvier, révèlent une incertitude de l'ordre de 20 % du besoin de chauffage à 95% de niveau de confiance. La température et le rayonnement direct sont les entrées les plus influentes sur l'incertitude des besoins de chauffage.

MOTS-CLÉS. Analyses de sensibilité et d'incertitude, Fiabilité, Données météorologiques, Bâtiment basse consommation

ABSTRACT. This paper proposes a new sampling method for dynamic inputs in order to allow their inclusion in the sensitivity and uncertainty analyses in dynamic thermal simulation. The approach is based on the weather data file associated with EnergyPlus, as TMY file. These inputs are the direct and the diffuse radiation, the temperature, the relative humidity, the wind speed and direction. The article presents the results of the sampling process and a comparison with real weather data over 10 years. The methodology is applied on a passive house (INCAS-DM). The results of the uncertainty analysis for the climate of Lyon in January, show a uncertainty of about 20% of the heating needs at a confidence level of 95 %. The temperature and the direct solar radiation are the most influential inputs on the uncertainty of heating needs.

KEYWORDS. Sensitivity and Uncertainty Analyses, Reliability, Weather Data, Low-consumption Building

1 INTRODUCTION

Dans le contexte actuel de course à la performance pour les bâtiments basse consommation d'énergie, on observe l'émergence d'un besoin industriel fort sur la thématique de la garantie de performance. Pour répondre à cette problématique, il est nécessaire d'estimer un intervalle de confiance sur les résultats de simulation et d'identifier les entrées incertaines responsables de l'incertitude de sortie. Les techniques d'analyse de sensibilité et d'incertitude (Saltelli et al. (2008)) permettent ce type d'étude. Leur utilisation se démocratise en thermique du bâtiment permettant un gain en transparence et en fiabilité en Simulation Thermique Dynamique (STD). A ce propos, Tian (2013) dresse un état de l'art des différentes applications dans le cadre du bâtiment et des systèmes. Cependant ces études se limitent aux entrées statiques telles que les paramètres associés à la composition du bâtiment modélisé (ex : épaisseur d'isolant). Les incertitudes relatives aux sollicitations comme la météo ne sont pas prises en compte. La variabilité annuelle de la météo est par exemple négligée. Le modèle de bâtiment est évalué à partir d'une année moyenne représentative du climat à long terme (TMY ou IWEC pour l'Europe). Pour le dimensionnement un fichier représentatif d'un extrême peut être utilisé, permettant d'évaluer le comportement limite du bâtiment. Or, on souhaite estimer une probabilité d'événement plutôt qu'une valeur moyenne de consommation ou une valeur extrême éventuellement peu probable sur 20 ans de fonctionnement et menant à un sur-dimensionnement des systèmes énergétiques. Afin d'aller vers la garantie de performance, une prise en compte de la variabilité annuelle des données météorologiques est primordiale dans l'évaluation d'un intervalle de confiance sur la sortie. Les variables météorologiques doivent être intégrées dans les analyses de sensibilité et d'incertitude. Cependant, la prise en compte des entrées temporelles et corrélées comme les données météorologiques est un verrou scientifique pour les analyses de sensibilité. Il s'agit d'un thème de recherche actuel de la communauté des statistiques comme le montre les travaux de Chastaing et al. (2012). Pour permettre d'évaluer les entrées responsables de l'incertitude de sortie, les méthodes d'analyse de sensibilité standard nécessitent une indépendance dans la génération des échantillons.

Ainsi, la complexité de la tâche réside dans l'échantillonnage de fichiers météo qui doivent être tout d'abord cohérents avec la physique des phénomènes météo, représentatifs de la variabilité annuelle d'un site donné, et qui doivent également respecter des contraintes statistiques pour permettre l'analyse de sensibilité, i.e. l'indépendance d'échantillonnage. Cet article propose une méthodologie pour répondre à ces problématiques. Cette stratégie est tout d'abord présentée puis une application sur le climat de Lyon permet de comparer l'incertitude générée avec des données mesurées. Enfin les variables météorologiques influentes sur les besoins de chauffage d'un bâtiment passif sont estimées par analyse de sensibilité.

2 MÉTHODOLOGIE POUR LA GESTION DES CORRÉLATIONS DANS L'ÉCHANTILLONNAGE DE LA MÉTÉO

2.1 PRINCIPE D'UN ÉCHANTILLONNAGE POUR LES ANALYSES D'INCERTITUDE ET DE SENSIBILITÉ

Le schéma en figure 1 illustre le principe de l'analyse d'incertitude dans le cas simple des entrées statiques. Un échantillon de N jeux de simulation est généré en tirant aléatoirement et indépendamment des valeurs de chaque entrée X_i à partir de la définition de leur incertitude (loi de distribution ou fonction de répartition) qui est représentée par une loi normale sur le schéma. N valeurs de la sortie (dans ce cas scalaire) sont obtenues après propagation de la matrice d'échantillon des entrées. A partir de ces N valeurs de Y on établit l'intervalle de confiance

(l'incertitude) sur la sortie. On calcule ensuite les indices de sensibilité pour identifier les entrées responsables de la dispersion de sortie.

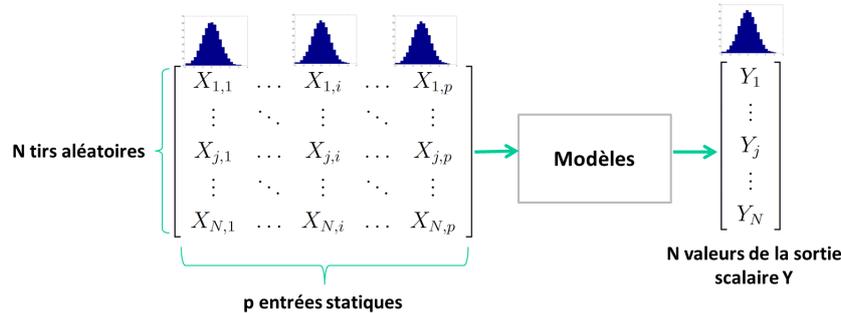


FIGURE 1. Schéma du principe de l'analyse d'incertitude.

Si on assimile le cas ci-dessus avec p entrées statiques à k variables météorologiques de N_t pas de temps (i.e. $N_t = 744$ pour un mois de 31 jours au pas de temps horaire), chaque $X_{j,i}$ représenté en figure 1 n'est plus une valeur mais un vecteur et chaque X_i est associée à une variable météorologique qui est une matrice de $N \times N_t$. Chaque matrice représente une variable météorologique : température, humidité, rayonnement, etc. Nous souhaitons évaluer l'influence de chaque groupe et donc chaque variable météorologique sur le mois entier et non l'influence à chaque pas de temps. Pour cela on évalue les indices de sensibilité par groupe de Sobol (Sobol et al. (2007)). Cette méthode permet d'évaluer la part de variance (les variations) de sortie induite par la variance de chaque groupe d'entrées (i.e. chaque variable météo). Une des conditions forte est l'indépendance dans l'échantillonnage de chaque groupe. On obtient l'effet seul du groupe ou de l'entrée X_i sur Y tel que S_i soit l'indice du premier ordre (Sobol (1993)) :

$$S_i = \frac{V(E[Y|X_i])}{V(Y)} \quad (1)$$

avec $V(\cdot)$ la variance et $E(\cdot)$ l'espérance mathématique. Les indices de sensibilité de Sobol sont robustes, populaires, intuitifs et particulièrement recommandés dans le cas de modèles complexes de grande dimension de système incertain comme la thermique du bâtiment.

2.2 MÉTHODE D'ÉCHANTILLONNAGE DÉVELOPPÉE

Afin de réaliser une analyse de sensibilité en prenant en compte les variables météorologiques, on doit échantillonner chaque variable. Par cela on entend générer pour chaque simulation des séquences météorologiques différentes pour induire une variance mais cohérentes avec la variabilité annuelle des données climatiques du site. La génération doit également être compatible avec les méthodes d'analyse de sensibilité standard en respectant l'hypothèse d'indépendance entre chaque groupe d'entrée. Dans ce but nous avons développé une méthode qui consiste à générer pour une variable donnée et pour chaque simulation une séquence temporelle aléatoire respectant l'auto-corrélation mais générée indépendamment des autres variables météo. On dispose de groupes de variable indépendants. Pour pouvoir respecter la physique du phénomène et les corrélations entre variables, on rajoute à chaque tir une structure corrélative (entre variable) qui est identique pour chaque séquence aléatoire d'une même variable et qui est donc "muette" pour l'analyse de sensibilité. La théorie de cette approche est présentée ci-dessous.

On considère un processus aléatoire que l'on note $x_i(\theta, t)$ où t représente le temps. La variable aléatoire θ représente le caractère stochastique de la fonction. On peut définir ce processus aléatoire par une moyenne $\bar{x}_i(t)$ à laquelle on rajoute une variable aléatoire $\varepsilon_i(\theta, t)$ telle que :

$$x_i(\theta, t) = \bar{x}_i(t) + \varepsilon_i(\theta, t). \quad (2)$$

On suppose que la moyenne est déterministe alors que la variable aléatoire est définie par une fonction de covariance (ou auto-corrélation) $C_{ii}(t, t')$ et une fonction de répartition $h_i(t)$. Comme \bar{x}_i est déterministe, la variabilité de x_i est due à ε_i . La fonction d'auto-corrélation représente la dépendance du signal avec lui-même. $C_{ii}(t, t')$ est supposée décroissante et fonction de $|t - t'|$. On peut extraire $C_{ii}(t, t')$ et $h_i(t)$ à partir d'une séquence météorologique donnée. Soit des fichiers météo au pas de temps horaire. On estime la moyenne horaire, la fonction d'auto-corrélation $C_{ii}(t_j, t_k)$ et on extrait la fonction de répartition empirique pour chaque pas de temps horaire $h_i(t_j)$, $t_j = 1, \dots, N_t$. Sachant que la moyenne horaire est fixée, la génération d'une séquence cohérente dépend de la génération de la variable aléatoire $\varepsilon_i(\theta, t)$. Pour cela, on utilise la procédure décrite par Iman et Conover (1982). Une application de la méthode est présentée ci-après.

3 APPLICATION DE LA MÉTHODE D'ÉCHANTILLONNAGE POUR LE CLIMAT DE LYON

3.1 EXTRACTION DES CARACTÉRISTIQUE STATISTIQUES À PARTIR D'UN IWEC

Les informations telles que la fonction d'auto-corrélation, la loi de distribution ainsi que la moyenne doivent être obtenues à partir d'un échantillon statistique de données météo. Dans le cas où plusieurs années mesurées sont disponibles la variabilité statistique est évidente. Or nous souhaitons extraire les informations statistiques des fichiers moyens représentatifs tels que le IWEC (International Weather for Energy Calculation), libres d'accès et associés aux logiciels de simulation.

Le procédé d'extraction commence par regrouper les 31 profils horaires du mois de janvier sur une seule journée. A partir de cet aléas de 31 points pour chaque pas de temps horaire, on calcule la fonction d'auto-corrélation, i.e. la dépendance entre chaque pas de temps. Ensuite on estime la moyenne horaire journalière que l'on extrait du signal pour déterminer les caractéristiques statistiques du résidu : il s'agit des 24 fonctions de répartition empiriques horaires qui définissent l'incertitude.

A partir de ces informations, on applique la méthode décrite en section 2.2 et on génère des journées aléatoires typiques de la période et du site sélectionné afin de construire des mois cohérents au regard de cette période (Goffart (2013)). On applique la méthode pour 6 variables, les prédominantes en STD : la température, le rayonnement diffus horizontal, le rayonnement direct normal, la vitesse, la direction de vent et l'humidité. On génère $N = 1000$ mois de janvier, on obtient donc 6 matrices de 1000×744 .

3.2 COMPARAISON DES ÉCHANTILLONS GÉNÉRÉS AVEC DES OBSERVATIONS MÉTÉO-FRANCE

On souhaite estimer la variabilité générée sur les 1000 échantillons et la situer face à la variabilité météorologique annuelle du site sélectionné, au mois de janvier. Pour cela on compare les normales saisonnières des 10 dernières années de Météo-France par rapport à la dispersion des échantillons. La comparaison est faite pour la température moyenne quotidienne (T_m) du mois de janvier ainsi que son ensoleillement mensuel (Ens), ces deux grandeurs sont calculées selon le protocole de l'Organisation Mondiale de la Météorologie. Les résultats sont regroupés sur la figure 2. Chaque graphe est composé de 10 ans d'observations Météo-France de 2004 à 2013 (croix bleues), de la moyenne Météo-France (ligne pointillée rouge) ainsi que la moyenne du IWEC (ligne pointillée bleue) qui est proche de la moyenne des échantillons. La dispersion des échantillons est représentée par l'intervalle à 1.96σ (i.e. 95 % de niveau de confiance) autour de la moyenne IWEC (lignes pointillées vertes).

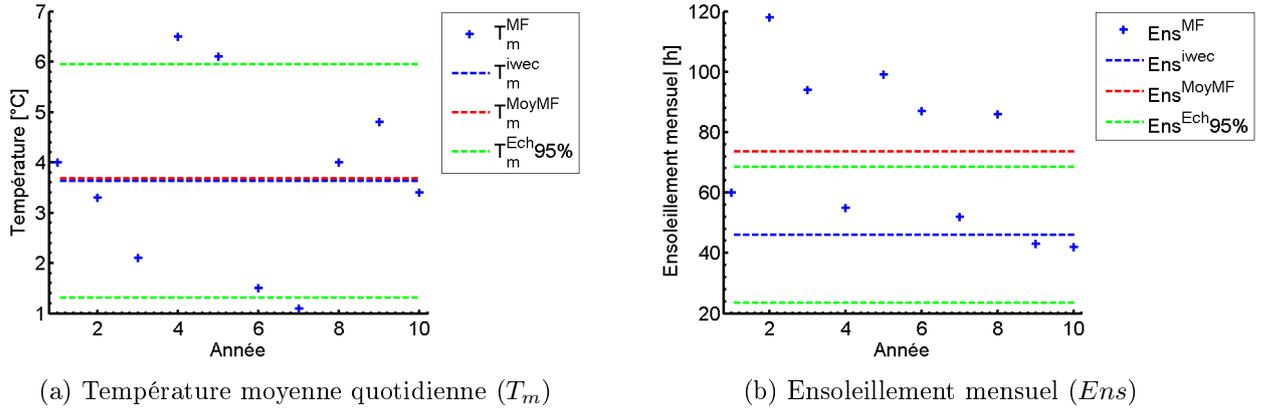


FIGURE 2. Comparaison de la dispersion des 1000 échantillons aux 10 ans d’observations Météo-France de 2004 à 2013 (croix bleues), de la moyenne Météo-France (ligne pointillée rouge) ainsi que la moyenne du IWEC (ligne pointillée bleue). La dispersion des échantillons est représentée par l’intervalle à 1.96σ autour de la moyenne IWEC (lignes pointillées vertes).

Pour la température en Figure 2a, on observe une valeur de la moyenne des observations Météo-France très proche de la valeur moyenne du IWEC (lignes pointillées rouge et bleue). De plus, la majorité des observations sont comprises dans les 95% des valeurs des échantillons, ce qui est cohérent avec une variabilité naturelle. Pour le cas de l’ensoleillement en figure 2b, 5 observations sur les 10 sont comprises dans la variabilité des échantillons. Néanmoins, on sous-évalue le rayonnement solaire dans les échantillons : la moyenne est 30 % plus faible que la valeur moyenne du IWEC. Cela peut être dû à la différence de période entre les observations pour la construction du IWEC (1982-1999) et les observations Météo-France(2004-2013). Cela peut être également dû aux sources de biais importantes pour la mesure et/ou la construction du rayonnement solaire des fichiers IWEC. L’amplitude de l’incertitude semble toutefois cohérente avec l’amplitude naturelle. Les valeurs des dispersions des échantillons sont regroupées dans le tableau 1.

Grandeur	Moyenne μ	Ecart type σ	Intervalle de confiance pour 95%
T_m [$^{\circ}C$]	3,42	1,23	$3,42 \pm 2,41$
Ens [h]	49,9	11,2	$49,9 \pm 22,0$

TABLE 1: Caractéristiques statistiques de la dispersions des 1000 échantillons pour T_m et Ens

La dispersion des échantillons est cohérente avec la variabilité naturelle des observations de Météo-France. La phase de comparaison mesure/échantillon présentée ici est la dernière étape d’un processus de vérification de la méthode, des études préliminaires (Goffart (2013)) ont tout d’abord vérifié que les échantillons respectent les fonctions d’auto-corrélation et de distribution du fichier source IWEC, l’indépendance dans l’échantillonnage de la variable aléatoire ($\varepsilon_i(\theta, t)$), ainsi que la cohérence physique des corrélations entre échantillons due à l’ajout de la moyenne horaire. On observe pour le dernier cas une légère perte d’information dans les corrélations, ce qui est un compromis nécessaire pour permettre le calcul des indices de sensibilité.

4 INCERTITUDES ET DONNÉES MÉTÉO LES PLUS INFLUENTES SUR LA MAISON INCAS-DM

Cette partie présente l’estimation des incertitudes par propagation des échantillons et le calcul des indices de sensibilité sur les besoins de chauffage d’une maison passive.

4.1 CAS ÉTUDIÉ : MAISON INCAS-DM

Le modèle utilisé dans cette étude est celui d'une des maisons présentes sur la plateforme INCAS de l'Institut National d'Énergie Solaire (INES) au Bourget-du-Lac en Savoie (Figure 3). La maison INCAS-DM est une maison basse consommation de stratégie d'économie d'énergie dite "passive" dont l'isolation des murs, des planchers et des plafonds ainsi que l'inertie est importante. La structure est de type Double-Mur (DM) puisqu'elle est composée de 15 cm de parpaing, 20 cm d'isolation et de nouveau de 15 cm de parpaing. Les gains solaires sont maximisés en hiver par les larges baies vitrées du côté Sud. En été le bâtiment se protège de ces gains par des protections solaires dues au débord de toit et au balcon (Figure 3). Le modèle EnergyPlus est constitué de 4 zones thermiques qui correspondent au 4 niveaux du bâtiment : le rez-de-jardin (RDJ), l'étage, le vide sanitaire et les combles, seules les zones RDJ et Etage sont chauffées. Les vitrages sont de type double vitrage sauf pour le Nord où ils sont en triple vitrage.



(a) Bâtiment réel



(b) Modèle EnergyPlus

FIGURE 3. Maisons INCAS-DM. Faces Sud (grandes baies vitrées et balcon) et Ouest.

Le modèle est simplifié notamment au niveau du système qui est pris comme idéal et aucun scénario n'est considéré, il n'y a pas de renouvellement d'air et seules des faibles infiltrations mesurées in-situ sont prises en compte (0,6 Vol/h). Nous souhaitons en effet dans cette étude, vérifier la réponse du bâtiment au niveau de l'enveloppe et minimiser les incertitudes liées au système complexe réel de cette maison qui est une VMC double flux. Il s'agit au niveau des réponses de ce bâtiment d'étudier les besoins de chauffage. On souhaite par cette étude illustrer sur un modèle simple la méthodologie de propagation des incertitudes et d'estimation de la sensibilité avec des variables temporelles corrélées. On étudie les besoins de chauffage de la zone thermique "RDJ" et "Etage" pour visualiser les différences de besoins entre zone.

4.2 RÉSULTATS DES INCERTITUDES ET INFLUENCES

L'incertitude sur chaque sortie nous permet d'évaluer l'intervalle de confiance sur les besoins de chauffage au mois de janvier dû à la variabilité naturelle de la météo générée en section 3.2. Les résultats, en 4a, sont sous forme de boxplot aussi appelé boîte à moustache : la "boîte" délimite les 50% d'informations et les "moustaches" un seuil de 99%. Cette représentation permet l'observation d'échantillon statistiques sans hypothèse forte sur la loi de probabilité. L'incertitude étant assimilable à une loi normale nous estimons moyenne et écart type ainsi que l'intervalle d'incertitude sur la performance du bâtiment à 95% de seuil de confiance, les résultats sont dans le tableau 2. La dispersion sur le besoin de chauffage due à la variabilité météo s'élève de près de 20% à 95% de niveau de confiance (tableau 2) de la valeur obtenue avec le fichier IWEC (étoile verte sur le graphe). On observe une légère différence d'incertitude entre le RDJ et l'Etage. Pour

comprendre cet écart, on estime les variables responsables des incertitudes de sortie pour les deux zones.

Les indices de sensibilité qui représentent l'effet seul de chaque variable sont estimés. Les résultats sont en figure 4b. Étant normalisés, les indices sont compris entre 0 et 1, plus l'indice est grand et plus l'effet de la variable sur la sortie est important. Dans le cas du RDJ, près de 70% des incertitudes de sortie sont expliquées par la variabilité de la température, le reste (30%) provient des incertitudes liées au rayonnement direct. La température et le rayonnement direct expliquent l'ensemble des variations de sorties et ce pour les deux zones RDJ et Etage. On observe une différence significative entre l'influence du rayonnement direct au RDJ où l'incertitude du rayonnement direct explique plus de 30% de l'intervalle de confiance sur les besoins de chauffage au RDJ contre trois fois moins à l'Etage. Cette différence s'explique par le pourcentage de surface vitrée au RDJ par rapport à l'Etage notamment au Sud, l'obstruction par le balcon des surfaces vitrées sud à l'étage diminue également l'influence des gains solaires.

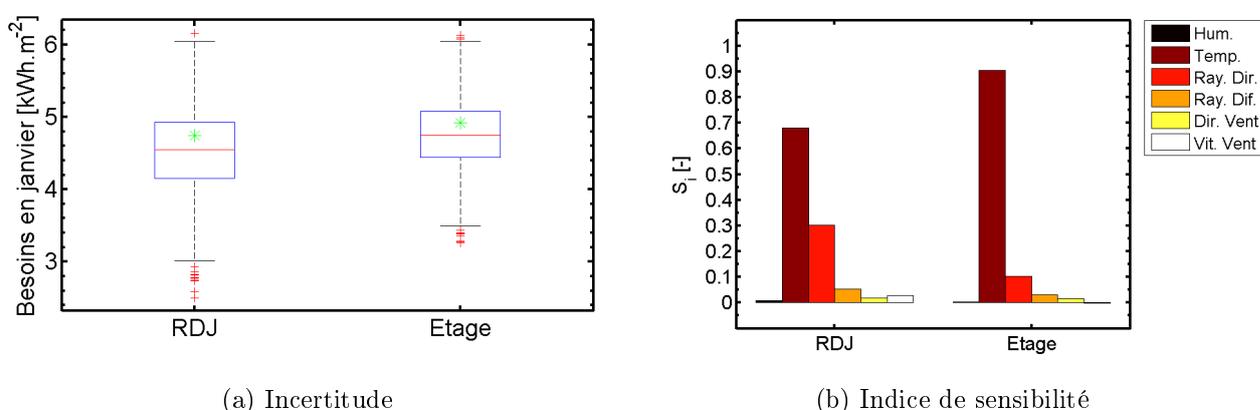


FIGURE 4. Résultats de l'étude pour les besoins de chauffage des deux zones RDJ et Etage du mois de janvier à Lyon. En a) l'incertitude sous forme de boxplot avec en étoile la valeur de référence (IWEC) et en b) les indices de sensibilité de chaque entrée sur la dispersion de sortie. La variabilité de la température et du rayonnement direct explique l'ensemble de la dispersion de sortie observée en a).

Besoins de chauffage	Moyenne μ	Écart type σ	Intervalle de confiance	Variation
RDJ [W/m^2]	4,54	0,564	4, 54 \pm 1, 106	24,4%
Etage [W/m^2]	4,76	0,479	4, 76 \pm 0, 939	19,7%

TABLE 2: Valeurs de dispersion du besoin de chauffage total et par zone pour Lyon au mois de janvier

La somme des indices de sensibilité est proche de 1 pour les deux zones. Nous pouvons en conclure que la totalité de l'information est présente et donc qu'il n'y a pas d'interactions entre variables dans le modèle. Ainsi, le calcul des effets principaux est suffisant et les effets totaux n'apporteraient pas d'informations supplémentaires.

5 CONCLUSION ET PERSPECTIVE

Les exemples de génération présentés ici font partie d'une démarche exploratoire sur la gestion des analyses de sensibilité et d'incertitude lorsque les entrées incertaines sont temporelles (auto-corrélées) et inter-corrélées. Nous avons généré une dispersion cohérente avec la variabilité naturelle pour chaque variable et respecté des contraintes d'indépendance de génération aléatoire. Les échantillons que l'on génère sont cohérents avec la variabilité naturelle et respectent la

physique sous-jacente des données météorologiques. On caractérise l'influence du bâtiment aux entrées météorologiques, permettant un gain de transparence sur la gestion bioclimatique du bâti mais également d'évaluer un intervalle de confiance sur les résultats en prenant en compte une variabilité annuelle cohérente de la météo. On obtient des premiers résultats encourageants.

La méthode générique développée peut être applicable quelque soit le modèle puisqu'elle est non intrusive et indépendante du modèle, et peut prendre en compte les entrées présentant une dépendance temporelle ou même spatiale en plus des corrélations entre variables (entrées fonctionnelles corrélées). Dans ce contexte on peut envisager d'évaluer les incertitudes des performances d'un bâtiment en fonction de sa localisation à partir d'une variabilité spatiale dans un contexte d'îlot de chaleur urbain ou bien de micro-climat.

Ce développement des méthodes statistiques dédiées à la STD peut permettre d'aller vers la garantie de performance dans le cas où le modèle prendrait en compte les consommations et non les besoins et que la prise en compte de l'occupant et des usages dans les entrées incertaines soit faite. La réduction de modèle est aussi une piste à explorer pour limiter le coût de simulation, bien que la gestion de l'aspect dynamique peut s'avérer délicate.

6 REMERCIEMENTS

Ces travaux ont été financés par le projet ANR FIABILITE n° ANR-10-HABISOL-0004

RÉFÉRENCES

- Chastaing, G., Gamboa, F., et Prieur, C. (2012). Generalized hoeffding-sobol decomposition for dependent variables. application to sensitivity analysis. *Electronic Journal of Statistics*, 6 :2420–2448.
- Goffart, J. (2013). *Impact de la variabilité des données météorologiques sur une maison basse consommation. Application des analyses de sensibilité pour des entrées temporelles*. PhD thesis, Université de Grenoble.
- Iman, R. et Conover, W. (1982). A distribution free approach to inducing a rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11 :311–334.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., et Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons.
- Sobol, I., Tarantola, S., Gatelli, D., Kucherenko, S., et Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety*, 92 :957–960.
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiment*, 1 :407–414.
- Tian, W. (2013). A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20 :411–419.