

## **Sujet de thèse : Méthodes de classification dynamique pour améliorer la précision de la consommation énergétique dans un contexte de rénovation de logements**

**Mots clés :** énergie, bâtiment, dynamique de comportement des occupants, classification de données de panel, modèles dynamiques à variables latentes, inférence variationnelle bayésienne

### **Contexte**

Le bâtiment fait partie des postes les plus significatifs en matière de consommation énergétique. L'amélioration des performances énergétiques constitue donc un enjeu majeur dans ce secteur. En France, de nombreuses campagnes de rénovation de bâtiments visant à répondre à cet objectif sont entreprises, notamment dans le secteur du logement social. Celles-ci s'appuient généralement sur des modèles numériques de simulation (ex. TRNSYS, EnergyPlus) qui permettent d'estimer la consommation énergétique en amont et en aval des opérations de rénovation, permettant ainsi d'évaluer en phase de conception, les gains énergétiques potentiels et d'ajuster les opérations en conséquence [5]. Ces modèles reposent sur des équations mathématiques traduisant les principes physiques associés aux mécanismes de transfert énergétique des bâtiments. Cependant, malgré le potentiel explicatif de ces modèles, un écart significatif est souvent observé entre les performances prédites et les performances réelles [3]. Cet écart est en partie lié à la prise en compte trop rigide du comportement réel des occupants des bâtiments qui, en réalité, est de nature complexe.

### **Objectifs**

Menée dans le cadre du projet I-Site Impulsion ANDRE, cette thèse vise à exploiter les données énergétiques hétérogènes issues d'un parc de logements sociaux géré par l'office public MC Habitat, pour décrire de manière fine les activités des ménages avant et après des opérations de rénovation. Les données qui seront exploitées sont relatives aux caractéristiques thermiques et architecturales des bâtiments, à la consommation énergétique observée au niveau des ménages (consommation d'électricité, ou de gaz, températures indoor, contrôle-commande des appareils de régulation) et au contexte météorologique (température outdoor, taux d'humidité, vitesse du vent, ensoleillement, données pluviométriques). Une partie de ces données sera collectée avant les rénovations à l'aide de capteurs *ad hoc* installés pour les besoins de l'étude, et la seconde partie sera issue d'équipements propres aux rénovations. La présente thèse pourra également bénéficier indirectement de données issues d'enquêtes visant à expliquer le comportement des ménages par leurs caractéristiques socioéconomiques et sociodémographiques, ainsi que par les caractéristiques propres des bâtiments.

Adossée au projet I-Site ANDRE, cette thèse vise à exploiter les données énergétiques hétérogènes issues d'un parc de logements sociaux géré par l'office public MC Habitat, pour décrire de manière fine les activités des ménages avant et après des opérations de rénovation. Les données qui seront exploitées sont relatives aux caractéristiques thermiques et architecturales des bâtiments, à la consommation énergétique observée au niveau des ménages (consommation d'électricité, ou de gaz, températures indoor, contrôle-commande des appareils de régulation) et au contexte météorologique (température outdoor, taux d'humidité, vitesse du vent, ensoleillement, données pluviométriques). Une partie de ces données sera collectée avant les rénovations à l'aide de capteurs *ad hoc* installés pour les besoins de l'étude, et la seconde partie sera issue d'équipements propres aux rénovations. La présente thèse pourra également bénéficier indirectement de données issues d'enquêtes visant à expliquer le comportement des ménages par leurs caractéristiques socioéconomiques et sociodémographiques, ainsi que par les caractéristiques propres des bâtiments.

L'objectif de la thèse est la réduction du biais de performance énergétique entre les prévisions initiales issues des modèles numériques et les consommations énergétiques observées en situation réelle, en se basant sur une caractérisation plus fine du comportement des ménages qui sera inféré à partir de données. En terme méthodologique, le principal problème abordé est celui de la prise en compte du caractère dynamique du comportement des occupants des bâtiments, qui peut être attribué à des facteurs internes et externes. Il s'agira dans un premier temps d'identifier, à l'aide de modèles statistiques à variables latentes, des classes de comportements homogènes au sens de la dynamique de variation de la consommation d'énergie relativement aux variables contextuelles qui devront être sélectionnées parmi les données disponibles (températures, régulation des appareils, humidité, ensoleillement...). Dans un second temps, l'intégration aux modèles physiques des classes de comportements obtenues dans l'étape précédente sera effectuée et les performances en termes de prévision énergétique seront analysées.

Ces travaux s'inscrivent dans la continuité des recherches menées au GRETTIA sur l'extraction des habitudes de consommation dans les réseaux intelligents d'électricité et d'eau [2] [6], ainsi que sur les modèles à variables latentes dynamiques [4]. Ils seront généralisés avec la prise en compte de facteurs endogènes et exogènes. Il faut noter que dans le domaine de la gestion de l'énergie des bâtiments, les profils d'usage traduisant le comportement des occupants sont souvent extraits de manière statique, sans nécessairement tenir compte de l'environnement dynamique [7] [9], ce qui n'est pas toujours réaliste en pratique. Pour l'estimation des paramètres du modèle statistique qui sera développé, une attention particulière sera portée aux méthodes d'inférence variationnelle bayésienne [1] qui se prêtent bien à la nature (markovienne) dynamique de ce type de modèle. Elles permettent d'approcher la loi a posteriori des variables latentes, qui est généralement difficile à calculer et à maximiser, par un minorant de celle-ci plus simple à formuler et plus rapide à optimiser. En outre, ces méthodes permettent aussi d'estimer directement les paramètres structurels tel que le nombre de classes. Des extensions « profondes » du modèle développé pourront également être étudiées au cours de la thèse (deep learning) à l'image des travaux initiés récemment pour les modèles de mélange [8].

Pour répondre aux objectifs opérationnels visés, la thèse pourra aussi s'appuyer sur le partenariat issu du projet I-site Impulsion ANDRE, notamment MC habitat (office public assurant la gestion d'un parc de logements sociaux et qui supervise les opérations de rénovations de ce parc), les laboratoires LATTIS et Lab'Urba (expertise sur la modélisation du comportement des ménages à partir de données d'enquêtes), ainsi que l'ESIEE et le LISIS (expertise sur la modélisation physique, énergétique et sur les méthodes inverses).

## **Déroulement**

Le planing prévu pour cette thèse est le suivant :

- 1) Etude bibliographique sur l'utilisation, dans les modèles numériques de simulation de la consommation énergétique de bâtiments, d'approches d'extraction automatique de profils d'usage à partir de données;
- 2) Prétraitement des données collectées avant et après les opérations de rénovation : préparation des données en vue des traitements statistiques ultérieurs (nettoyage, comblement des valeurs manquantes, interpolation, croisement de séries temporelles issues de différentes sources) ;
- 3) Modélisation statistique à variables latentes en vue de l'extraction automatique de classes associées aux principaux comportements des occupants des bâtiments ; mise en œuvre de méthodes bayésiennes variationnelles pour l'estimation des paramètres de ces modèles ;

4) Intégration au processus de prévision énergétique de la méthodologie d'extraction des profils d'activités dynamiques des ménages.

### Références bibliographiques

- [1] Blei D. M., Kucukelbir A., and McAuliffe J. D. (2017). Variational inference: a review for statisticians. *Journal of the American Statistical Association*, vol. 112 (518), pp. 859-877.
- [2] Cheifetz N., Noumir Z., Samé A., Sandraz A.-C., Féliers C. and Heim V. (2017). Modeling and Clustering Water Demand Patterns from Real-World Smart Meter Data. *DWES* 10, 75-82.
- [3] De Wilde Pieter (2014). The gap between predicted and measured energy performance of buildings: a framework for investigation. *Automation in Construction*, vol 41, pp 40-49.
- [4] El Assaad H., Samé A., Govaert G., and Aknin P. (2016). A variational Expectation-Maximization algorithm for temporal data clustering. *CSDA*, vol. 103, pp. 206-228.
- [5] Marrone P., Gori P., Asdrubali F., Evangelisti L., Calcagnini L., Grazieschi G. (2018). Energy benchmarking in educational buildings through cluster analysis of energy retrofitting. *Energies*, 11(3), 649.
- [6] Melzi F. N., Samé A., Zayani M. H. and Oukhellou L. (2017). A Dedicated mixture model for clustering smart meter data: identification and analysis of electricity consumption behaviors. *Energies*, 10 (10).
- [7] Pan S., Wang X., Wie Y., et al. (2017). Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in shanghai residences. *Building Simulation*, vol. 10 (6), pp. 889-898.
- [8] Viroli C., and McLachlan J. G. (2017). Deep Gaussian mixture models. *Statistics and Computing*. <https://doi.org/10.1007/s11222-017-9793-z>.
- [9] Yu Z., Fung B. C. M., Haghighat F., Yoshino H., Morofsky E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, vol. 43(6), pp. 1409-1417.

### Profil du candidat

Le candidat recherché pour cette thèse sera titulaire d'un diplôme équivalent Master 2 orienté Science de données ou Statistique, avec un intérêt pour les applications. Une bonne connaissance des langages R, Matlab et/ou Python est également demandée.

**Contact** : Allou Samé (allou.same@ifsttar.fr) ; Latifa Oukhellou (latifa.oukhellou@ifsttar.fr)