# Post-doctorate position : Multi-scale Bayesian prediction in the presence of categorical and hidden variables – application to the prediction of energy consumption in buildings

**Context** : This project is part of the chair MIAI / BALTEEC (Bayesian Active Learning Techniques for Energy Efficient buildings) that involves 3 universities (Clermont-Ferrand, Grenoble, Chambery), private companies (URBS, Heliocity) and IFPEN.

There is a large interest in the prediction of energy consumption in housing, both in private houses and in apartments to improve knowledge of construction assets and define renovation policies. 20% of the housing in France has undergone a diagnostic of energy performance in the last 5 years. This constitutes an important open data source that should help in predicting what happens in the 80% remaining cases.

Predicting energy consumption in buildings offers interesting mathematical challenges and opportunities : it is a problem where the inputs are mixed, i.e., both continuous and categorical; not all the input data is always known; the output depends on the occupant behavior, which is usually unknown and treated as an additional source of uncertainty; and statistical models can cooperate with physical (thermal) models.

Predicting energy consumption has been formulated as Bayesian learning problems where a target building is compared to a typology of buildings (Araldi et al., 2021; Sokol et al., 2017; Grossouvre et al., 2024). One advantage of the Bayesian approach is to provide a way to quantify the prediction uncertainty.

**Planned work** : the work planned is primarily in the field of applied mathematics and computational statistics and concerns design of experiments, Bayesian models and kernel methods (Gaussian process regression) for classification and/or regression (Rasmussen and Williams 2006, Steinwart and Christmann 2008). It is secondarily a work in energy consumption at the housing or building scale.

Two different problems underlie this project:

1) Bayesian prediction with mixed (continuous/categorical) and partly hidden variables. Bayesian models aim to predict a specific response (e.g., the energy consumption of a housing) with associated uncertainty. In our setting, some of the covariates are not observed at the time of prediction, but knowledge about their distribution and their effect on the outcome of interest can be gathered through existing data (e.g., recent energy performance assessments). This problem can be seen as a problem of successive Bayesian predictions, inferring the non-observed covariates before inferring the response.

2) Multi-scale Bayesian prediction. We will consider models which have a large scale output that partly depends on lower scale outputs: this modeling choice is inspired from buildings (large scale) whose total energy consumption depends in part on the consumption of the apartments (lower scale) it contains. Observations exist at both

scales, but need not be exhaustive. How to take advantage of such a nested, multi-scale, data structure in a Bayesian prediction ? Upscaling (predicting the large scale from partial low scales covariates) and downscaling (the inverse) are both of practical interest.

Both research directions are of interest. Either one or both will be pursued depending on the candidate's expertise and preferences

**Profile of candidate** : PhD in applied mathematics (probability / statistics / optimization) or PhD in engineering with a strong taste for statistical models (typically Bayesian models, Gaussian process Regression), motivation to work on environmental-related issues in particular energy in buildings.

**Team** : the collaborators are Didier Rullière (LIMOS), Rodolphe Le Riche (LIMOS), Antoine Chatalic (GIPSA-lab) and Marc Grossouvre (URBS). Frequent interactions with other members of the BALTEEC chair are anticipated.

**Duration, starting date** : 1 year, can start in March 2026 or later until the position is filled

**Location** : LIMOS laboratory in Clermont-Ferrand or Saint-Etienne, France. Possibilities to be located in the Grenoble area (to discuss).

**Salary** : up to 3700 euros (before taxes, "brut") = 2900 euros (after taxes, "net"), depending on experience

**Contacts** : Rodolphe Le Riche (rodolphe.le_riche@uca.fr), Marc Grossouvre (marcgrossouvre@urbs.fr), Clémentine Prieur (co-chair BALTEEC, clementine.prieur@univ-grenoble-alpes.fr).

**References**
- Araldi, A., Emsellem, D., Fusco, G., Tettamanzi, A., & Overal, D. (2021). Exploring building typologies through fast iterative Bayesian clustering. In SAGEO'2021 (pp. 113-124).
- Marc Grossouvre , Didier Rullière and Jonathan Villot, Predicting missing Energy Performance Certificates: Spatial interpolation of mixture distributions, Energy and IA, 2024.
- C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006.
- Julia Sokol, Carlos Cerezo Davila, Christoph F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy and Buildings, Volume 134, 2017.
- Ingo Steinwart and Andreas Christmann, Support Vector Machines Springer, 2008.
- Junlu Yang, Hanning Wang, Linmiao Cheng, Zhi Gao, Fusuo Xu, A review of resistance–capacitance thermal network model in urban building energy simulations, Energy and Buildings, Volume 323, 2024.